

GODUNOV METHODS

Theory and Applications

Volume II

Edited by E. F. Toro

GODUNOV METHODS

THEORY AND APPLICATIONS

GODUNOV METHODS

THEORY AND APPLICATIONS

Edited by

E. F. Toro

*Manchester Metropolitan University
Manchester, UK*

Springer Science+Business Media, LLC

Library of Congress Cataloging-in-Publication Data

Godunov methods: theory and applications/edited by E.F. Toro.

p. cm.

Includes bibliographical references and index.

ISBN 978-1-4613-5183-2 ISBN 978-1-4615-0663-8 (eBook)

DOI 10.1007/978-1-4615-0663-8

1. Fluid dynamics—Congresses. 2. Conservation laws (Mathematics)—Congresses. 3.

Differential equations, Hyperbolic—Numerical solutions—Congresses. I. Toro, E. F. II.

International Conference on Godunov Methods: Theory and Applications (1999: Oxford, England)

QA911 .G57 2001

532.05—dc21

2001038336

Proceedings of an International Conference on Godunov Methods: Theory and Applications, held October 18–22, 1999, in Oxford, UK, to honor Professor S. K. Godunov in the year of his 70th birthday and to review more than 40 years of research on Godunov methods

© 2001 Springer Science+Business Media New York

Originally published by Kluwer Academic / Plenum Publishers in 2001

Softcover reprint of the hardcover 1st edition 2001

<http://www.wkap.nl/>

10 9 8 7 6 5 4 3 2 1

A C.I.P. record for this book is available from the Library of Congress

All rights reserved

No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise, without written permission from the Publisher

PREFACE

This edited review book on Godunov methods contains 97 articles, all of which were presented at the international conference on *Godunov Methods: Theory and Applications*, held at Oxford in October 1999, to commemorate the 70th birthday of the Russian mathematician Sergei K. Godunov. The meeting enjoyed the participation of 140 scientists from 20 countries; one of the participants commented: *every one is here*, meaning that virtually everybody who had made a significant contribution to the general area of numerical methods for hyperbolic conservation laws, along the lines first proposed by Godunov in the fifties, was present at the meeting. Sadly, there were important absentees, who due to personal circumstance could not attend this very exciting gathering. The central theme of the meeting, and of this book, was numerical methods for hyperbolic conservation laws following Godunov's key ideas contained in his celebrated paper of 1959. But Godunov's contributions to science are not restricted to *Godunov's method*. He has indeed made very significant contributions to other areas in applied and computational mathematics; as one of the plenary speakers put it: *I know no other living scientist who has made so many contributions to so diverse areas of research*. The breadth of topics covered in the conference, such as Thermodynamics, theory of conservation laws, stability theory and numerical linear algebra, reflected Godunov' wide-ranging contributions to science. The fundamental contributions of Godunov to the theoretical and numerical aspects of hyperbolic conservation laws have played a key role in the research of the subject for half a century. Moreover, the ever increasing interest on the subject has undergone an explosive development in the last two decades or so; as an example, it may be appropriate to mention here that the European community has identified this topic as a key research area and has provided substantial financial support for basic and applied research over the last few years. The stimulus for further research activity does not come from mathematicians and numerical analysts alone. Hyperbolic conservation laws play a central role in mathematical modelling in

several distinct disciplines of science and technology. Application areas include compressible, single (and multiphase) fluid dynamics, shock waves, meteorology, elasticity, magnetohydrodynamics, relativity, and many others. The successes in the design and application of new and improved numerical methods for hyperbolic conservation laws in the last twenty years have made a dramatic impact in these application areas.

From the point of view of numerical methods for hyperbolic conservation laws, the most well-known contribution of Godunov is the idea of utilising the solution of local Riemann problems in the construction of discretisation schemes; this is known today as Godunov's method. This approach, coupled to ways of extending the accuracy above *one* in smooth regions, has become very successful. Moreover, the utilisation of Riemann problem solutions is today not restricted to just the *finite volume method*. The Riemann problem is also used in other numerical approaches, such as *front tracking*. More recently the Riemann problem has been incorporated into two very distinct numerical approaches, namely the *discontinuous Galerkin finite element method* and the *smooth particle hydrodynamics* method. The so-called Godunov's theorem (this is of course one of his many theorems) is another example of Godunov's contributions to numerical methods. In simple terms, Godunov's theorem says that *all (linear) methods of accuracy greater than 1 will produce spurious oscillations in the vicinity of large gradients*; 50 years after the statement of the theorem, the difficulty of reconciling (a) the absence of spurious oscillations and (b) high accuracy, remain to be resolved for numerical methods intended for realistic problems. Notable progress has been made on this and related topics but there is still much work to be done both on theoretical aspects of hyperbolic conservation laws and on the design, analysis and application of new and better numerical methods.

The papers included in this book more or less summarise the state of the art on the subject and provide the basis for future research. The 97 papers cover a very wide range of topics such as design and analysis of numerical schemes, applications to compressible and incompressible fluid dynamics, multi-phase flows, combustion problems, astrophysics, environmental fluid dynamics, detonations waves and many others. The papers

have been organised in strict alphabetical order, according to the first author's name. There are long and short papers, which correspond to plenary and contributed presentations at the conference. All contributed papers were accepted after a careful refereeing process. Of the plenary presentation papers included here, there is one on *Godunov Methods*, by Peter K Sweby. This paper gives a useful introductory overview of the current state of Godunov-type methods.

This one-off meeting was possible, fundamentally due to the enthusiasm of all participants who, due to the very limited number of sponsors I managed to enlist, had to bear most of the organisation costs. I am very pleased to thank the support given by The London Mathematical Society (LMS), The Defence and Evaluation Research Agency (DERA), John Wiley and Sons Limited, Springer-Verlag and Numeritek Limited. My sincere thanks to Steve Higgins, Chris Brand and Gaynor Swarbrick for their help with the organisation of the conference and the handling and compilation of the papers for this volume.

E F Toro OBE

Manchester UK 2001

CONTENTS

Olešnik's E-Condition from the Viewpoint of Numerics <i>H. Aiso</i>	1
On Some New Results for Residual Distribution Schemes <i>R. Abgrall and T. J. Barth</i>	27
Simulations of Relativistic Jets with Genesis <i>M. A. Aloy, J. M. Ibanyez, J. M. Marti, J. L. Gomez, and E. Mueller</i>	45
Relativistic Jets from Collapsars <i>M. A. Aloy, E. Mueller, J. M. Ibanyez, J. M. Marti, and A. Macfadyen</i>	53
Exact Computation in Numerical Linear Algebra: The Discrete Fourier Transform <i>J. A. D. W. Anderson and P. K. Sweby</i>	61
Comparative Study of HLL, HLLC and Hybrid Riemann Solvers in Unsteady Compressible Flows <i>A. Bagabir and D. Drikakis</i>	69
A New Reconstruction Technique for the Euler Equations of Gas Dynamics with Source Terms <i>P. Bartsch and A. Borzi</i>	77
Colella-Glaz Splitting Scheme for Thermally Perfect Gases <i>A. Beccantini</i>	89
Meshless Particle Methods: Recent Developments for Nonlinear Conservation Laws in Bounded Domain <i>B. Ben Moussa</i>	97
Application of Wave-propagation Algorithm to Two-dimensional Thermoelastic Wave Propagation in Inhomogeneous Media <i>A. Berezovski and G. A. Maugin</i>	109
Unstructured Mesh Solvers for Hyperbolic PDEs with Source Terms: Error Estimates and Mesh Quality	117

M. Berzins and L. J. K. Durbeck

- Constancy Preserving, Conservative Methods for Free-surface Models 125

L. Bonaventura and E. Gross

- Hyperbolic-elliptic Splitting for the Pseudo-compressible Euler Equations 135

A. Bonfiglioli

- Godunov Solution of Shallow Water Equations on Curvilinear and Quadtree Grids 141

A. G. L. Borthwick, M. Fujihara, and B. D. Rogers

- A High-order-accurate Reconstruction for the Computation of Compressible Flows on Cell-vertex Triangular Grids 149

L. A. Catalano

- Numerical Experiments with Multilevel Schemes for Conservation Laws 155

G. Chiavassa and R. Donat

- Volume-of-fluid Methods for Partial Differential Equations 161

P. Colella

- Some New Godunov and Related Relaxation Methods for Two-phase Flow Problems 179

F. Coquel, E. Godlewski, B. Perthame, A. In, and P. Rascle

- Development of Genuinely Multi-dimensional Upwind Residual Distribution Schemes for the System of Eight Wave Ideal

- Magnetohydrodynamic Equations on Unstructured Grids 189

A. Csik, H. Deconinck, and S. Poedts

- Application of TVD High Resolution Schemes to the Viscous Shock Tube Problem 197

V. Daru and C. Tenaud

- Comparison of Numerical Solvers with Godunov Scheme for Multicomponent Turbulent Flows 203

E. Declercq

- Godunov-type Schemes for the MHD Equations 209

<i>A. Dedner, D. Kröner, C. Rohde, and M. Wesenberg</i>	
Absorbing Boundary Conditions for Astrophysical MHD Simulations	217
<i>A. Dedner, D. Kröner, M. Wesenberg, and I. Sofronov</i>	
About Kinetic Schemes Built in Axisymmetrical and Spherical Geometries	225
<i>S. Dellacherie</i>	
Lagrangian Systems of Conservation Laws and Approximate Riemann Solvers	233
<i>B. Després</i>	
Intermediate Shocks in 3D MHD Bow Shock Flows	247
<i>H. De Sterck and S. Poedts</i>	
A Second Order Godunov-type Scheme for Naval Hydrodynamics	253
<i>A. Di Mascio, R. Broglia, and B. Favini</i>	
Uniformly High-order Methods for Unsteady Incompressible Flows	263
<i>D. Drikakis</i>	
Application of the Finite Volume Method with Osher Scheme and Split Technique for Different Types of Flow in a Channel	285
<i>K. S. Erduran and V. Kutija</i>	
A-priori Estimates for a Semi-Lagrangian Scheme for the Wave Equation	293
<i>M. Falcone and R. Ferretti</i>	
Interstellar Shock Structures in Weakly Ionised Gases	301
<i>S. A. G. E. Falle</i>	
The Ghost Fluid Method for Numerical Treatment of Discontinuities and Interfaces	309
<i>R. P. Fedkiw</i>	
A Hybrid Primitive-Conservative Upwind Scheme for the Drift Flux Model	319
<i>K. K. Fjelde and K. H. Karlsen</i>	

Numerical Simulations of Relativistic Wind Accretion onto Black Holes Using Godunov-type Methods <i>J. A. Font, J. M. Ibáñez, and P. Papadopoulos</i>	327
A Second Order Accurate, Space-time Limited, BDF Scheme for the Linear Advection Equation <i>S. A. Forth</i>	335
Multidimensional Upwind Schemes: Application to Hydraulics <i>P. García-Navarro, M. Hubbard, and P. Brufau</i>	343
HELMIT - A New Interface Reconstruction Algorithm <i>R. Giddings</i>	367
A Godunov-type Method for Studying the Linearised Stability of a Flow. Application to the Richtmyer-Meshkov Instability <i>E. Godlewski, M. Olazabal, and P. A. Raviart</i>	377
Thermodynamics, Conservation Laws and their Rotation Invariance <i>S. K. Godunov</i>	399
A New Limiter that Improves TVD-MUSCL Schemes <i>L. Gozalo and R. Abgrall</i>	411
Exact Roe Linearisation for van der Waals' Gas <i>A. Guardone and L. Quartapelle</i>	419
The Godunov-Ryabenkii Condition: The Beginning of a New Stability Theory <i>B. Gustafsson</i>	425
A Front Tracking Method for Hybrid Grids <i>D. Hänel, L. Tran, and R. Vilsmeier</i>	445
A Problem of Classical Shock Capturing Finite Volume Schemes in Hypersonic Flows <i>V. Hannemann</i>	453
Orientation Effects on Bent Extragalactic Jets <i>S. Higgins, T. O'Brien, and J. Dunlop</i>	461

Operator Splitting for Convection-dominated Nonlinear Partial Differential Equations <i>H. Holden, K. H. Karlsen, K. A. Lie, and N. H. Risebro</i>	469
Balancing Source Terms and Flux Gradients in Finite Volume Schemes <i>M. E. Hubbard and P. Garcia-Navarro</i>	477
Riemann Solvers in General Relativistic Hydrodynamics <i>J. M. Ibáñez, M. A. Aloy, J. A. Font, J. M. Martí, J. A. Miralles, and J. A. Pons</i>	485
A Fully Adaptive Multiresolution Scheme for Shock Computations <i>M. K. Kaibara and S. M. Gomes</i>	497
Application of a Godunov-type ALE-method to Underwater Shock Waves <i>A. Klomfass, P. Neuwald, and K. Thoma</i>	505
Numerical Simulation of 2-D Two-phase Flows with Interface <i>S. Kokh and G. Allaire</i>	513
Relativistic MHD Simulations Using a Godunov-type Method <i>S. Komissarov</i>	519
Godunov Type Methods on Unstructured Grids and Local Mesh Refinement <i>D. Kröner and T. Gessner</i>	527
3D Visualization of Shock Waves Using Volume Rendering <i>J. O. Langseth</i>	549
Gas Flows Generated by Propellant Burning <i>C. A. Lowe and J. F. Clarke</i>	557
Finite Volume Evolution Galerkin Methods for Multidimensional Hyperbolic Systems <i>M. Lukáčová-Medvidová, K. W. Morton, and G. Warnecke</i>	571
The Numerical Simulation of Relativistic Fluid Flow with Strong Shocks <i>A. Marquina</i>	577

An Artificial Compression Procedure Via Flux Correction <i>V. Martínez</i>	595
A Second-order Time-splitting Technique for Advection-Dispersion Equation on Unstructured Grids <i>A. Mazzia, L. Bergamaschi, and M. Putti</i>	603
Towards Implicit Godunov Method: Exact Linearisation of the Numerical Flux <i>I. Men'shov and Y. Nakamura</i>	611
Mass Flux Computation as a Key to the Carbuncle Phenomenon <i>J.-M. Moschetta</i>	623
On the Positivity of FVS Schemes <i>J.-M. Moschetta and J. Gressier</i>	631
The Carbuncle Phenomenon: a Genuine Euler Instability? <i>J.-M. Moschetta, J. Gressier, J.C. Robinet, and G. Casalis</i>	639
A Godunov-type Solver for the Maxwell Equations with Divergence Cleaning <i>C.-D. Munz, P. Omnes ,and R. Schneider</i>	647
Convergence of Kinetic Approximation to Nonlinear Parabolic Problems <i>G.Naldi, L. Pareschi, and G. Toscani</i>	655
On Options for the Numerical Modelling of the Diffusion Term in River Pollution Simulations <i>S. Neelz, S. G. Wallis, and J. R. Manson</i>	663
Multidimensional Flux-vector-splitting and High-resolution Characteristic Schemes <i>S. Noelle</i>	671
A Comparison of Roe, VFFC and AUSM+ Schemes for Two-phase Water/Steam Flows <i>H. Paillyere, A. Kumbaro, C. Viozat, S. Clerc, A. Broquet, and C. Corre</i>	677
Low Dissipation Entropy Fix for Positivity Preserving Roe's Scheme	685

M. Pelanti, L. Quartapelle, and L. Vigevano

Bicharacteristic Methods for Multidimensional Hyperbolic Systems <i>M. H. Pham, R. Rudgyard, and E. Süli</i>	691
An exact Riemann Solver for Multidimensional Special Relativistic Hydrodynamics <i>J. A. Pons, J. M. Marti, and E. Mueller</i>	699
Experience with the Osher Scheme for Applied Aerodynamics <i>N. Qin</i>	707
A High-resolution Godunov Method for Modelling Anomalous Fluid Behaviour <i>W. J. Rider and J. W. Bates</i>	717
Towards Godunov-type Methods for Hyperbolic Conservation Laws with Stiff Relaxation <i>P. L. Roe and J. A. F. Hittinger</i>	725
Thermodynamics and Hyperbolic Systems of Balance Laws in Continuum Mechanics <i>E. I. Romensky</i>	745
Development and Application of High-resolution Adaptive Numerical Techniques in Shock Wave Research Center <i>T. Saito, P. Voinovich, E. Timofeev, and K. Takayama</i>	763
Interfaces, Detonation Waves, Cavitation and the Multi-phase Godunov Method <i>R. Saurel</i>	785
One-dimensional Calculation of Unsteady Open Channel Flow Using Adaptive Mesh Refinement <i>J. Schramm, S. Enk, and J. Koenegter</i>	809
Error Estimates for Godunov-type Schemes in the Presence of Source Terms <i>H. J. Schroll</i>	815
A Multi-dimensional Euler Solver <i>R. Schwane</i>	823

MPDATA—A Multipass Donor Cell Solver for Geophysical Flows <i>P. K. Smolarkiewicz and L. G. Margolin</i>	833
On the Hyperbolic Nature of Two-phase Flow Equations: Characteristic Analysis and Related Numerical Methods <i>H. Städke, B. Worth, and G. Franchello</i>	841
A Method of Lines Flux-difference Splitting Finite Volume Approach for 1D and 2D River Flow Problems <i>G. Steinebach and A. Q. T. Ngo</i>	863
A Simple Smoothing TVD Scheme on Structured and Unstructured Grids <i>M. Sun and K. Takayama</i>	873
Godunov Methods <i>P. K. Sweby</i>	879
Centred Unsplit Finite Volume Schemes for Multidimensional Hyperbolic Conservation Laws <i>E. F. Toro and W. Hu</i>	899
Towards Very High Order Godunov Schemes <i>E. F. Toro, R. Millington, and L. A. M. Nejad</i>	907
Model Hyperbolic Systems with Source Terms: Exact and Numerical Solutions <i>E. F. Toro and M. E. Vazquez-Cendon</i>	941
A Godunov-type Method for Capturing Water Waves <i>E. H. van Brummelen and B. Koren</i>	949
A Staggered Scheme for Hyperbolic Conservation Laws Applied to Computation of Flow with Cavitation <i>D. R. van der Heul, C. Vuik, and P. Wesseling</i>	969
An Expert System to Control the CFL Number of Implicit Upwind Methods <i>D. Vanderstraeten</i>	977
Discontinuous Galerkin Methods for Hyperbolic Partial Differential Equations	985

J. J. W. van der Vegt, H. van der Ven, and O. J. Boelens

Solving Incompressible Two-phase Flows with a Coupled TVD Interface Tracking/Local Mesh Refinement Method <i>S. Vincent and J. P. Caltagirone</i>	1007
A Large Timestep Godunov-type Model of Global Atmospheric Chemistry and Transport <i>B. M.-J. B. D. Walker and N. Nikiforakis</i>	1015
Approximate Riemann Solvers, Godunov Schemes and Contact Discontinuities <i>B. Wendroff</i>	1023
A Unified Method for Compressible and Incompressible Flows with General Equation of State <i>P. Wesseling and D. R. van der Heul</i>	1057
Wave Interactions in Non-linear Strings <i>R. Young</i>	1065
Index	1073

OLEI $\check{\text{N}}$ NIK'S E-CONDITION FROM THE VIEWPOINT OF NUMERICS

AISO, HIDEAKI

*Computational Sciences Div., National Aerospace Laboratory,
Jindaiji-Higashi 7-44-1 Chofu TOKYO 182-0040 JAPAN,
E-mail:aiso@nal.go.jp*

Abstract. We are interested in difference schemes for scalar conservation laws. The consistency with Olei $\check{\text{N}}$ nik's E-condition is discussed. We prove that some class of difference schemes are consistent with the E-condition and discuss the quality of numerical result from the viewpoint of the E-condition. The consistency with E-condition gives the convergence proof in $L^\infty \cap L^1_{loc}$ -category as well.

1. Introduction

The concept of consistency, which means the property that numerical solutions obtained by a difference scheme inherit some important property of the exact solution, is used to analyze difference schemes. The convergence property, which guarantees that a numerical solution converges to the exact solution as the difference increments approach to zero, is an example of consistency. In the case of conservation laws, a large part of the discussion is devoted to the convergence property, while theoretical proof is given mainly in the case of scalar conservation laws and brings some implication to the system's case. Once the convergence is proved, it is natural to expect that the numerical solution is near enough to the exact solution with the difference increments small enough.

But in numerical computation the difference increments are always some finite numbers so that we never obtain the convergence limit of numerical solution. The numerical results often show some kind of inconvenience (numerical inconvenience) like smearing, over- or under-shoot, unnatural flexion of data (*i.e.* loss of smoothness) and so on, even if the convergence to exact solution is theoretically guaranteed. In the scalar case, it is not difficult to recognize such numerical inconvenience and it might be possi-

ble even to remove such inconvenience a posteriori (sometimes “by hand”). But the machinery that produces those phenomena is inherited to the system’s case, where it is difficult to correct those kind of error a posteriori and it is sometimes impossible even to recognize the occurrence of such phenomena. Therefore it is important to analyze the numerical inconvenience, while analysis of this kind have not yet been developed enough. But we may expect that some mathematical tools to analyze the convergence can be applied to the numerical inconvenience with a slight change of the viewpoint.

We discuss difference schemes approximating scalar conservation laws

$$u_t + f(u)_x = 0.$$

The discussion on convergence is usually based on two kinds of consistency; the consistency with weak solution (*i.e.* the convergence to a weak solution which is derived from some compactness concept such as TVD-property) and the consistency with entropy condition. It is related to the fact that the existence of exact solution is guaranteed by the concept of weak solution and the uniqueness is guaranteed by the entropy condition. The entropy condition has several different definitions which are theoretically equal. Although the definition using the entropy inequality

$$U_t + F_x \leq 0 \text{ for any entropy pair } (U, F)$$

is popular, we remember another definition, so called “Oleinik’s E-condition”

$$\frac{u(x+h, t) - u(x, t)}{h} < \frac{E}{t} \text{ for some constant } E,$$

which is used in the work (Oleinik, 1957) that gives the first proof of existence and uniqueness of entropy solution to scalar conservation law. The E-condition seems to govern the behavior of solution more directly than the entropy inequality. It implies that the analysis of the consistency with E-condition would give more information on numerical behavior of difference schemes than that of the consistency with entropy inequality.

We here consider difference schemes for scalar conservation laws from the viewpoint of the consistency with E-condition. We prove that some class of difference schemes are consistent with the E-condition. Godunov scheme and some TVD-schemes with Harten-like entropy fix are included. We also discuss the numerical behavior. We obtain some understanding of the machinery that produces flexion or jump (loss of smoothness) of numerical solution around the sonic points. This kind of numerical inconvenience is discussed by Roe (Roe, 1992) from a different point of view.

From the theoretical viewpoint, we note that the consistency with E-condition guarantees the convergence in $L^\infty \cap L^1_{loc}$ -category because E-condition itself bounds the variation of $u(*, t), t > 0$. (See (Oleinik, 1957), cf. for example, (Aiso, 1993))

2. Scalar Conservation Laws and Difference Schemes

We are concerned with difference schemes that approximate the initial value problem of conservation law

$$\begin{cases} u_t + f(u)_x = 0, -\infty < x < \infty, 0 < t < \infty, \\ u(x, 0) = u_0(t), -\infty < x < \infty. \end{cases} \quad (1)$$

We assume that the flux function f is strictly convex, i.e. $f'' > c > 0$ for some constant c , and that the initial data u_0 is of $L^\infty \cap L^1_{loc}$ and satisfies $M_l \leq u_0 \leq M_u$. For simplicity we assume $f(0) = f'(0) = 0$.

We suppose difference schemes in the viscosity form

$$u_i^{n+1} = u_i^n - \frac{\lambda}{2} \{f(u_{i+1}^n) - f(u_{i-1}^n)\} + \frac{\lambda}{2} \left\{ a_{i+\frac{1}{2}}^n (u_{i+1}^n - u_i^n) - a_{i-\frac{1}{2}}^n (u_i^n - u_{i-1}^n) \right\}. \quad (2)$$

The form consists of the central difference and the numerical viscosity. The coefficient $\lambda = \frac{\Delta t}{\Delta x}$, the ratio of the time increment Δt to the space increment Δx , should satisfy the natural CFL-condition $\lambda \sup |f'| \leq 1$. Throughout this article we assume the condition

$$\lambda \sup_{M_l \leq s \leq M_u} |f'(s)| \leq 1 \quad (3)$$

as the CFL-condition, because all the difference schemes discussed here satisfy the property that

$$\inf_j u_i^0 \leq u_i^n \leq \sup_j u_j^0, \quad \text{for every } u_i^n. \quad (4)$$

For convenience, we define $q_{i+\frac{1}{2}}^n$ by

$$q_{i+\frac{1}{2}}^n = \int_0^1 f'(u_i^n + (u_{i+1}^n - u_i^n)\theta) d\theta = \frac{f(u_{i+1}^n) - f(u_i^n)}{u_{i+1}^n - u_i^n}. \quad (5)$$

Difference schemes are specified by the choice of numerical viscosity coefficients $a_{i+\frac{1}{2}}^n$. Examples are;

$$a_{i+\frac{1}{2}}^n = a^{LF} \equiv \frac{1}{\lambda}, \quad (6)$$

$$a_{i+\frac{1}{2}}^n = a^{EO}(u_i^n, u_{i+1}^n) \equiv \int_0^1 |f'(u_i^n + (u_{i+1}^n - u_i^n)\theta)| d\theta, \quad (7)$$

$$a_{i+\frac{1}{2}}^n = a^G(u_i^n, u_{i+1}^n) \equiv \max_{(s-u_i^n)(s-u_{i+1}^n)} \frac{f(u_i^n) + f(u_{i+1}^n) - 2f(s)}{u_{i+1}^n - u_i^n}, \quad (8)$$

$$a_{i+\frac{1}{2}}^n = a^{MR}(u_i^n, u_{i+1}^n) \equiv \left| q_{i+\frac{1}{2}}^n \right|. \quad (9)$$

(6), (7), (8) and (9) give Lax-Friedrichs scheme, Enquist-Osher scheme, Godunov scheme, and Murmann-Roe scheme (upwind scheme without additional viscosity), respectively. The well known Harten's first order TVD scheme with entropy-fix is given by

$$a_{i+\frac{1}{2}}^n = \frac{1}{\lambda} Q \left(\lambda q_{i+\frac{1}{2}}^n \right), \quad (10)$$

where Q is a function given by

$$Q(s) = \max \{ |s|, \epsilon \} \text{ or } Q(s) = \begin{cases} |s|, & |s| \geq \epsilon, \\ \frac{s^2}{2\epsilon} + \frac{\epsilon}{2}, & |s| \leq \epsilon \end{cases}$$

with a positive constant ϵ less than 1.

The viscosity form (2) is equivalent to the conservation form

$$u_{i+1}^n = u_i^n - \lambda \left\{ \bar{f}_{i+\frac{1}{2}}^n - \bar{f}_{i-\frac{1}{2}}^n \right\} \quad (11)$$

if each numerical flux $\bar{f}_{i+\frac{1}{2}}^n$ is given by

$$\bar{f}_{i+\frac{1}{2}}^n = \frac{1}{2} \{ f(u_i^n) + f(u_{i+1}^n) \} - \frac{1}{2} a_{i+\frac{1}{2}}^n (u_{i+1}^n - u_i^n). \quad (12)$$

We review the solution to problem (1) briefly. Even with smooth initial data u_0 , the existence of global smooth solution $u = u(x, t)$, $-\infty < x < \infty$, $0 \leq t < \infty$ might be violated because of the nonlinearity of flux function f . We introduce the concept of weak solution to guarantee the existence of global solution, while the uniqueness of solution is lost because the problem may allow infinitely many weak solutions. Then we impose an additional condition, which is called the entropy condition, to choose a unique solution from the infinitely many weak solutions. (Oleinik, 1957; Smoller, 1982; Vol'pert, 1967; Vol'pert, 1985) The unique solution is called the entropy solution.

We have several different statements to define the entropy condition, although they are equivalent in the sense that they select the same solution. We here remember the following two definitions.

(Definition by the Entropy Inequality) *We say that a weak solution $u = u(x, t)$ satisfies the entropy condition if the weak solution u satisfies the entropy inequality*

$$U(u)_t + F(u)_x \leq 0 \quad (13)$$

for any entropy pair (U, F) , where an entropy pair means a pair (U, F) of two functions $U(u)$ and $F(u)$ satisfying 1) U is convex, and 2) $F' = U'f'$. The entropy inequality (13) is understood in the sense of distribution.

(Definition by the E-condition) *We say that a weak solution $u = u(x, t)$ satisfies the entropy condition if the weak solution u satisfies the following inequality (E-condition)*

$$\frac{u(x+h, t) - u(x, t)}{h} < \frac{E}{t} \quad (14)$$

for some positive constant E .

In the work (Oleĭnik, 1957) an approximate solution is constructed with Lax-Friedrichs scheme and the scheme's consistency with E-condition is shown by proving the inequality

$$\frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x} < \frac{E}{n\Delta t} \text{ for some constant } E. \quad (15)$$

It means that the approximate solution's convergence limit gives the entropy solution. The inequality (15) may be called a numerical E-condition.

The work (Oleĭnik, 1957) is followed by several works in which the consistency with entropy condition is discussed. (For example, see (Aiso, 1993; Aiso, 1995; Aiso, 1996; Crandall and Majda, 1980; Engquist and Osher, 1980; Godunov, 1959; Osher, 1984; Osher and Tadmor, 1988; Tadmor, 1984).) In most of them the discussion is based on a numerical version of entropy inequality (13) (or numerical entropy inequality) like

$$\begin{aligned} & U(u_{i+1}^n) - U(u_i^n) + \frac{\lambda}{2} \{F(u_{i+1}^n) - F(u_{i-1}^n)\} \\ & - \frac{\lambda}{2} \left\{ A_{i+\frac{1}{2}}^n (u_{i+1}^n - u_i^n) - A_{i-\frac{1}{2}}^n (u_i^n - u_{i-1}^n) \right\} \leq 0. \end{aligned} \quad (16)$$

A numerical entropy inequality is a convenient tool to discuss the consistency with entropy condition because the convergence of approximate solutions is usually considered in the weak form, the same manner as the entropy inequality is understood. The fact that difference schemes are naturally understood with the concept of finite volume is another reason. On the other hand, it is not always so good an idea to discuss a numerical E-condition because some difference schemes converging to the entropy solution violates

the property that $\sup_{-\infty < i < \infty} \{u_{i+k}^n - u_i^n\}$ decreases as n increases, where k is a fixed positive integer ($k = 1$ or 2 for almost all practical cases), while it does not completely deny the existence of constant $E > 0$ and integer $k > 0$ satisfying the inequality

$$\frac{u_{i+k}^n - u_i^n}{k\Delta x} < \frac{E}{n\Delta t}. \quad (17)$$

But the E-condition (14) governs the behavior of solution to the problem (1) more directly than the entropy inequality (13) does. In the entropy solution to the problem (1), we can consider curves $x = x(t)$;

$$\frac{dx}{dt} = f'(u(x, t)) \quad (18)$$

called the characteristics. Each of the characteristics originates at some point on x -axis (*i.e.* $t = 0$). If the initial value has discontinuity, infinitely many characteristics may have a common originating point on x -axis so that a centered rarefaction is formed there. As t increases, a curve may vanish only at the shock discontinuity. The property of curves implies that

$$\frac{u(x+h, t) - u(x, t)}{h} \leq \frac{1}{f''(u(x, t))} + O(h) \quad (19)$$

and that the optimal value of constant E in the right-hand of the E-condition (14) is $E_0 = \frac{1}{\inf |f''|}$. Therefore we expect to obtain some information on the behavior of numerical solution by considering the optimal value of E for numerical E-condition like (15) or (17). It is one of the advantage when we discuss the consistency with E-condition. On the contrary, we can not expect so much information on the behavior of numerical solution when we discuss the consistency with entropy inequality.

First we discuss the consistency with E-condition in the case of Godunov scheme, and then TVD schemes with Harten-like entropy fix. We finally discuss the behavior of numerical solution given by those schemes as well.

3. Godunov Scheme

First we review that Lax-Friedrichs scheme is consistent with E-condition.

THEOREM 1 *Lax-Friedrichs scheme, which is determined by (2) and (6), satisfies the inequality (15):*

$$\frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x} < \frac{E}{n\Delta t},$$

where

$$E = 2E_0 = \frac{2}{\inf |f''|}.$$

The fact is already proved in (Oleĭnik, 1957). The scheme is a staggered, i.e. each u_i^{n+1} depends on u_{i-1}^n and u_{i+1}^n but not on u_i^n . Then $\{u_i^n\}_{i+n:\text{odd}}$ and $\{u_i^n\}_{i+n:\text{even}}$ are independent of each other. Therefore it is natural to estimate rather $u_{i+1}^n - u_{i-1}^n$ than $u_{i+1}^n - u_i^n$.

We review the proof briefly. Let us introduce the notation $z_i^n = u_{i+1}^n - u_{i-1}^n$. The definition of scheme immediately gives

$$z_i^{n+1} = \frac{1}{2}z_{i+1}^n - \frac{\lambda}{2}\{f(u_{i+2}^n) - f(u_i^n)\} + \frac{1}{2}z_{i-1}^n - \frac{\lambda}{2}\{f(u_{i-2}^n) - f(u_i^n)\}. \quad (20)$$

By the assumption $f'' > c$ we obtain

$$f(v+h) - f(v) = f'(v)h + \frac{c}{2}h^2, \quad (21)$$

and

$$\begin{aligned} z_i^{n+1} &\leq \frac{1}{2}\{1 + \lambda f'(u_i^n)\}z_{i-1}^n - \frac{\lambda c}{4}(z_{i-1}^n)^2 + \frac{1}{2}\{1 - \lambda f'(u_i^n)\}z_{i+1}^n - \frac{\lambda c}{4}(z_{i+1}^n)^2 \\ &\leq \max(z_{i-1}^n, z_{i+1}^n) - \frac{\lambda c}{4} \{\max(z_{i-1}^n, z_{i+1}^n)\}^2. \end{aligned} \quad (22)$$

When $M_l \leq v < w \leq M_u$, the inequalities

$$\begin{aligned} f(w) - f(v) &\geq f'(v)(w-v) + \frac{c}{2}(w-v)^2 \geq -\frac{1}{\lambda}(w-v) + \frac{c}{2}(w-v)^2, \\ f(w) - f(v) &\leq f'(w)(w-v) - \frac{c}{2}(w-v)^2 \leq \frac{1}{\lambda}(w-v) - \frac{c}{2}(w-v)^2 \end{aligned} \quad (23)$$

directly yield

$$-\frac{1}{\lambda}(w-v) + \frac{c}{2}(w-v)^2 \leq \frac{1}{\lambda}(w-v) - \frac{c}{2}(w-v)^2 \quad (24)$$

and then

$$w-v \leq \frac{2}{\lambda c}. \quad (25)$$

The assumed property (4) implies that

$$z_i^n < \frac{2}{\lambda c} \text{ for each } z_i^n. \quad (26)$$

We note that the quadratic function $s - \frac{\lambda c}{4}s^2$ of s is increasing over the interval $(-\infty, \frac{2}{\lambda c})$ and maps $(-\infty, \frac{2}{\lambda c})$ to $(-\infty, \frac{1}{\lambda c})$. Then we obtain

$$z^{n+1} \leq z^n - \frac{\lambda c}{4}(z^n)^2 \quad (27)$$

for the sequence $\{z^n\}$, where $z^n = \sup_i z_i^n$.

Define the function $z = z(\tau), \tau > 0$ by

$$z'(\tau) = -\frac{\lambda}{4}\{z(\tau)\}^2, \quad z(0) = z^0. \quad (28)$$

We easily observe the fact

$$z^n < z(n), n \geq 1 \quad (29)$$

and the following explicit description of $z(\tau)$.

$$z(\tau) = \frac{1}{\frac{\lambda c}{4}\tau + \frac{1}{z^0}}. \quad (30)$$

Then it follows that

$$z^n \leq \frac{1}{\frac{\lambda c}{4}n + \frac{1}{z^0}} < \frac{1}{\frac{\lambda c}{4}n} = \frac{4\Delta x}{n\Delta t \cdot c} \quad (31)$$

and

$$\frac{u_{i+1}^n - u_{i-1}^n}{2\Delta t} \leq \frac{z^n}{2\Delta t} < \frac{2}{n\Delta t \cdot c} \quad (32)$$

This completes the proof.

We see that the machinery of estimate comes mainly from the convexity of flux function f .

THEOREM 2 *Godunov scheme, which is determined by (2) and (8), satisfies the inequality*

$$\frac{u_{i+1}^n - u_i^n}{\Delta x} < \frac{E}{n\Delta t}, \quad (33)$$

where

$$E = 4E_0 = \frac{4}{\inf |f''|}. \quad (34)$$

If $u_i^n \cdot u_{i+1}^n \geq 0$ (i.e. the sonic point 0 is not between u_i^n and u_{i+1}^n), the constant E in the right-hand side of (33) can be $2E_0$ instead of (34).

The proof of theorem owes the convexity of function f as well. But it is a little more complicated because the explicit switching procedure works in the scheme according to the direction of characteristics.

The numerical viscosity coefficient of scheme is described as

$$a_{i+\frac{1}{2}}^n = a^G(u_i^n, u_{i+1}^n) = \begin{cases} |f'(u_i^n)|, & u_i^n = u_{i+1}^n, \\ \left|q_{i+\frac{1}{2}}^n\right|, & u_{i+1}^n < u_i^n \text{ or } u_i^n < u_{i+1}^n \leq 0 \text{ or } 0 \leq u_i^n < u_{i+1}^n, \\ \frac{f(u_i^n) + f(u_{i+1}^n)}{u_{i+1}^n - u_i^n}, & u_i^n < 0 < u_{i+1}^n. \end{cases} \quad (35)$$

The scheme is written in the conservation form (11) with the following numerical flux.

$$\bar{f}_{i+\frac{1}{2}}^n = \bar{f}^G(u_i^n, u_{i+1}^n) = \begin{cases} \max\{f(u_{i+1}^n), f(u_i^n)\}, & u_{i+1}^n \leq u_i^n, \\ \min_{u_i^n \leq s \leq u_{i+1}^n} f(s), & u_i^n < u_{i+1}^n. \end{cases} \quad (36)$$

The viscosity form (2) directly yields the following relation.

$$\begin{aligned} u_{i+1}^{n+1} - u_i^{n+1} &= \frac{\lambda}{2} \left(a_{i-\frac{1}{2}}^n + q_{i-\frac{1}{2}}^n \right) (u_i^n - u_{i-1}^n) + \left(1 - \lambda a_{i+\frac{1}{2}}^n \right) (u_{i+1}^n - u_i^n) \\ &\quad + \frac{\lambda}{2} \left(a_{i+\frac{3}{2}}^n - q_{i+\frac{3}{2}}^n \right) (u_{i+2}^n - u_{i+1}^n). \end{aligned} \quad (37)$$

Let $z_{i+\frac{1}{2}}^n$, $z_{i+\frac{1}{2}, \pm}^n$ and z^n be determined by

$$z_{i+\frac{1}{2}}^n = (u_{i+1}^n - u_i^n)^+, \quad z_{i+\frac{1}{2}, -}^n = \{(u_i^n)^- - (u_{i+1}^n)^-\}^+, \quad z_{i+\frac{1}{2}, +}^n = \{(u_{i+1}^n)^+ - (u_i^n)^+\}^+ \quad (38)$$

and

$$z^n = \max_i \left\{ z_{i+\frac{1}{2}, -}^n, z_{i+\frac{1}{2}, +}^n \right\}, \quad (39)$$

where $s^+ = \max(s, 0)$ and $s^- = \max(-s, 0)$. The relation $z_{i+\frac{1}{2}}^n = z_{i+\frac{1}{2}, -}^n + z_{i+\frac{1}{2}, +}^n$ is clear. It is easy to see

1. $z_{i+\frac{1}{2}, -}^n = -u_i^n = 0 - u_i^n$, $z_{i+\frac{1}{2}, -}^n = u_{i+1}^n = u_{i+1}^n - 0$ if $u_i^n < 0 < u_{i+1}^n$,
2. $z_{i+\frac{1}{2}, -}^n = u_{i+1}^n - u_i^n$, $z_{i+\frac{1}{2}, +}^n = 0$ if $u_i^n \leq u_{i+1}^n \leq 0$,
3. $z_{i+\frac{1}{2}, -}^n = 0$, $z_{i+\frac{1}{2}, +}^n = u_{i+1}^n - u_i^n$ if $0 \leq u_i^n \leq u_{i+1}^n$,
4. $z_{i+\frac{1}{2}, -}^n = z_{i+\frac{1}{2}, +}^n = 0$ otherwise.

We first construct an inductive estimate $\{z^n\}_{n=0,1,2,3,\dots}$. For the purpose we estimate $z_{i+\frac{1}{2}, \pm}^{n+1}$ in terms of $z_{i-\frac{1}{2}, \pm}^n$, $z_{i+\frac{1}{2}, \pm}^n$ and $z_{i+\frac{3}{2}, \pm}^n$.

We note that $u_i^n < 0 < u_{i+1}^n$ holds if $u_i^{n+1} < 0 < u_{i+1}^{n+1}$ and that, even if $u_i^n < 0 < u_{i+1}^n$ is satisfied but not $u_i^{n+1} < 0 < u_{i+1}^{n+1}$, the relation

$$\max(0 - u_i^{n+1}, u_{i+1}^{n+1} - 0) \geq \max\left(z_{i+\frac{1}{2},-}^{n+1}, z_{i-\frac{1}{2},+}^{n+1}\right)$$

still holds. This fact means that we may estimate $0 - u_i^{n+1}$ and $u_{i+1}^{n+1} - 0$ instead of $z_{i+\frac{1}{2},-}^{n+1}$ and $z_{i-\frac{1}{2},+}^{n+1}$ if $u_i^n < 0 < u_{i+1}^n$. We also note that $z_{i+\frac{1}{2},\pm}^n \leq \frac{1}{\lambda c}$, is derived similarly to (26) in the proof of theorem 1 and that the quadratic function $s - \frac{\lambda c}{2}s^2$ or $s - \frac{\lambda c}{4}s^2$ of s is increasing if $s \leq \frac{1}{\lambda c}$.

We consider the following four cases.

- (C0) None of $u_i^n - u_{i-1}^n, u_{i+1}^n - u_i^n$ and $u_{i+2}^n - u_{i+1}^n$ is positive.
- (C1) Only one of $u_i^n - u_{i-1}^n, u_{i+1}^n - u_i^n$ and $u_{i+2}^n - u_{i+1}^n$ is positive.
- (C2) Just two of $u_i^n - u_{i-1}^n, u_{i+1}^n - u_i^n$ and $u_{i+2}^n - u_{i+1}^n$ are positive.
- (C3) All of $u_i^n - u_{i-1}^n, u_{i+1}^n - u_i^n$ and $u_{i+2}^n - u_{i+1}^n$ are positive, i.e. $u_{i-1}^n < u_i^n < u_{i+1}^n < u_{i+2}^n$.

The case (C0) is trivial. It is clear that $z_{i+\frac{1}{2},\pm}^{n+1} = 0$.

The case (C3) consists of the following subcases (C3-1)-(C3-3).

- (C3-1) $0 \leq u_i^n$.
- (C3-2) $u_i^n < 0 < u_{i+1}^n$.
- (C3-3) $u_{i+1}^n \leq 0$.

In (C3-1), $u_{i+1}^{n+1} > u_i^{n+1}$ comes from the relation (37). We observe

$$\bar{f}_{i-\frac{1}{2}}^n = f(\max(u_{i-1}^n, 0)), \quad \bar{f}_{i+\frac{1}{2}}^n = f(u_i^n) < \frac{1}{\lambda}u_i^n, \quad \bar{f}_{i+\frac{3}{2}}^n = f(u_{i+1}^n), \quad (40)$$

and

$$u_i^{n+1} = u_i^n - \lambda\{f(u_i^n) - f(\max(u_{i-1}^n, 0))\} \geq u_i^n - \lambda f(u_i^n) > 0. \quad (41)$$

Then we obtain $z_{i+\frac{1}{2},-}^{n+1} = 0$ and

$$\begin{aligned} z_{i+\frac{1}{2},+}^{n+1} &= u_{i+1}^{n+1} - u_i^{n+1} \\ &\leq \{1 - \lambda f'(u_i^n)\}(u_{i+1}^n - u_i^n) - \frac{\lambda c}{2}(u_{i+1}^n - u_i^n)^2 \\ &\quad + \lambda f'(u_i^n)\{u_i^n - \max(u_{i-1}^n, 0)\} - \frac{\lambda c}{2}\{u_i^n - \max(u_{i-1}^n, 0)\}^2 \\ &= \{1 - \lambda f'(u_i^n)\}z_{i+\frac{1}{2},+}^n - \frac{\lambda c}{2}\left(z_{i+\frac{1}{2},+}^n\right)^2 + \lambda f'(u_i^n)z_{i-\frac{1}{2},+}^n - \frac{\lambda c}{2}\left(z_{i-\frac{1}{2},+}^n\right)^2 \\ &\leq \max\left(z_{i-\frac{1}{2},+}^n, z_{i+\frac{1}{2},+}^n\right) - \frac{\lambda c}{2}\left\{\max\left(z_{i-\frac{1}{2},+}^n, z_{i+\frac{1}{2},+}^n\right)\right\}^2, \end{aligned} \quad (42)$$

where we use the property (21).

In (C3-2), we observe

$$\bar{f}_{i-\frac{1}{2}}^n = f(u_i^n) < \frac{1}{\lambda}(-u_i^n), \quad \bar{f}_{i+\frac{1}{2}}^n = f(0) = 0, \quad \bar{f}_{i+\frac{3}{2}}^n = f(u_{i+1}^n) < \frac{1}{\lambda}u_{i+1}^n \quad (43)$$

and

$$u_i^{n+1} = u_i^n + \lambda f(u_i^n) < 0 < u_{i+1}^n - \lambda f(u_{i+1}^n) = u_{i+1}^{n+1}. \quad (44)$$

By (21), we see $f(u_i^n) \geq \frac{1}{2}c(u_i^n)^2$ and $f(u_{i+1}^n) \geq \frac{1}{2}c(u_{i+1}^n)^2$. Then we obtain

$$z_{i+\frac{1}{2},-}^{n+1} = 0 - u_i^{n+1} = 0 - \{u_i^n + \lambda f(u_i^n)\} \leq z_{i+\frac{1}{2},-}^n - \frac{\lambda c}{2} \left(z_{i+\frac{1}{2},-}^n \right)^2 \quad (45)$$

and

$$z_{i+\frac{1}{2},+}^{n+1} = u_{i+1}^{n+1} - 0 = u_{i+1}^n - 0 - \lambda f(u_i^n) \leq z_{i+\frac{1}{2},+}^n - \frac{\lambda c}{2} \left(z_{i+\frac{1}{2},+}^n \right)^2, \quad (46)$$

which means

$$\max \left(z_{i+\frac{1}{2},\pm}^{n+1} \right) \leq \max \left(z_{i+\frac{1}{2},\pm}^n \right) - \frac{\lambda c}{2} \left(\max z_{i+\frac{1}{2},\pm}^n \right)^2. \quad (47)$$

(C3-3) is similar to (C3-1). In this subcase we obtain $z_{i+\frac{1}{2},+}^{n+1} = 0$ and

$$z_{i+\frac{1}{2},-}^{n+1} \leq \max \left(z_{i+\frac{1}{2},-}^n, z_{i+\frac{3}{2},-}^n \right) - \frac{\lambda c}{2} \left\{ \max \left(z_{i+\frac{1}{2},-}^n, z_{i+\frac{3}{2},-}^n \right) \right\}^2. \quad (48)$$

In the case (C1), $\max(z_{i+\frac{1}{2},\pm}^{n+1})$ is less than $\max(z_{i-\frac{1}{2},\pm}^n) - \frac{\lambda c}{2} \left\{ \max(z_{i-\frac{1}{2},\pm}^n) \right\}^2$, $\max(z_{i+\frac{1}{2},\pm}^n) - \frac{\lambda c}{2} \left\{ \max(z_{i+\frac{1}{2},\pm}^n) \right\}^2$ or $\max(z_{i+\frac{3}{2},\pm}^n) - \frac{\lambda c}{2} \left\{ \max(z_{i+\frac{3}{2},\pm}^n) \right\}^2$, if $u_{i-1}^n < u_i^n$, $u_i^n < u_{i+1}^n$ or $u_{i+1}^n < u_{i+2}^n$, respectively.

The case (C2) consists of the three subcases.

(C2-1) $u_{i-1}^n < u_i^n < u_{i+1}^n$

(C2-2) $u_i^n < u_{i+1}^n < u_{i+2}^n$

(C2-3) $u_{i-1}^n < u_i^n$, $u_{i+1}^n < u_{i+2}^n$

By the relation (37), we easily see that

$$u_{i+1}^{n+1} - u_i^{n+1} \leq \frac{\lambda}{2} \left(a_{i-\frac{1}{2}}^n + q_{i-\frac{1}{2}}^n \right) (u_i^n - u_{i-1}^n) + \left(1 - \lambda a_{i+\frac{1}{2}}^n \right) (u_{i+1}^n - u_i^n) \quad (49)$$

in (C2-1), and that

$$u_{i+1}^{n+1} - u_i^{n+1} \leq \left(1 - \lambda a_{i+\frac{1}{2}}^n\right) (u_{i+1}^n - u_i^n) + \frac{\lambda}{2} \left(a_{i+\frac{3}{2}}^n - q_{i+\frac{3}{2}}^n\right) (u_{i+2}^n - u_{i+1}^n) \quad (50)$$

in (C2-2). Then the discussion given in the case (C3) can be applied and it is easy to see that

$$\max(z_{i+\frac{1}{2}, \pm}^{n+1}) \leq \max(z_{i-\frac{1}{2}, \pm}^n, z_{i+\frac{1}{2}, \pm}^n, z_{i+\frac{3}{2}, \pm}^n) - \frac{\lambda c}{2} \left\{ \max(z_{i-\frac{1}{2}, \pm}^n, z_{i+\frac{1}{2}, \pm}^n, z_{i+\frac{3}{2}, \pm}^n) \right\}^2. \quad (51)$$

The subcase (C2-3) is a little more complicated. If $0 \leq u_{i+1}^n \leq u_i^n$, we observe

$$\bar{f}_{i-\frac{1}{2}}^n = f(\max(u_{i-1}^n, 0)), \quad \bar{f}_{i+\frac{1}{2}}^n = f(u_i^n), \quad \bar{f}_{i+\frac{3}{2}}^n = f(u_{i+1}^n). \quad (52)$$

Then we obtain that

$$\begin{aligned} \max(z_{i+\frac{1}{2}, \pm}^{n+1}) &\leq u_{i+1}^{n+1} - u_i^{n+1} \\ &= u_{i+1}^n - u_i^n - \lambda \{f(u_{i+1}^n) - f(u_i^n)\} + \lambda \{f(u_i^n) - f(\max(u_{i-1}^n, 0))\} \\ &= \left(1 - \lambda q_{i+\frac{1}{2}}^n\right) (u_{i+1}^n - u_i^n) + \lambda \{f(u_i^n) - f(\max(u_{i-1}^n, 0))\} \\ &\leq \lambda \{f(u_i^n) - f(\max(u_{i-1}^n, 0))\} \\ &\leq \lambda f'(u_i^n) (u_i^n - \max(u_{i-1}^n, 0)) - \frac{\lambda c}{2} (u_i^n - \max(u_{i-1}^n, 0))^2 \\ &\leq z_{i-\frac{1}{2}, +}^n - \frac{\lambda c}{2} (z_{i-\frac{1}{2}, +}^n)^2. \end{aligned} \quad (53)$$

If $u_{i+1}^n \leq u_i^n \leq 0$, we have

$$\max(z_{i+\frac{1}{2}, \pm}^{n+1}) \leq z_{i+\frac{3}{2}, -}^n - \frac{\lambda c}{2} (z_{i+\frac{3}{2}, -}^n)^2. \quad (54)$$

by similar discussion.

If $u_{i+1}^n < 0 < u_i^n$, we observe

$$\begin{aligned} \bar{f}_{i-\frac{1}{2}}^n &= f(\max(u_{i-1}^n, 0)), \\ \bar{f}_{i+\frac{1}{2}}^n &= \max\{f(u_i^n), f(u_{i+1}^n)\} = \frac{1}{2} \{f(u_i^n) + f(u_{i+1}^n)\} + \frac{1}{2} |f(u_i^n) - f(u_{i+1}^n)|, \\ \bar{f}_{i+\frac{3}{2}}^n &= f(\min(u_{i+2}^n, 0)) \end{aligned}$$

and

$$\begin{aligned} u_{i+1}^{n+1} - u_i^{n+1} &= -(u_i^n - u_{i+1}^n) + \lambda \{f(u_i^n) - f(\max(u_{i-1}^n, 0))\} \\ &\quad + \lambda \{f(u_{i+1}^n) - f(\min(u_{i+2}^n, 0))\} + \lambda |f(u_{i+1}^n) - f(u_i^n)|. \end{aligned} \quad (55)$$

Here it is easy to see that $|u_i^n| = u_i^n$ or $|u_{i+1}^n| = u_{i+1}^n$ is larger than $\lambda |f(u_{i+1}^n) - f(u_i^n)|$. Then we obtain

$$\begin{aligned} u_{i+1}^{n+1} - u_i^{n+1} &\leq -u_i^n + \lambda \{f(u_i^n) - f(\max(u_{i-1}^n, 0))\} + \lambda \{f(u_{i+1}^n) - f(\min(u_{i+2}^n, 0))\} \\ &\leq -u_i^n + \lambda f(u_i^n) + \lambda \{f(u_{i+1}^n) - f(\min(u_{i+2}^n, 0))\} \\ &\leq \lambda \{f(u_{i+1}^n) - f(\min(u_{i+2}^n, 0))\} \end{aligned} \quad (56)$$

or

$$\begin{aligned} u_{i+1}^{n+1} - u_i^{n+1} &\leq u_{i+1}^n + \lambda \{f(u_{i+1}^n) - f(\min(u_{i+2}^n, 0))\} + \lambda \{f(u_i^n) - f(\max(u_{i-1}^n, 0))\} \\ &\leq u_{i+1}^n + \lambda f(u_{i+1}^n) + \lambda \{f(u_i^n) - f(\max(u_{i-1}^n, 0))\} \\ &\leq \lambda \{f(u_i^n) - f(\max(u_{i-1}^n, 0))\}, \end{aligned} \quad (57)$$

which implies

$$\max \left(z_{i+\frac{1}{2}, \pm}^{n+1} \right) \leq \max \left(z_{i-\frac{1}{2}, +}^n, z_{i+\frac{3}{2}, -}^n \right) - \frac{\lambda c}{2} \left\{ \max \left(z_{i-\frac{1}{2}, +}^n, z_{i+\frac{3}{2}, -}^n \right) \right\}^2. \quad (58)$$

From the discussion above for the cases (C0)-(C3), we finally obtain the inductive estimate

$$z^{n+1} \leq z^n - \frac{\lambda c}{2} (z^n)^2 \quad (59)$$

for the sequence $\{z^n\}$. The inductive estimate gives

$$z^n \leq \frac{2\Delta x}{n\Delta t \cdot c} \quad (60)$$

by similar discussion to (27)-(31) in the proof of theorem 1. The relation

$$u_{i+1}^n - u_i^n \leq z_{i+\frac{1}{2}}^n = z_{i+\frac{1}{2}, -}^n + z_{i+\frac{1}{2}, +}^n \leq 2z^n \quad (61)$$

yields the estimate (33) with (34). If the sonic point 0 is not between u_i^n and u_{i+1}^n , we have

$$u_{i+1}^n - u_i^n \leq z_{i+\frac{1}{2}}^n = \max z_{i+\frac{1}{2}, \pm}^n \leq z^n \quad (62)$$

and conclude $E = 2E_0$ instead of (34).

We complete the proof of theorem.

In the sense that they are based on Riemann problem occurring at each contact of neighboring finite volumes, Lax-Friedrichs scheme and Godunov scheme are from a similar idea. But the proof of the consistency with E-condition is much easier in the case of Lax-Friedrichs scheme because the scheme avoids explicit switching procedure employing the staggered mesh system. While the upwindness is a natural idea to construct difference schemes for conservation laws, upwind difference schemes usually consists of some explicit switching procedure which makes the discussion on the consistency with E-condition complicated. But it is still possible to obtain a result on the consistency with E-condition for some class of upwind difference schemes. It is shown in the following section.

4. TVD Schemes with Entropy Fix

We suppose a difference scheme (2) with each coefficient $a_{i+\frac{1}{2}}^n$ given by

$$a_{i+\frac{1}{2}}^n = \begin{cases} \left| q_{i+\frac{1}{2}}^n \right| & \text{if } u_i^n \geq u_{i+1}^n, \\ \frac{1}{\lambda} Q_\epsilon(\lambda q_{i+\frac{1}{2}}^n) & \text{if } u_i^n < u_{i+1}^n, \end{cases} \quad (63)$$

where

$$Q_\epsilon(s) = \max\{|s|, \epsilon\}. \quad (64)$$

THEOREM 3 Let ϵ be fixed so that $0 < \epsilon \leq \frac{1}{2}$. Then the difference scheme determined by (2), (63) and (64) satisfies the the numerical E-condition (33);

$$\frac{u_{i+1}^n - u_i^n}{\Delta x} < \frac{E}{n\Delta t},$$

with some constant E large enough.

Furthermore, for an arbitrary positive constant η , there exists a positive integer $N = N(\epsilon, \eta)$ so that the numerical E-condition (15) is valid with the constant $E = (4 + \eta)E_0$ if $n \geq N$. N does not depend on Δx nor Δt . When $\sup_i \{u_{i+1}^0 - u_i^0\}$ is small enough and ϵ is large enough, it is possible even to take $\eta = 0$ in this statement.

COROLLARY 4 Let $Q = Q(s)$, $-1 \leq s \leq 1$ be a function satisfying the following condition.

- 1) Q is an even function, i.e. $Q(-s) = Q(s)$.
- 2) Q is a convex function.
- 3) $Q(s) = |s|$ if $|s| \geq \frac{1}{2}$.
- 4) $Q(0) > 0$

(Note that $Q(s)$ takes the minimum at $s = 0$ because of 1) and 2).)

Then the difference scheme (2) with each numerical viscosity coefficient $a_{i+\frac{1}{2}}^n$ given by

$$a_{i+\frac{1}{2}}^n = \begin{cases} \left| q_{i+\frac{1}{2}}^n \right| & \text{if } u_i^n \geq u_{i+1}^n, \\ \frac{1}{\lambda} Q(\lambda q_{i+\frac{1}{2}}^n) & \text{if } u_i^n < u_{i+1}^n \end{cases} \quad (65)$$

the numerical E-condition (33);

$$\frac{u_{i+1}^n - u_i^n}{\Delta x} < \frac{E}{n\Delta t},$$

with some constant E large enough.

Furthermore, for an arbitrary positive constant η , there exists a positive integer $N = N(\epsilon, \eta)$ so that the numerical E-condition (15) is valid with the constant $E = (4 + \eta)E_0$ if $n \geq N$. N does not depend on Δx nor Δt . When $\sup_i u_{i+1}^0 - u_i^0$ is small enough, it is possible even to take $\eta = 0$ for some appropriate Q .

The corollary derives from the theorem and the fact that the function Q can be written in the form of convex combination of the family $\{Q_\epsilon\}_{0 < \epsilon \leq \frac{1}{2}}$ of functions Q_ϵ given by (64).

We sketch the proof of theorem.

Let us introduce the following notation.

$$z_{i+\frac{1}{2}}^n = (u_{i+1}^n - u_i^n)^+ = \max(u_{i+1}^n - u_i^n, 0), z^n = \sup_i \left\{ z_{i+\frac{1}{2}}^n \right\}. \quad (66)$$

We first show the following lemma.

LEMMA 5 *The sequence $\{z^n\}_{n=0,1,2,\dots}$ satisfies*

$$z^{n+1} \leq \max \left\{ z^n - \frac{\lambda c}{4} (z^n)^2, (1 - \alpha \epsilon) z^n \right\}, \quad (67)$$

where α is a positive constant; $\alpha = \frac{3-\sqrt{5}}{2}$.

We estimate $z_{i+\frac{1}{2}}^{n+1}$ in terms of $z_{i-\frac{1}{2}}^n$, $z_{i+\frac{1}{2}}^n$ and $z_{i+\frac{3}{2}}^n$ to prove the lemma.

As in the proof of theorem 2, we have the following four cases.

- (C0) None of $u_i^n - u_{i-1}^n$, $u_{i+1}^n - u_i^n$ and $u_{i+2}^n - u_{i+1}^n$ is positive.
- (C1) Only one of $u_i^n - u_{i-1}^n$, $u_{i+1}^n - u_i^n$ and $u_{i+2}^n - u_{i+1}^n$ is positive.
- (C2) Just two of $u_i^n - u_{i-1}^n$, $u_{i+1}^n - u_i^n$ and $u_{i+2}^n - u_{i+1}^n$ are positive.
- (C3) All of $u_i^n - u_{i-1}^n$, $u_{i+1}^n - u_i^n$ and $u_{i+2}^n - u_{i+1}^n$ are positive,
i.e. $u_{i-1}^n < u_i^n < u_{i+1}^n < u_{i+2}^n$.

The case (C0) is obvious.

The case (C3) consists of the ten subcases

$$(C3-1) q_{i-\frac{1}{2}}^n < q_{i+\frac{1}{2}}^n < q_{i+\frac{3}{2}}^n \leq -\frac{\epsilon}{\lambda}, \quad (C3-2) q_{i-\frac{1}{2}}^n < q_{i+\frac{1}{2}}^n \leq -\frac{\epsilon}{\lambda} < q_{i+\frac{3}{2}}^n < \frac{\epsilon}{\lambda},$$

$$(C3-3) q_{i-\frac{1}{2}}^n < q_{i+\frac{1}{2}}^n \leq -\frac{\epsilon}{\lambda}, \frac{\epsilon}{\lambda} \leq q_{i+\frac{3}{2}}^n, \quad (C3-4) q_{i-\frac{1}{2}}^n \leq -\frac{\epsilon}{\lambda} < q_{i+\frac{1}{2}}^n < q_{i+\frac{3}{2}}^n < \frac{\epsilon}{\lambda},$$

$$(C3-5) q_{i-\frac{1}{2}}^n \leq -\frac{\epsilon}{\lambda} < q_{i+\frac{1}{2}}^n < \frac{\epsilon}{\lambda} \leq q_{i+\frac{3}{2}}^n, \quad (C3-6) q_{i-\frac{1}{2}}^n \leq -\frac{\epsilon}{\lambda}, \frac{\epsilon}{\lambda} \leq q_{i+\frac{1}{2}}^n < q_{i+\frac{3}{2}}^n,$$

$$(C3-7) -\frac{\epsilon}{\lambda} < q_{i-\frac{1}{2}}^n < q_{i+\frac{1}{2}}^n < q_{i+\frac{3}{2}}^n < \frac{\epsilon}{\lambda}, \quad (C3-8) -\frac{\epsilon}{\lambda} < q_{i-\frac{1}{2}}^n < q_{i+\frac{1}{2}}^n < \frac{\epsilon}{\lambda} \leq q_{i+\frac{3}{2}}^n,$$

$$(C3-9) -\frac{\epsilon}{\lambda} < q_{i-\frac{1}{2}}^n < \frac{\epsilon}{\lambda} \leq q_{i+\frac{1}{2}}^n < q_{i+\frac{3}{2}}^n, \quad (C3-10) \frac{\epsilon}{\lambda} \leq q_{i-\frac{1}{2}}^n < q_{i+\frac{1}{2}}^n < q_{i+\frac{3}{2}}^n,$$

where we note the relation $q_{i-\frac{1}{2}}^n < q_{i+\frac{1}{2}}^n < q_{i+\frac{3}{2}}^n$.

If $\left| q_{i-\frac{1}{2}}^n \right|, \left| q_{i+\frac{1}{2}}^n \right|, \left| q_{i+\frac{3}{2}}^n \right| \geq \frac{\epsilon}{\lambda}$, i.e. in the subcases (C3-1), (C3-3), (C3-6) and (C3-10), we see that

$$\begin{aligned} z_{i+\frac{1}{2}}^{n+1} &= u_{i+1}^{n+1} - u_i^{n+1} \leq \max\left(z_{i-\frac{1}{2}}^n, z_{i+\frac{1}{2}}^n, z_{i+\frac{3}{2}}^n\right) - \frac{\lambda c}{2} \left\{ \max\left(z_{i-\frac{1}{2}}^n, z_{i+\frac{1}{2}}^n, z_{i+\frac{3}{2}}^n\right) \right\}^2 \\ &< \max\left(z_{i-\frac{1}{2}}^n, z_{i+\frac{1}{2}}^n, z_{i+\frac{3}{2}}^n\right) - \frac{\lambda c}{4} \left\{ \max\left(z_{i-\frac{1}{2}}^n, z_{i+\frac{1}{2}}^n, z_{i+\frac{3}{2}}^n\right) \right\}^2 \end{aligned} \quad (68)$$

by similar discussion to the case (C3) in the proof of theorem 2.

In the subcase (C3-5),

$$z_{i+\frac{1}{2}}^{n+1} = u_{i+1}^{n+1} - u_i^{n+1} \leq (1 - \epsilon)(u_{i+1}^n - u_i^n) < (1 - \alpha\epsilon)z_{i+\frac{1}{2}}^n \quad (69)$$

is easily derived.

In the subcase (C3-2), $z_{i+\frac{1}{2}}^{n+1} = u_{i+1}^{n+1} - u_i^{n+1}$ is estimated as follows.

$$\begin{aligned} z_{i+\frac{1}{2}}^{n+1} &= \left(1 + \lambda q_{i+\frac{1}{2}}^n\right)(u_{i+1}^n - u_i^n) + \left(\frac{\epsilon}{2} - \frac{\lambda}{2} q_{i+\frac{3}{2}}^n\right)(u_{i+2}^n - u_{i+1}^n) \\ &\leq \left(1 - \frac{\epsilon}{2} + \frac{\lambda}{2} q_{i+\frac{1}{2}}^n\right)(u_{i+1}^n - u_i^n) + \left(\frac{\epsilon}{2} - \frac{\lambda}{2} q_{i+\frac{3}{2}}^n\right)(u_{i+2}^n - u_{i+1}^n) \\ &\leq \left(1 - \frac{\epsilon}{2} + \frac{\lambda}{2} f'(u_{i+1}^n)\right)(u_{i+1}^n - u_i^n) - \frac{\lambda c}{4}(u_{i+1}^n - u_i^n)^2 \\ &\quad + \left(\frac{\epsilon}{2} - \frac{\lambda}{2} f'(u_{i+1}^n)\right)(u_{i+2}^n - u_{i+1}^n) - \frac{\lambda c}{4}(u_{i+2}^n - u_{i+1}^n)^2 \\ &\leq \max\left(z_{i+\frac{1}{2}}^n, z_{i+\frac{3}{2}}^n\right) - \frac{\lambda c}{4} \left\{ \max\left(z_{i+\frac{1}{2}}^n, z_{i+\frac{3}{2}}^n\right) \right\}^2 \end{aligned} \quad (70)$$

A similar estimate

$$z_{i+\frac{1}{2}}^{n+1} \leq \max \left(z_{i-\frac{1}{2}}^n, z_{i+\frac{1}{2}}^n \right) - \frac{\lambda c}{4} \left\{ \max \left(z_{i-\frac{1}{2}}^n, z_{i+\frac{1}{2}}^n \right) \right\}^2 \quad (71)$$

is obtained in (C3-9).

The remaining subcases (C3-4), (C3-7) and (C3-8) are a little more complicated. In (C3-7), we first observe

$$\begin{aligned} z_{i+\frac{1}{2}}^{n+1} &= u_{i+1}^n - u_i^{n+1} \\ &= \left(\frac{\epsilon}{2} + \frac{\lambda}{2} q_{i-\frac{1}{2}}^n \right) (u_i^n - u_{i-1}^n) + (1-\epsilon)(u_{i+1}^n - u_i^n) + \left(\frac{\epsilon}{2} - \frac{\lambda}{2} q_{i+\frac{3}{2}}^n \right) (u_{i+2}^n - u_{i+1}^n) \\ &\leq \left(\frac{\epsilon}{2} + \frac{\lambda}{2} f'(u_i^n) \right) (u_i^n - u_{i-1}^n) - \frac{\lambda c}{4} (u_i^n - u_{i-1}^n)^2 + (1-\epsilon)(u_{i+1}^n - u_i^n) \\ &\quad + \left(\frac{\epsilon}{2} - \frac{\lambda}{2} f'(u_{i+1}^n) \right) (u_{i+2}^n - u_{i+1}^n) - \frac{\lambda c}{4} (u_{i+2}^n - u_{i+1}^n)^2. \end{aligned} \quad (72)$$

Then, by the fact that

$$f'(u_i^n) \leq f' \left(\frac{u_i^n + u_{i+1}^n}{2} \right) - \frac{c}{2} (u_{i+1}^n - u_i^n), \quad f'(u_{i+1}^n) \geq f' \left(\frac{u_i^n + u_{i+1}^n}{2} \right) + \frac{c}{2} (u_{i+1}^n - u_i^n),$$

and

$$\left| f' \left(\frac{u_i^n + u_{i+1}^n}{2} \right) \right| < \frac{\epsilon}{\lambda},$$

we obtain

$$\begin{aligned} z_{i+\frac{1}{2}}^{n+1} &\leq \left(\frac{\epsilon}{2} + \frac{\lambda}{2} f' \left(\frac{u_i^n + u_{i+1}^n}{2} \right) \right) (u_i^n - u_{i-1}^n) + (1-\epsilon)(u_{i+1}^n - u_i^n) \\ &\quad + \left(\frac{\epsilon}{2} - \frac{\lambda}{2} f' \left(\frac{u_i^n + u_{i+1}^n}{2} \right) \right) (u_{i+2}^n - u_{i+1}^n) \\ &\quad - \frac{\lambda c}{4} (u_i^n - u_{i-1}^n)^2 - \frac{\lambda c}{4} (u_{i+1}^n - u_i^n) (u_i^n - u_{i-1}^n) \\ &\quad - \frac{\lambda c}{4} (u_{i+2}^n - u_{i+1}^n)^2 - \frac{\lambda c}{4} (u_{i+1}^n - u_i^n) (u_{i+2}^n - u_{i+1}^n) \\ &\leq (1-\epsilon) z_{i+\frac{1}{2}}^n + \epsilon \cdot \max \left(z_{i-\frac{1}{2}}^n, z_{i+\frac{3}{2}}^n \right) \\ &\quad - \frac{\lambda c}{4} \left\{ \max \left(z_{i-\frac{1}{2}}^n, z_{i+\frac{3}{2}}^n \right) \right\}^2 - \frac{\lambda c}{4} z_{i+\frac{1}{2}}^n \cdot \max \left(z_{i-\frac{1}{2}}^n, z_{i+\frac{3}{2}}^n \right). \end{aligned} \quad (73)$$

It implies that

$$z_{i+\frac{1}{2}}^{n+1} \leq \begin{cases} \max \left(z_{i-\frac{1}{2}}^n, z_{i+\frac{1}{2}}^n, z_{i+\frac{3}{2}}^n \right) - \frac{\lambda c}{4} \left\{ \max \left(z_{i-\frac{1}{2}}^n, z_{i+\frac{1}{2}}^n, z_{i+\frac{3}{2}}^n \right) \right\}^2 & \text{if } \max \left(z_{i-\frac{1}{2}}^n, z_{i+\frac{3}{2}}^n \right) \geq \frac{\sqrt{5}-1}{2} z_{i+\frac{1}{2}}^n, \\ \left(1 - \frac{3-\sqrt{5}}{2} \epsilon \right) z_{i+\frac{1}{2}}^n & \text{otherwise.} \end{cases} \quad (74)$$

In the subcase (C3-4), observing

$$z_{i+\frac{1}{2}}^{n+1} = u_{i+1}^{n+1} - u_i^{n+1} = (1 - \epsilon) z_{i+\frac{1}{2}}^n + \left(\frac{\epsilon}{2} - \frac{\lambda}{2} q_{i+\frac{3}{2}}^n \right) z_{i+\frac{3}{2}}^n \quad (75)$$

and

$$\begin{aligned} q_{i+\frac{1}{2}}^n &= \frac{f(u_{i+2}^n) - f(u_{i+1}^n)}{u_{i+2}^n - u_{i+1}^n} \geq \frac{f'(u_{i+1}^n)(u_{i+2}^n - u_{i+1}^n) + \frac{c}{2}(u_{i+2}^n - u_{i+1}^n)^2}{u_{i+2}^n - u_{i+1}^n} \\ &= f'(u_{i+1}^n) + \frac{c}{2}(u_{i+2}^n - u_{i+1}^n) \geq \frac{\epsilon}{\lambda} + \frac{c}{2} z_{i+\frac{3}{2}}^n, \end{aligned} \quad (76)$$

we achieve the following estimate.

$$\begin{aligned} z_{i+\frac{1}{2}}^{n+1} &\leq (1 - \epsilon) z_{i+\frac{1}{2}}^n + \epsilon \cdot z_{i+\frac{3}{2}}^n + \frac{\lambda c}{2} \left(z_{i+\frac{3}{2}}^n \right)^2 \\ &\leq \begin{cases} \max(z_{i+\frac{1}{2}}^n, z_{i+\frac{3}{2}}^n) - \frac{\lambda c}{4} \left\{ \max(z_{i+\frac{1}{2}}^n, z_{i+\frac{3}{2}}^n) \right\}^2 & \text{if } z_{i+\frac{3}{2}}^n \geq \frac{\sqrt{5}-1}{2} z_{i+\frac{1}{2}}^n, \\ \left(1 - \frac{3-\sqrt{5}}{2} \epsilon \right) z_{i+\frac{1}{2}}^n & \text{otherwise.} \end{cases} \end{aligned} \quad (77)$$

Similar discussion is applied to (C3-8) and we obtain

$$z_{i+\frac{1}{2}}^{n+1} \leq \begin{cases} \max(z_{i-\frac{1}{2}}^n, z_{i+\frac{1}{2}}^n) - \frac{\lambda c}{4} \left\{ \max(z_{i-\frac{1}{2}}^n, z_{i+\frac{1}{2}}^n) \right\}^2 & \text{if } z_{i-\frac{1}{2}}^n \geq \frac{\sqrt{5}-1}{2} z_{i+\frac{1}{2}}^n, \\ \left(1 - \frac{3-\sqrt{5}}{2} \epsilon \right) z_{i+\frac{1}{2}}^n & \text{otherwise.} \end{cases} \quad (78)$$

In (C1) $u_{i+1}^n - u_i^n$ is bounded by either of $\frac{\lambda}{2} \left(a_{i-\frac{1}{2}}^n + q_{i-\frac{1}{2}}^n \right) (u_i^n - u_{i-1}^n)$, $\left(1 - \lambda a_{i+\frac{1}{2}}^n \right) (u_{i+1}^n - u_i^n)$ or $\frac{\lambda}{2} \left(a_{i+\frac{3}{2}}^n + q_{i+\frac{3}{2}}^n \right) (u_{i+2}^n - u_{i+1}^n)$. Then we easily observe that $z_{i+\frac{1}{2}}^{n+1}$ is bounded by either of $z_{i-\frac{1}{2}}^n - \frac{\lambda c}{2} \left(z_{i-\frac{1}{2}}^n \right)^2$, $z_{i+\frac{3}{2}}^n - \frac{\lambda c}{2} \left(z_{i+\frac{3}{2}}^n \right)^2$ or $(1 - \epsilon) z_{i+\frac{1}{2}}^n$.

The remaining case (C2) consists of the following subcases.

(C2-1) $u_{i-1}^n < u_i^n < u_{i+1}^n$.

(C2-2) $u_i^n < u_{i+1}^n < u_{i+2}^n$.

(C2-3) $u_{i-1}^n < u_i^n, u_{i+1}^n < u_{i+2}^n$.

All the machinery used to discuss the subcases (C2-1) and (C2-2) already appears in the case (C3). We conclude

$$z_{i+\frac{1}{2}}^{n+1} \leq \max(z_{i-\frac{1}{2}}^n, z_{i+\frac{1}{2}}^n, z_{i+\frac{3}{2}}^n) - \frac{\lambda c}{4} \left\{ \max(z_{i-\frac{1}{2}}^n, z_{i+\frac{1}{2}}^n, z_{i+\frac{3}{2}}^n) \right\}^2 \text{ or } \left(1 - \frac{3-\sqrt{5}}{2} \epsilon \right) z_{i+\frac{1}{2}}^n \quad (79)$$

with a little less difficulty than (C3).

In the subcase (C2-3), we observe the followings.

- If $q_{i-\frac{1}{2}}^n \leq 0 \leq q_{i+\frac{3}{2}}^n$,

$$u_{i+1}^{n+1} - u_i^{n+1} \leq \frac{\epsilon}{2}(u_i^n - u_{i-1}^n) + \frac{\epsilon}{2}(u_{i+2}^n - u_{i+1}^n) \leq (1 - \epsilon) \max \left(z_{i-\frac{1}{2}}^n, z_{i+\frac{3}{2}}^n \right)$$

- If $q_{i-\frac{1}{2}}^n \leq -\frac{\epsilon}{\lambda}$, it is clear that $a_{i+\frac{1}{2}}^n + q_{i+\frac{1}{2}}^n = 0$. Then we observe

$$u_{i+1}^{n+1} - u_i^{n+1} \leq \frac{\lambda}{2} \left(a_{i+\frac{3}{2}}^n - q_{i+\frac{3}{2}}^n \right) (u_{i+2}^n - u_{i+1}^n) \leq z_{i+\frac{3}{2}}^n - \frac{\lambda c}{4} \left(z_{i+\frac{3}{2}}^n \right)^2.$$

- If $q_{i+\frac{1}{2}}^n \geq \frac{\epsilon}{\lambda}$, similar discussion to that above gives

$$u_{i+1}^{n+1} - u_i^{n+1} \leq z_{i-\frac{1}{2}}^n - \frac{\lambda c}{4} \left(z_{i-\frac{1}{2}}^n \right)^2.$$

- If $q_{i+\frac{1}{2}}^n \leq 0 \leq q_{i-\frac{1}{2}}^n$, we easily observe $u_i^n > 0$ and $u_{i+1}^n < \bar{u} < u_{i-1}^n$, where \bar{u} is defined so that $\bar{u} < 0$ and $f(\bar{u}) = f(u_i^n)$. We also observe

$$|f(u_{i+1}^n) - f(\bar{u})| < \frac{1}{\lambda} (\bar{u} - u_{i+1}^n)$$

and

$$\bar{u} - u_{i+1}^n > \lambda |f(u_{i+1}^n) - f(\bar{u})| = \lambda |f(u_{i+1}^n) - f(u_i^n)| = -\lambda q_{i+\frac{1}{2}}^n (u_{i+1}^n - u_i^n).$$

Then we conclude

$$\begin{aligned} & \frac{\lambda}{2} \left(a_{i-\frac{1}{2}}^n + q_{i-\frac{1}{2}}^n \right) (u_i^n - u_{i-1}^n) \leq u_i^n - u_{i-1}^n \\ & < u_i^n - \bar{u} = (u_i^n - u_{i+1}^n) + (u_{i+1}^n - \bar{u}) < - \left(1 - \lambda a_{i+\frac{1}{2}}^n \right) (u_{i+1}^n - u_i^n), \end{aligned}$$

and

$$u_{i+1}^{n+1} - u_i^{n+1} \leq \frac{\lambda}{2} \left(a_{i+\frac{3}{2}}^n - q_{i+\frac{3}{2}}^n \right) (u_{i+2}^n - u_{i+1}^n) \leq z_{i+\frac{3}{2}}^n - \frac{\lambda c}{4} \left(z_{i+\frac{3}{2}}^n \right)^2.$$

- If $q_{i+\frac{3}{2}}^n \leq 0 \leq q_{i+\frac{1}{2}}^n$, we obtain

$$\frac{\lambda}{2} \left(a_{i+\frac{3}{2}}^n - q_{i+\frac{3}{2}}^n \right) (u_{i+2}^n - u_{i+1}^n) \leq - \left(1 - \lambda a_{i+\frac{1}{2}}^n \right) (u_{i+1}^n - u_i^n)$$

and

$$u_{i+1}^{n+1} - u_i^{n+1} \leq z_{i-\frac{1}{2}}^n - \frac{\lambda c}{4} \left(z_{i-\frac{1}{2}}^n \right)^2$$

as well.

After the observation above, we still have the following subcases that remains to be considered.

$$(C2-3-1) \quad -\frac{\epsilon}{\lambda} < q_{i-\frac{1}{2}}^n < 0 \text{ and } q_{i+\frac{1}{2}}^n, q_{i+\frac{3}{2}}^n < 0,$$

$$(C2-3-2) \quad 0 < q_{i+\frac{3}{2}}^n < \frac{\epsilon}{\lambda} \text{ and } 0 < q_{i-\frac{1}{2}}^n, q_{i+\frac{1}{2}}^n.$$

In the subcase (C2-3-1), we observe

$$\begin{aligned} \left(1 - \lambda a_{i+\frac{1}{2}}^n\right) (u_{i+1}^n - u_i^n) &= \left(1 + \lambda q_{i+\frac{1}{2}}^n\right) (u_{i+1}^n - u_i^n) \\ &\leq \frac{1}{2} \left(1 + \lambda q_{i+\frac{1}{2}}^n\right) (u_{i+1}^n - u_i^n) \leq \frac{\epsilon}{2} (u_{i+1}^n - u_i^n) + \frac{\lambda}{2} \{f(u_{i+1}^n) - f(u_i^n)\}, \end{aligned} \quad (80)$$

$$\begin{aligned} \left(1 - \lambda a_{i+\frac{1}{2}}^n\right) (u_{i+1}^n - u_i^n) &= u_{i+1}^n - u_i^n + \lambda \{f(u_{i+1}^n) - f(u_i^n)\} \\ &= (u_{i+1}^n - u_{i-1}^n) + (u_{i-1}^n - u_i^n) + \lambda \{f(u_{i+1}^n) - f(u_{i-1}^n)\} + \lambda \{f(u_{i-1}^n) - f(u_i^n)\} \\ &= \left(1 + \lambda \frac{f(u_{i+1}^n) - f(u_{i-1}^n)}{u_{i+1}^n - u_{i-1}^n}\right) (u_{i+1}^n - u_{i-1}^n) + \left(1 + \lambda q_{i-\frac{1}{2}}^n\right) (u_{i-1}^n - u_i^n) \end{aligned} \quad (81)$$

and

$$\begin{aligned} \frac{\lambda}{2} \left(a_{i-\frac{1}{2}}^n + q_{i-\frac{1}{2}}^n\right) (u_i^n - u_{i-1}^n) &= \left(\frac{\epsilon}{2} + \frac{\lambda}{2} q_{i-\frac{1}{2}}^n\right) (u_i^n - u_{i-1}^n) \\ &= \frac{\epsilon}{2} (u_i^n - u_{i-1}^n) + \frac{\lambda}{2} \{f(u_i^n) - f(u_{i-1}^n)\}. \end{aligned} \quad (82)$$

We obtain the following estimate.

- If $u_{i-1}^n \leq u_{i+1}^n$,

$$\begin{aligned} u_{i+1}^{n+1} - u_i^{n+1} &= \frac{\lambda}{2} \left(a_{i-\frac{1}{2}}^n + q_{i-\frac{1}{2}}^n\right) (u_i^n - u_{i-1}^n) + \left(1 - \lambda a_{i+\frac{1}{2}}^n\right) (u_{i+1}^n - u_i^n) \\ &\quad + \frac{\lambda}{2} \left(a_{i+\frac{3}{2}}^n - q_{i+\frac{3}{2}}^n\right) (u_{i+2}^n - u_{i+1}^n) \\ &\leq \frac{\epsilon}{2} (u_i^n - u_{i-1}^n) + \frac{\lambda}{2} \{f(u_i^n) - f(u_{i-1}^n)\} \\ &\quad + \frac{\epsilon}{2} (u_{i+1}^n - u_i^n) + \frac{\lambda}{2} \{f(u_{i+1}^n) - f(u_i^n)\} + \frac{\lambda}{2} \left(a_{i+\frac{3}{2}}^n - q_{i+\frac{3}{2}}^n\right) (u_{i+2}^n - u_{i+1}^n) \\ &\leq \frac{\epsilon}{2} (u_{i+1}^n - u_{i-1}^n) + \frac{\lambda}{2} \{f(u_{i+1}^n) - f(u_{i-1}^n)\} + \frac{\lambda}{2} \left(a_{i+\frac{3}{2}}^n - q_{i+\frac{3}{2}}^n\right) (u_{i+2}^n - u_{i+1}^n) \\ &\leq \frac{1}{2} \{1 + \lambda f'(u_{i+1}^n)\} (u_{i+1}^n - u_{i-1}^n) - \frac{\lambda c}{4} (u_{i+1}^n - u_{i-1}^n)^2 \\ &\quad + \frac{\lambda}{2} \left\{a_{i+\frac{3}{2}}^n - f'(u_{i+1}^n)\right\} (u_{i+2}^n - u_{i+1}^n) - \frac{\lambda c}{4} (u_{i+2}^n - u_{i+1}^n)^2 \end{aligned}$$

$$\begin{aligned}
&\leq \max(u_{i+1}^n - u_{i-1}^n, u_{i+2}^n - u_{i+1}^n) - \frac{\lambda c}{4} \{ \max(u_{i+1}^n - u_{i-1}^n, u_{i+2}^n - u_{i+1}^n) \}^2 \\
&\leq \max\left(z_{i-\frac{1}{2}}^n, z_{i+\frac{3}{2}}^n\right) - \frac{\lambda c}{4} \left\{ \max\left(z_{i-\frac{1}{2}}^n, z_{i+\frac{3}{2}}^n\right) \right\}^2
\end{aligned} \tag{83}$$

with the use of (80) and (82).

- If $u_{i+1}^n < u_{i-1}^n$,

$$\begin{aligned}
&u_{i+1}^{n+1} - u_i^{n+1} \\
&= \left(1 + \lambda \frac{f(u_{i+1}^n) - f(u_{i-1}^n)}{u_{i+1}^n - u_{i-1}^n}\right) (u_{i+1}^n - u_{i-1}^n) + \left(1 + \lambda q_{i-\frac{1}{2}}^n\right) (u_{i-1}^n - u_i^n) \\
&\quad + \left(\frac{\epsilon}{2} + \frac{\lambda}{2} q_{i-\frac{1}{2}}^n\right) (u_i^n - u_{i-1}^n) + \frac{\lambda}{2} \left(a_{i+\frac{3}{2}}^n - q_{i+\frac{3}{2}}^n\right) (u_{i+2}^n - u_{i+1}^n) \\
&= \left(1 + \lambda \frac{f(u_{i+1}^n) - f(u_{i-1}^n)}{u_{i+1}^n - u_{i-1}^n}\right) (u_{i+1}^n - u_{i-1}^n) - \left(1 - \frac{\epsilon}{2} + \frac{\lambda}{2} q_{i-\frac{1}{2}}^n\right) (u_i^n - u_{i-1}^n) \\
&\quad + \frac{\lambda}{2} \left(a_{i+\frac{3}{2}}^n - q_{i+\frac{3}{2}}^n\right) (u_{i+2}^n - u_{i+1}^n) \\
&\leq \frac{\lambda}{2} \left(a_{i+\frac{3}{2}}^n - q_{i+\frac{3}{2}}^n\right) (u_{i+2}^n - u_{i+1}^n) \leq z_{i+\frac{3}{2}}^n - \frac{\lambda c}{4} \left(z_{i+\frac{3}{2}}^n\right)^2
\end{aligned} \tag{84}$$

with the use of (81) and (82).

Then we conclude

$$z_{i+\frac{1}{2}}^{n+1} \leq \max\left(z_{i-\frac{1}{2}}^n, z_{i+\frac{3}{2}}^n\right) - \frac{\lambda c}{4} \left\{ \max\left(z_{i-\frac{1}{2}}^n, z_{i+\frac{3}{2}}^n\right) \right\}^2. \tag{85}$$

In (C2-3-2), A similar discussion gives the same estimate as (85).

Finally we conclude

$$z^{n+1} \leq \max\left\{z^n - \frac{\lambda c}{4}(z^n)^2, (1 - \alpha\epsilon)z^n\right\}, \tag{86}$$

and complete the proof of lemma.

Let us return to the proof of theorem.

Define $z = z(\tau)$, $\tau \geq 0$ by

$$\begin{cases} z'(\tau) = \max\left\{-\frac{\lambda}{4}\{z(\tau)\}^2, -\alpha\epsilon z(\tau)\right\} \\ z(0) = z^0. \end{cases} \tag{87}$$

It is easy to see that

$$z^n < z(n), n \geq 1. \tag{88}$$

The function $z = z(\tau)$ is explicitly described as follows.

$$\text{When } 0 < z^0 \leq \frac{4\alpha\epsilon}{\lambda c},$$

$$z(\tau) = \frac{\frac{4}{\lambda c}}{\tau + \frac{4}{\lambda c z^0}},$$

$$\text{when } z^0 > \frac{4\alpha\epsilon}{\lambda c},$$

$$z(\tau) = \begin{cases} z^0 e^{-\alpha\epsilon\tau}, & 0 \leq \tau \leq \frac{1}{\alpha\epsilon} \log \frac{\lambda c z^0}{4\alpha\epsilon} \\ \frac{\frac{4}{\lambda c}}{\tau + \frac{1}{\alpha\epsilon} \left(1 - \log \frac{\lambda c z^0}{4\alpha\epsilon}\right)}, & \tau > \frac{1}{\alpha\epsilon} \log \frac{\lambda c z^0}{4\alpha\epsilon}. \end{cases} \quad (89)$$

From the explicit description of $z(\tau)$, we observe that

$$\sup_{0 \leq \tau < \infty} \tau z(\tau) = \frac{4}{\lambda c} \max \left(1, \frac{\lambda c z^0}{4\alpha\epsilon e} \right) \quad (90)$$

and

$$z^n < z(n) \leq \frac{\frac{4}{\lambda c} \max \left(1, \frac{\lambda c z^0}{4\alpha\epsilon e} \right)}{n}, \quad n \geq 1, \quad (91)$$

which directly gives the numerical E-condition (33)

$$\frac{u_{i+1}^n - u_i^n}{\Delta x} < \frac{E}{n\Delta t},$$

with

$$E = \frac{4}{c} \max \left(1, \frac{\lambda c z^0}{4\alpha\epsilon e} \right) = 4E_0 \max \left(1, \frac{\lambda c z^0}{4\alpha\epsilon e} \right). \quad (92)$$

Furthermore the fact

$$\lim_{t \rightarrow +\infty} \tau z(\tau) = \frac{4}{\lambda c} \quad (93)$$

means that, for an arbitrary $\eta > 0$, some constant $N > 0$ exists so that

$$\tau z(\tau) < \frac{4 + \eta}{\lambda c}, \quad \tau > N. \quad (94)$$

The inequality (94) means that, if $n > N$, the numerical E-condition (33)

$$\frac{u_{i+1}^n - u_i^n}{\Delta x} < \frac{E}{n\Delta t},$$

holds with

$$E = \frac{4 + \eta}{c} = (4 + \eta)E_0. \quad (95)$$

Finally suppose that the condition

$$z^0 < \frac{4e\alpha\epsilon}{\lambda c}$$

is satisfied, the inequality

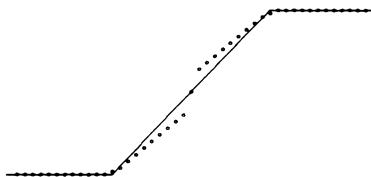
$$z(\tau) \leq \frac{1}{\tau} \cdot \frac{4}{\lambda c}$$

holds for $\tau > \max \left(0, \frac{1}{\alpha\epsilon} \log \frac{\lambda c z^0}{4\alpha\epsilon} \right)$. It means that the statement above is still valid with $\eta = 0$.

This completes the proof of theorem 3.

5. Conclusion

It is well known that numerical phenomena like flexion or jump often happen around the sonic points even if the exact solution is smooth there. The phenomena are observed especially in the region of rarefaction. A typical example is shown in the following figure.



Even if such a kind of numerical inconvenience stays as far as the difference increments $\Delta x, \Delta t > 0$ are arbitrary small, we may prove the convergence to exact solution. It means that the numerical inconvenience is another thing while it is related to the convergence property. The problem of numerical inconvenience is recognized when we are interested in the quality of numerical data calculated with the difference increments finite. In fact, the difference schemes discussed in the theorems above converge to the entropy solution (The consistency with E-condition itself guarantees the convergence to the entropy solution, or the convergence is already proved through the consistency with entropy inequality, see (Aiso, 1993)), but we may still recognize the numerical inconvenience.

We may expect that this kind of numerical inconvenience would not happen if a difference scheme realizes the numerical E-condition (17) with the constant $E = E_0 = \frac{1}{\inf |f''|}$. On the contrary, the value of constant E

in the numerical E-condition (17) implies the possibility or magnitude of numerical jump around sonic points. The larger the constant E , the bigger jump we may expect. The theorem 2 tells that we may take $E = 4E_0$ in general but that the assumption that no sonic point is between the data u_i^n and u_{i+1}^n of neighboring nodes cuts down it as small as $E = 2E_0$. It explain the experience that we have smaller numerical jumps around sonic points and better numerical results if we choose numerical initial data $\{u_i^0\}$ so that they take the value of sonic point ($u_i^0 = 0$) at each node around which the initial data u_0 take the value of sonic point ($u_0(x) = 0$).

The theorem 3 requires that the class of difference schemes have two kinds of machinery to make numerical jumps around sonic points. One comes from the convexity of flux function. It is similar to the theorem 2. The other comes from the short of numerical viscosity for entropy fix. The proof of theorem implies that the numerical inconvenience from the second machinery may be recognized only when the time step n is small. The important point is that we can avoid the numerical inconvenience from the second machinery in the region $t > T$ ($T > 0$) by taking Δt so small that $\bar{n} = \left[\frac{T}{\Delta t} \right]$ is as large enough as the theorem requires. Even though such numerical inconvenience is much remarkable for n small and ϵ small, we do not have to care unless we are interested in the initial layer. Therefore we conclude that the first machinery, which also happens in the case of theorem 2, is more essential and that some new idea is still needed to reduce the effect of machinery.

An implication for the system's case is the following.

- The numerical machinery observed in the case of theorem 2 happens in the system's case as well. It is still essential.
- The second numerical machinery of the theorem 3 is not restricted into the initial layer because the interaction of shock waves or contact may create a new centered rarefaction wave at any positive time $t > 0$. Therefore we need more careful choice of additional numerical viscosity for entropy fix.

Finally, we mention that the theorems 2 and 3 also implies the convergence of schemes in the category of $L^\infty \cap L_{loc}^1$. It is a little stronger than those in the category of BV , where the problem is often considered. Generally, schemes included in the theorem 3 are not monotone. It means that the theorem gives the convergence proof in $L^\infty \cap L_{loc}^1$ -category without assuming the monotonicity of schemes, while the work (Crandall and Majda, 1980) shows that monotone schemes converge to the entropy solution in the category.

References

- Aiso H (1993). Admissibility of difference approximation for scalar conservation laws. *Hiroshima Math. J.*, **23**(1), pp 15–61.
- Aiso H (1995). Higher Order-Accurate Difference Approximation for Scalar Conservation Laws and the Consistency with Entropy Condition. *Collection of Technical Papers, 6th International Symposium on Computational Fluid Dynamics*, 7–12.
- Aiso H (1996). A General Class of Higher Order-Accurate Difference Approximations for Scalar Conservation Laws Converging to the Entropy Solution. *Computational Fluid Dynamics '96 (Proceedings of Third ECCOMAS Computational Fluid Dynamics Conference.)*, 937–943.
- Crandall M and Majda A (1980). Monotone difference approximations for scalar conservation laws. *Math. Comp.*, **34**, pp 1–21.
- Engquist B and Osher S (1980). Stable and entropy satisfying approximations for transonic flow calculations. *Math. Comp.*, **34**, pp 45–75.
- Godunov S K (1959). Finite difference method for numerical computation of discontinuous solutions of the equations of fluid dynamics (in Russian). *Mat. Sb. (N.S.)*, **47**, pp 251–306.
- Olešnik O (1957). Discontinuous solutions of nonlinear differential equations. *Uspekhi Mat. Nauk. (N.S.)*, **12**, pp 3–73. English transl. in *Amer. Math. Soc. Transl., Ser. 2, vol. 26*, 95–172.
- Osher S (1984). Riemann solvers, the entropy condition, and difference approximations. *SIAM J. Numer. Anal.*, **21**(2), pp 217–235.
- Osher S and Tadmor E (1988). On the convergence of difference approximations to scalar conservation laws. *Math. Comp.*, **50**, pp 19–51.
- Roe P L (1992). Sonic Flux Formulae. *SIAM J. Sci. Statist. Comput.*, **13**, pp 611–630.
- Smoller J (1982). *Shock waves and reaction-diffusion equations*. Springer-Verlag, New York.
- Tadmor E (1984). Numerical viscosity and the entropy condition for conservative difference schemes. *Math. Comp.*, **43**, pp 369–382.
- Vol'pert A I (1967). Spaces BV and quasi-linear equations. *Math. USSR-Sb.*, **2**, pp 225–267.
- Vol'pert A I (1985). *Analysis in classes of discontinuous functions and of mathematical physics*. Martinus Nijhoff Publishers.

NEW RESULTS FOR RESIDUAL DISTRIBUTION SCHEMES

R. ABGRALL

*Mathématiques Appliquées,
Université Bordeaux I,
33 405 Talence Cedex,
France
Email: abgrall@math.u-bordeaux.fr*

AND

T. J. BARTH

*NASA Ames Research Center,
Information Sciences Directorate,
Moffett Field,
California, 94035-1000 USA.
Email: bath@nas.nasa.gov*

Abstract. In this paper, we present new results concerning the construction of upwind residual distribution schemes on unstructured meshes. These schemes are tailored to the steady state numerical solution of systems of first order conservation laws in more than one space dimension. The schemes under consideration have been previously introduced by P.L. Roe, H. Deconinck and their coworkers (Roe, 87; Roe, 90; Struijs, Deconinck and Roe, 91; Roe and Sidilkover, 92; Deconinck, Roe and Struijs, 93; Deconinck, Struijs, Bourgeois and Roe, 93) and developed specifically for use on simplicial (triangulated) meshes. In the present work, two separate but related topics are addressed: (1) how to construct mathematically well-founded variants of these schemes that are second order accurate at steady state, and (2) how to generalise these new schemes to more general systems of conservation laws and/or general (non-simplicial) element meshes. Numerical examples of transonic Euler flow are provided to verify the analysis and validate the solution quality improvements obtained with the new schemes.

1. Introduction

The development of accurate and truly robust numerical solution techniques for first order conservation laws on arbitrary meshes in multiple space dimensions has remained an elusive problem. Many currently used schemes employ sophisticated high resolution discretization techniques developed for structured meshes in the 70's-80's by van Leer, Roe, Osher, Harten, Yee, Sweby, and many others. The list of contributors to this technology is enormous. Some of the most significant contributions have been collected in (Hussaini, van Leer and Van Rosendal, 97). Even so, the quality of computed solutions is still questionable : some apparently simple problems such as computing the steady state lift and drag of an airfoil is still a difficult task. One reason is that high resolution schemes capable of accurately computing discontinuities still suffer from the damaging effects of spurious entropy production in smooth solution regimes. In part, this comes from the fact that these methods are usually devised for scalar 1D problems and then extended to multiple space dimensions and systems of conservation laws but their construction still relies too heavily on "1D ideas". Another criticism of current methods is their sensitivity to mesh distortions. In practice, it is still very difficult to construct meshes in 3D with very good quality. Consequently, the quality of the solution itself suffers and maybe questionable in many cases. Hence, it is natural to try to develop methods that do not intrinsically depend on regularity of the mesh. In the present work, our attention is focused on steady state solutions of the Euler equations on simplicial (triangulated) meshes. Even so, we demand that the techniques developed here generalize to more general systems such as magneto-hydrodynamics, Euler equations with conventional forms of chemistry, etc. as well as general (non-simplicial) meshes.

To overcome the above mentioned obstacles, some researchers over the last several years have tried to incorporate some of the favorable ideas contained in the 1D high resolution schemes (e.g. upwinding) into a finite element like framework. Some major contributions has been made by P.L. Roe, H. Deconinck, D. Sidilkover and their coauthors (Roe, 87; Roe, 90; Struijs, Deconinck and Roe, 91; Roe and Sidilkover, 92; Deconinck,Roe and Struijs, 93; Deconinck, Struijs, Bourgeois and Roe, 93). These so-called "fluctuation splitting" schemes were first developed for scalar transport equations and then formally extended to systems (see (Struijs, Deconinck and Roe, 91; Deconinck, Struijs, Bourgeois and Roe, 93) for example) by appealing to as much flow physics as possible. The fluctuation splitting schemes share many common features with the SUPG and Galerkin least-squares schemes of Hughes (Hughes and Mallet, 86) and the streamline diffusion method of Johnson (Johnson, 87). In the case of scalar advection, these schemes pro-

duce upwind biasing along streamlines that does not depend intrinsically on the mesh geometry. One advantage of these methods is that, at least for scalar equations, one can construct a second order accurate scheme on triangular meshes with a very compact stencil, i.e. the discretization stencil only uses only adjacent neighbors.

In the first part of the paper, we state some notation and remind the reader of some well known facts that are useful in later discussion. We then recall two important examples of upwind residual schemes, the N- and LDA- schemes, and state their properties. Both rely on a linearisation for the Euler equations, so we discuss the linearisation problem. In particular, we introduce the concept of linearisation via quadrature. Although the resulting schemes are not strictly speaking conservative, the schemes nevertheless exhibit a kind of Lax-Wendroff theorem that ensures that solutions, in the limit of quadrature points and mesh refinement, are indeed weak solutions of the Euler equations. We then show how to construct genuinely second order schemes for steady state problem which appeal to a local entropy principle. Numerical examples are given to illustrate the new techniques and to demonstrate the efficiency of our approach.

2. Notations

We are interested in the numerical approximation of the Euler equations of Fluid Mechanics in a domain Ω with boundary conditions,

$$\begin{aligned} \frac{\partial W}{\partial t} + \operatorname{div} \mathcal{F}(W) &= 0 \quad t > 0 \text{ and } x \in \Omega \\ W(x, 0) &= W_0(x) \quad x \in \Omega \\ \text{Boundary conditions} &\quad x \in \partial\Omega \end{aligned} \tag{1}$$

The flux $\mathcal{F} = (\mathcal{F}, \mathcal{G})$ and the conserved variables are given by

$$\begin{aligned} W &= (\rho, \rho u, \rho v, E)^T, F(W) = (\rho u, \rho u^2 + p, \rho uv, u(E + p))^T \\ \text{and } G(W) &= (\rho v, \rho uv, \rho v^2 + p, v(E + p))^T \end{aligned}$$

where ρ is the density, u and v are the components of the velocity, $E = \rho\epsilon + \frac{1}{2}\rho(u^2 + v^2)$ is the total energy. The system is closed by the equation of state relating the pressure p to the conserved variables, here we choose

$$p = (\gamma - 1) \left(E - \frac{1}{2}\rho(u^2 + v^2) \right) = (\gamma - 1)\rho\epsilon.$$

The ratio of specific heats γ is kept constant, $\gamma = 1.4$ in the applications.

The system (1) has to be supplemented by the entropy inequality which translates the second law of thermodynamics,

$$\frac{\partial S}{\partial t} + \frac{\partial(uS)}{\partial x} + \frac{\partial(vS)}{\partial y} \leq 0 \text{ on } \Omega. \tag{2}$$

The mathematical entropy is given by $S = -\rho h(s)$ (Harten, 83), where s is the physical entropy

$$s = c_v \log \left(\frac{p}{\rho^\gamma} \right) + s_0 \quad (3)$$

and h is any real valued function such that $h' > 0$ and $\frac{h''}{h'} < \gamma^{-1}$. In the practical examples, we take $h(x) = x$. If the flow is smooth, (3) is equivalent to

$$\frac{\partial s}{\partial t} + u \frac{\partial s}{\partial x} + v \frac{\partial s}{\partial y} = 0 (\geq 0) \quad (4)$$

and E. Tadmor has shown (Tadmor, 87) that the solution (if it is bounded) enjoys the minimum principle $s(x, t) \geq \min_{||y-x|| \leq t ||\vec{u}||_\infty} s(y, 0)$ where $||x||$ is the Euclidian norm of x and $||\vec{u}||_\infty$ is the L^∞ norm of the velocity field.

Since S is convex, the mapping $W \mapsto V = S_W$ is one-to-one on their domain of definition. It is well known that the flux \mathcal{F} when expressed in terms of the entropy variable V has symmetric Jacobian matrices.

Throughout the paper, we consider a two dimensional computational domain Ω that is triangulated by triangles¹. The set of triangles is $\{T_j\}_{j=1,\dots,nt}$. The mesh points are $\{M_i\}_{i=1,\dots,n_s}$. The vertices of a triangle T are $M_{i_1}, M_{i_2}, M_{i_3}$. When there is no ambiguity, they are denoted by their index in the list $\{M_i\}_{i=1,\dots,n_s}$, namely i_1, i_2, i_3 or simply by 1, 2, 3. To discretise (1), we consider the following residual scheme

$$|C_i| \frac{W_i^{n+1} - W_i^n}{\Delta t} + \sum_{T, M_i \in T} \Phi_i^T = 0. \quad (5)$$

In this equation, W_i^n is an approximation of $W(M_i, t_n)$, $|C_i|$ is the area of the dual control volume (see Figure 1) and the residual Φ_i^T are function of W_i^n and its neighboring values. The residuals are supposed to fulfill the following condition :

$$\sum_{M_i \in T} \Phi_i^T = \int_T \operatorname{div} \mathcal{F}^h dx \text{ for any triangle } T \quad (6)$$

where \mathcal{F}^h is an approximation of \mathcal{F} . In (Abgrall and Mer, 98), it is shown that under reasonable assumptions on the Φ_i^T 's (continuity, convergence of \mathcal{F}^h towards \mathcal{F} , continuity of \mathcal{F}^h on the edges of T) and the classical assumption of the Lax-Wendroff theorem (Lax and Wendroff, 60), the numerical solution converges to a weak solution of (1).

¹though our methods can be applied to more general meshes, for example meshes which elements are quadrilaterals (hexahedron in 3D)

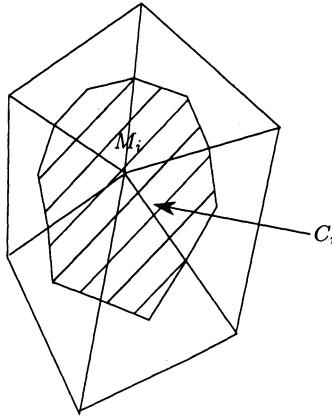


Figure 1. The dual cell is obtained by joining the midpoints of the edges starting from M_i and the centroids of the triangles containing M_i as a vertex.

3. Linearisation via the Z- and V- variables

The first step is to design a *conservative* linearisation of the Euler equations. This is a generalisation of the 1D linearisation procedure by P.L. Roe (Roe, 81) or its generalisation by A. Harten (Harten, Lax and van Leer, 83). In this section, we show how to proceed in the genuinely multidimensional case.

3.1. LINEARISATION IN THE Z VARIABLE

This a direct extension of the Roe linearisation in 1D, done in (Deconinck, Roe and Struijs, 93). The key remark is that the vector of conserved variables and the flux are quadratic in term of $Z := (\sqrt{\rho}, \sqrt{\rho}\vec{u}, \sqrt{\rho}H)$ in the cas of a perfect gaz. We can write

$$W = \frac{1}{2}D(Z).Z \quad , \quad \mathcal{F} = \frac{1}{2}\mathcal{R}(Z).Z$$

where $D(Z)$ is a matrix that depends linearly in Z , as well as R and S in $R(Z) = (R(Z), S(Z))$. The matrix D is invertible as soon as $\rho > 0$. One can interpolate linearly Z in any triangle T , and the following identity can be written (where Z^h is the interpolation)

$$\int_T \operatorname{div} \mathcal{F}(Z^h) dx = \mathcal{R}(\bar{Z}) D^{-1}(\bar{Z}) \int_T \nabla W(Z^h) dx$$

where the matrices are evaluated at $\bar{Z} = \frac{Z_1+Z_2+Z_3}{3}$. In (Abgrall and Barth, 99), it is shown that if $\rho_1 > 0$, $\rho_2 > 0$ and $\rho_3 > 0$, the averaged matrices $\mathcal{R}(\bar{Z}) D^{-1}(\bar{Z})$ have real eigenvalues and are diagonalisable.

The main restriction of this construction is that it is limited to a calorically perfect gas and to triangular type meshes.

3.2. LINEARISATION VIA THE V VARIABLES

In order to motivate this new linearisation, consider the problem of finding one on quadrilaterals. A simple way of proceeding would be to cut the element into two triangles, but there are several ways of cutting the element. The solution is non generic. Another way of proceeding would be to interpolate the data, say the conservative variables, on the element, $W^h = \sum_{vert.} W_i \Lambda_i$ (the Λ_i being the basis function, for example Q^1), and then to write

$$\begin{aligned} \int_Q \operatorname{div} \mathcal{F}(W^h) dx dy &= \int_Q A(W^h) W_x^h + A(W^h) W_y^h dx dy \\ &= \sum_{vert.} \left(\int_Q A(W^h)(\Lambda_j)_x + B(W^h)(\Lambda_j)_x dx dy \right) W_j \\ &= \sum_{vert.} K_j W_j \end{aligned} \tag{7}$$

with $K_j = \int_Q A(W^h)(\Lambda_j)_x + B(W^h)(\Lambda_j)_x dx dy$: we have exactly the same algebra as in the Z variable case, but we have to compute K_j and these matrices must be diagonalisable in R .

To reach that aim, we first consider the formulation in entropy variables : necessarily, the K_j matrices are diagonalisable in R because they are symmetric. But the problem of computing K_j remains.

The computation of K_j is very complex, but an alternative is to use quadrature formula. We illustrate this for the V variables interpolated on triangles by a linear interpolation

Consider an approximation of (7) using NQ -point numerical quadrature

$$\int_Q H(x) dx = \sum_{q_l \text{ quadrature points}} \omega_l H(q_l) + R_{NQ+1} \tag{8}$$

where ω_l are quadrature weights, q_l quadrature positions, and R_{NQ+1} is the numerical remainder term, NQ the number of quadrature points and $H(x) = \mathbf{A} = (A(V^h), B(V^h))$. This renders the scheme nonconservative in space.

We consider the two assumptions

Assumption (H1) Let \mathcal{T}_h be a regular triangulation. For all $C \in R$, there exists $C' \in R$ which depends only on W_0 and of \mathcal{T}_h such that $\forall V \in$

\mathcal{X}^h , $\|V\|_{L^\infty(R^2)} \leq C$ we have

$$\forall T, \forall i, M_i \in T, \left\| \Phi_i^T \right\| \leq C' h \sum_{M_j \in T} \|V_j - V_i\|. \quad (9)$$

Assumption (H2) $\forall V \in \mathcal{X}^h$,

$$\sum_{i=1}^3 \Phi_i^T(V) = |T| \left(\sum_{q_l \in Q} \omega_l \mathcal{A}(\pi_h V(q_l)) \right) \nabla \pi_h V,$$

where \mathcal{A} is the Jacobian matrix of the flux \mathcal{F} in the V variables.

Then, following (Hou and Le Floch, 94; Geiben, Kröner, and Rokyta, 93; Abgrall and Mer, 98), we get

Theorem 3.1 Let us consider an initial condition $W_0 \in L^\infty(R^d)^m$ and $\tau > 0$. Let W_h be an approximation. We assume that the scheme satisfies (H1) et (H2). We assume there exists a constant C that depends only on W_0 and a function $W \in L^2(R^d \times R^+)$ such that

$$\sup_h \sup_{x,y,t} \|W_h(x,y,t)\| \leq C$$

$$\lim_h \|W - W_h\|_{L^2_{loc}(R^d \times R^+)^m} = 0$$

Last, we assume that there exists a locally bounded, positive measure μ such that $\|\nabla \pi_h V_h\|$ tends to μ in the sense of distributions. Then W satisfies

$$\begin{aligned} & \left| \int_{Q \times [0,\tau]} \left(\frac{\partial \varphi}{\partial t} W(x,t) dx dt + \nabla_x \varphi(x,t) \cdot \vec{F}(W(x,t)) \right) dx dt \right. \\ & \quad \left. + \int_Q \varphi(x,0) W_0(x) dx \right| \\ & \leq \frac{C(\mathcal{T}_h, \mathcal{F})}{(k+1)!} \langle |\varphi|, \mu \rangle. \end{aligned}$$

where $C(\mathcal{T}_h, \mathcal{F})$ is a constant that only depends on \mathcal{T}_h and $\|D_V^{k+1} \mathcal{F}\|$ in $\{V, \|V\| \leq \sup_h V_h\}$.

How can this result be useful ? In practice, one compute a solution with a certain accuracy that is difficult to evaluate. The more quadrature points are used, the smaller the measure $\frac{\mu}{(p+1)!}$, hence one can reasonably expect that the error term is rapidly negligible compared the the computational error. If the mesh is regular, one can also expect that $C(\mathcal{T}_h, \mathcal{F})$ is not too

large because the entropy is analytic in term of W . More precisely, we expect that far from vacum, the error term is rapidly small.

The second remark is that the measure μ should be localised on the discontinuity of the solution. Elsewhere it is 0. This can be seen from the proof. Hence, it is not necessary to use a uniformly high order quadrature formula : most of the time, a quadrature formula that is at least one order of accuracy more than the expected accuracy of the solution is enough. This is the strategy that we have taken in the numerical results.

4. The system N- and LDA- schemes

Two important schemes are the N and the LDA (Low Diffusion Advection) schemes, (van der Weide and Deconinck, 96). They have been first derived for scalar convection problems and then extended formally to systems. We recall them in their system version.

4.1. THE SYSTEM N SCHEME

We set $\Phi_i^T = K_i^+ (\widetilde{W}_i - \widetilde{W})$ where $K_i = \bar{A}n_x^i + \bar{B}n_y^i$ and $\widetilde{W}_i = D(\bar{Z})Z_i$ ². In order to recover the conservation relation, we must have

$$\left(\sum_{i=1,3} K_i^- \right) \widetilde{W} = \sum_{i=1,3} K_i^- \widetilde{W}_i \quad (10)$$

To define \widetilde{W} , one has *a priori* to invert the matrix

$$\sum_{i=1,3} K_i^-,$$

in some cases, this may be impossible. When it is possible, we denote by N the matrix

$$N = \left(\sum_{i=1,3} K_i^- \right)^{-1}.$$

However, we show in (Abgrall and Barth, 99) that, for the Euler equations, $K_i^+ \widetilde{W}$ is always defined.

More precisely, the space of state R^4 can be written as $R^4 = Rr_0 \oplus H$ where

$$r_0 = \left(1, \bar{u}, \bar{v}, \frac{1}{2} (\bar{u}^2 + \bar{v}^2) \right)^T.$$

²(n_x^i, n_y^i) = \vec{n}_i is the inward normal opposite to node M_i , $i=1,3$ in T .

The space H is the image of $\pi(\mathbf{W}) = \mathbf{W} - \frac{\langle \mathbf{W}, \mathbf{v}_0 \rangle}{\langle \mathbf{r}_0, \mathbf{v}_0 \rangle} \mathbf{r}_0$. The vector v_0 is $\nabla_W s$ evaluated at the average state, it is the (common) left eigenvector of \bar{A} and \bar{B} . We can give a meaning to the inverse of $\sum_{i=1,3} K_i^-$, denoted by N . The proof is valid for a linearised symmetric system.

The N scheme is linearly dissipative when the system is symmetrisable. More precisely, if the linearisation is carried out in the entropy variables V , T. Barth (Barth, 97; Abgrall and Barth, 99) has shown that,

Lemma 4.1 *If the matrices K_i are symmetric, one has*

$$\sum_{M_i \in T} \langle V_i, \Phi_i^T \rangle = \frac{1}{2} \sum_{M_i \in T} \langle V_i, K_i V_i \rangle + Q_N(V_1, V_2, V_3) \quad (11)$$

where

$$\begin{aligned} 2Q_N(V_1, V_2, V_3) &= - \langle \Phi^T N, \Phi^T \rangle \\ &+ \sum_{M_i \in T} \left(\langle V_i, K_i^+ V_i \rangle - \langle K_i^+ V_i, M K_i^+ V_i \rangle \right) \\ &+ \sum_{M_i \in T} \left(\langle V_i, -K_i^- V_i \rangle - \langle -K_i^+ V_i, M - K_i^- V_i \rangle \right). \end{aligned} \quad (12)$$

The quadratic form Q_N is positive : the N scheme is locally dissipative.

He shows that each of the three terms in (12) is positive. In the case of a linear problem, the sum of $\langle V_i, K_i V_i \rangle$ cancels, and we get a global energy stability result for the N-scheme. If we had a linearisation in the *entropy variable* $V = \nabla_W S$, we could interpret $\frac{1}{2} \sum_{M_i \in T} \langle V_i, K_i V_i \rangle$ as

$$\int_{\partial T} (\langle V, K_n V \rangle)^h d\sigma$$

where the “energy” $\langle V, K_n V \rangle$ is piecewise linearly interpolated by $(\langle V, K_n V \rangle)^h$. Hence, using exactly the same technique as in (Abgrall and Mer, 98), if the conditions of Lax Wendroff theorem are true, then the limit of the numerical solutions satisfies an entropy inequality, namely

$$\frac{\partial S}{\partial t} + \operatorname{div}(uS) \leq 0.$$

4.2. THE SYSTEM LDA SCHEME

The straightforward extension of the scalar LDA scheme is given by

$$\Phi_i^T = -K_i^+ N \Phi^T \quad (13)$$

where the N matrix is given by (12). The conservation relation is obviously satisfied. This scheme is upwind and LP. More precisely, when the linearisation is carried out in the entropy variables, we have

Lemma 4.2 *Let us set*

$$V^+ = -N \left(\sum_{i=1,3} K_i^+ V_i \right) \quad \text{and} \quad V^- = N \left(\sum_{i=1,3} K_i^- V_i \right).$$

Then we have

$$\sum_{M_i \in T} \langle V_i, \Phi_i^T \rangle = \frac{1}{2} \sum_{M_i \in T} \langle V_i, K_i V_i \rangle + Q_{LDA}(V_1, V_2, V_3) \quad (14)$$

with

$$\begin{aligned} 2Q_{LDA}(V_1, V_2, V_3) &= \sum_{i=1}^3 \langle V^+ - \tilde{V}_i, K_i^+ (V^+ - \tilde{V}_i) \rangle \\ &+ \sum_{i=1}^3 \langle \tilde{V}_i - V^-, K_i^+ (\tilde{V}_i - V^-) \rangle \\ &+ \sum_{M_i \in T} \left(\langle V_i, K_i^+ V_i \rangle - \langle K_i^+ V_i, M K_i^+ V_i \rangle \right) \\ &+ \sum_{M_i \in T} \left(\langle V_i, -K_i^- V_i \rangle - \langle -K_i^+ V_i, M - K_i^- V_i \rangle \right). \end{aligned} \quad (15)$$

4.3. COMPARISON BETWEEN THE LDA AND N SCHEMES

It is also possible to compare Q_N and Q_{LDA} for a symmetrisable system when the linearisation is done via the entropy variables,

Lemma 4.3 *We have*

$$Q_{LDA}(V_1, V_2, V_3) \leq Q_N(V_1, V_2, V_3)$$

This result states that the N scheme is more dissipative than the LDA scheme.

5. Construction of schemes that are second order at steady state

5.1. A LP POSITIVE SCALAR SCHEME

In this section, we consider the scalar equation

$$\frac{\partial u}{\partial t} + \langle \vec{\lambda}, \nabla u \rangle = 0. \quad (16)$$

In (Abgrall and Barth, 99)), we show that the PSI scheme (for Positive Streamwise Invariant) can be recovered by setting the residual to

$$\Phi_i = l\Phi_i^N + (1-l)\Phi_i^{LDA}$$

and using obvious notations. The limiter l is defined by
 $l = \min(1, \max(\varphi(r_1), \varphi(r_2)))$ where

$$\varphi(x) = \begin{cases} \frac{r}{r-1} & \text{if } r \leq 0 \\ 0. & \text{else.} \end{cases} \quad (17)$$

and $r_j = \frac{\Phi_j^{LDA}}{\Phi_j^N}$.

The PSI scheme has several remarkable properties. First, it is positive, that is if u_i^n for any i , then $u_i^{n+1} \geq 0$ as is the scalar N scheme. This property is true under a CFL like condition. Second, this scheme is upwind, that is if $\langle \lambda, \vec{n}_i \rangle < 0$, then $\Phi_i = 0$. Third, it is linear preserving (LP), that is, if $\Phi = 0$ then $\Phi_i = 0$. This property ensures that the scheme is formally second order accurate *at steady state*.

5.2. A LP STABLE SCHEME FOR (1)

5.2.1. *Comments on the system N scheme*

All the numerical experiments that have been conducted with the system N scheme indicate it is a very stable, robust and monotone scheme. By saying it is monotone, we mean that whatever strong the discontinuities are, there are no pre- or post- discontinuity oscillations. In particular, it is a numerical fact that the physical entropy follows the following semi-discrete relation

$$|C_i| \left(\frac{ds}{dt} \right)_i + \sum_{T, M_i \in T} \sum_{M_j \in T} c_{ij}^T (s_i - s_j) \geq 0 \quad (18)$$

for some *positive* numbers c_{ij} .

Since r_0 is a common eigenvector of \bar{A} and \bar{B} , we have

$$\langle v_i, \pi' \Phi_i^{N,T} \rangle = \sum_{M_j \in T} c_{ij}^T \langle v_i, \pi' (\mathbf{W}_i - \mathbf{W}_j) \rangle. \quad (19)$$

where the c_{ij}^T are the coefficients for the scalar N scheme where $\lambda = \bar{u}$, i.e.

$$k_j = \frac{1}{2} \langle \bar{u}, \vec{n}_j \rangle, \quad c_{ij}^T = \frac{k_i^+ k_j^-}{\sum_{j=1,3} k_j^-}$$

$$\text{and } \pi' = Id - \pi$$

Equation (19) states that the system N scheme is *positive* on the (linearised) entropy wave.

Throughout the paper, we assume that a similar relation does exists on the shear and accoustic waves modes ; more precisely, we assume that a relationship of the type

$$\langle v_i, \pi \Phi_i^{N,T} \rangle \leq \sum_{M_j \in T} c_{ij}^T \langle v_i, \pi (\mathbf{W}_i - \mathbf{W}_j) \rangle \quad (20)$$

holds even if we have been unable to prove it. The first inequality (19) states a monotone behavior of the projection of Φ_i^T on the entropy wave. The second inequality (20) states the same for the projection of Φ_i^T on the accoustic and shear modes.

5.3. CONSTRUCTION OF AN ENTROPY STABLE LP SCHEME

In the following analysis, we set

$$v_i = \nabla s(W_i)$$

and we consider the semi-discrete scheme

$$|C_i| \left(\frac{dW}{dt} \right)_i + \sum_{T, M_i \in T} \Phi_i^T = 0$$

where

$$\Phi_i^T = \ell \Phi_i^{N,T} + (\text{Id} - \ell) \Phi_i^{\text{LDA},T}. \quad (21)$$

Here ℓ is a matrix which structure has to be defined in order to get a monotone entropy stable scheme.

We first left multiply $\left(\frac{dW}{dt} \right)_i$ by v_i , and we get

$$|C_i| \left(\frac{ds}{dt} \right)_i + \sum_{T, M_i \in T} \langle v_i, \Phi_i^T \rangle = 0$$

The idea is to construct ℓ in such a way that

$$\frac{\langle v_i, \Phi_i^T \rangle}{\langle v_i, \Phi_i^{N,T} \rangle} \geq 0. \quad (22)$$

If it is possible, by combining this inequality to (19-20), provided that $\frac{\langle v_i, \Phi_i^T \rangle}{\langle v_i, \Phi_i^{N,T} \rangle}$ is bounded, we recover formally a bound on the solution, under a CFL like condition. Now we show how it is possible to construct such a matrix ℓ .

To begin with, recall the decomposition of the state space, $R^4 = Rr_0 \oplus H$ given, for any state $W \in R^4$ by

$$W = \pi(\mathbf{W}) + \pi'(\mathbf{W}), \quad \pi'(\mathbf{W}) = \mathbf{l}(\mathbf{W})\mathbf{r}_0 = \frac{\langle \mathbf{W}, \mathbf{v}_0 \rangle}{\langle \mathbf{r}_0, \mathbf{v}_0 \rangle} \mathbf{r}_0$$

$$\text{and } y = \pi(\mathbf{W}) = \mathbf{W} - \frac{\langle \mathbf{W}, \mathbf{v}_0 \rangle}{\langle \mathbf{r}_0, \mathbf{v}_0 \rangle} \mathbf{r}_0.$$

The vectors r_0 and v_0 are evaluated for an averaged set of Jacobian matrices \bar{A} and \bar{B} . From a physical point of view, $\mathbf{l}(W)$ is the component of W on the entropy wave r_0 , while $\pi(\mathbf{W})$ is the sum of the acoustic and shear waves.

The key remark is to notice that the N scheme and the *LDA* scheme have a simple expression for this decomposition because r_0 is a common eigenvector of \bar{A} and \bar{B} . More precisely, we have

$$\begin{aligned} \Phi_i^N &= \sum_{j=1; j \neq i}^3 K_i^+ N K_j^- \pi' (\tilde{\mathbf{W}}_i - \tilde{\mathbf{W}}_j) + \sum_{j=1; j \neq i}^3 \mathbf{K}_i^+ \mathbf{N} \mathbf{K}_j^- \pi (\tilde{\mathbf{W}}_i - \tilde{\mathbf{W}}_j) \\ &= \left(\frac{\sum_{j=1; j \neq i}^3 k_i^+ k_j^- l(W_i - W_j)}{\sum_{j=1,3} k_j^-} \right) r_0 + \sum_{j=1; j \neq i}^3 K_i^+ N K_j^- \pi (\tilde{\mathbf{W}}_i - \tilde{\mathbf{W}}_j) \end{aligned}$$

and a similar expression for the LDA scheme. For this reason, we set

$$\ell = \mathbf{l}_1 \pi' + \mathbf{l}_2 \pi. \quad (23)$$

In other words, the matrix ℓ has two components. One acts only on the components on the entropy wave, and the other component only plays on

the shear/accoustic waves. Then we evaluate the entropy production within a single triangle,

$$\begin{aligned}
\langle v_i, \Phi_i \rangle &= \langle v_i, \ell \Phi_i^N \rangle + \langle v_i, (\mathbf{Id} - \ell) \Phi_i^{\text{LDA}} \rangle \\
&= (l_1 \langle v_i, \pi'(\Phi_i^N) \rangle + (1 - l_1) \langle v_i, \pi'(\Phi_i^{\text{LDA}}) \rangle) \\
&\quad + (l_2 \langle v_i, \pi(\Phi_i^N) \rangle + (1 - l_2) \langle v_i, \pi(\Phi_i^{\text{LDA}}) \rangle) \\
&= \left(l_1 + (1 - l_1) \frac{\langle v_i, \pi'(\Phi_i^{\text{LDA}}) \rangle}{\langle v_i, \pi'(\Phi_i^N) \rangle} \right) \langle v_i, \pi'(\Phi_i^N) \rangle \\
&\quad + \left(l_2 + (1 - l_2) \frac{\langle v_i, \pi(\Phi_i^{\text{LDA}}) \rangle}{\langle v_i, \pi(\Phi_i^N) \rangle} \right) \langle v_i, \pi(\Phi_i^N) \rangle
\end{aligned}$$

We define l_1 and l_2 by the two conditions

– For $i = 1, 3$,

$$l_1 + (1 - l_1) \frac{\langle v_i, \pi'(\Phi_i^{\text{LDA}}) \rangle}{\langle v_i, \pi'(\Phi_i^N) \rangle} \geq 0,$$

– For $i = 1, 3$,

$$l_2 + (1 - l_2) \frac{\langle v_i, \pi(\Phi_i^{\text{LDA}}) \rangle}{\langle v_i, \pi(\Phi_i^N) \rangle} \geq 0.$$

Following the developments of section 5.1, we set

$$\begin{aligned}
l_1 &= \min(1, \max(\varphi(r'_1), \varphi(r'_2), \varphi(r'_3))) \\
l_2 &= \min(1, \max(\varphi(r_1), \varphi(r_2), \varphi(r_3)))
\end{aligned} \tag{24}$$

where $r_i = \frac{\langle v_i, \pi(\Phi_i^{\text{LDA}}) \rangle}{\langle v_i, \pi(\Phi_i^N) \rangle}$ and $r'_i = \frac{\langle v_i, \pi'(\Phi_i^{\text{LDA}}) \rangle}{\langle v_i, \pi'(\Phi_i^N) \rangle}$ for $i = 1, 3$.

6. Numerical tests

We illustrate our methods on a classical example, the computation of a flow over a NACA0012, the Mach number at infinity is $M_\infty = 0.85$ and the angle of attack is 1° .

6.1. LINEARISATION ISSUES

We provide the pressure coefficient for N scheme in entropy variables (Figure 2-a). An adaptive quadrature formula has been used, namely a one point quadrature in the smooth regions, and a four point one elsewhere.

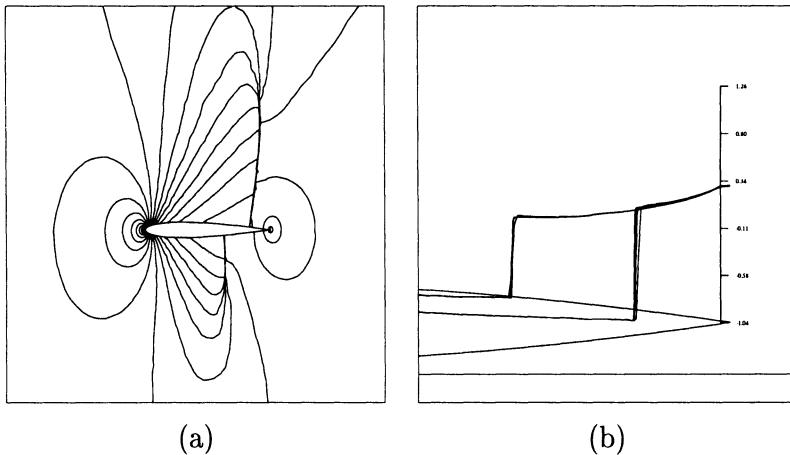


Figure 2. (a) After mesh refinement, isovales of cp , $\min = -1.02, \max = 1.22$. (b) Comparison between the numerical scheme in the V variables with 4 quadrature points and the solution obtained by the conservative scheme

The smoothness is estimated by computing the variation of the mathematical entropy within one triangle. If it is low, the solution is regular. The Figure 2-b shows cuts of the pressure coefficient on the body surface given by four computations. Three of them are obtained with this non conservative scheme with various refined meshes, the last one by the conservative scheme (in Z variables), on the most refined mesh. It is clear that the results are very close, it is difficult to tell which one is the best.

6.2. ACCURACY ISSUE

Here we compare a standard MUSCL scheme on triangular meshes to the blended scheme presented above. The Figure 3 represents the Mach number contours. The slip line is much better represented by the blended scheme. A carefull look at the solutions shows that the shock is also thiner. Last, the entropy production of the new scheme is also much less important, see (Abgrall and Barth, 99) for details.

7. Conclusion

In this paper, we have presented recent results on upwind fluctuation distributive schemes : a second order extension of the PSI scheme on systems, and a new linearisation proceedure. In both case, the results are very encouraging. More details can be found in (Abgrall and Barth, 99; Abgrall and Barth, 99) : the procedures are discussed in much more details, and other numerical examples assess our claims.

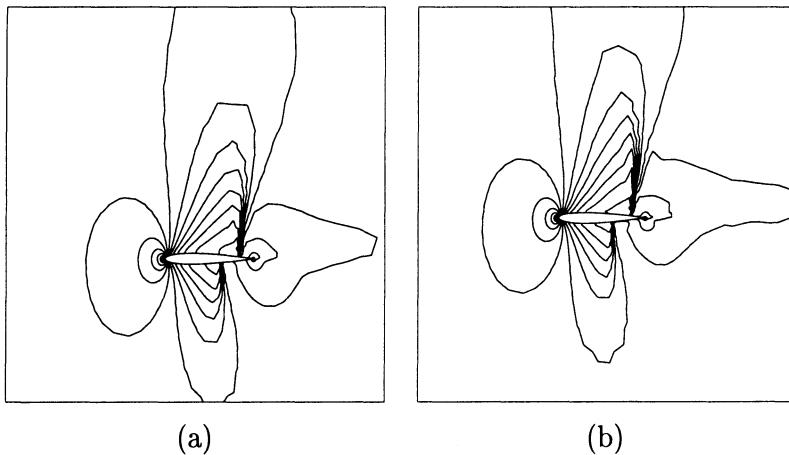


Figure 3. Pressure coefficient contours for the MUSCL scheme (a) and the blended scheme (b).

References

- R. Abgrall and T.J. Barth. Linearisation via the entropy variables : application to residual distributive schemes. , 1999. in preparation.
- R. Abgrall and K. Mer. Un théorème de type Lax–Wendroff pour les schémas distributifs. Technical Report 98010, Mathématiques Appliquées de Bordeaux, Mars 1998.
- T.J. Barth. Some working notes on the N scheme. Private communication, 1997.
- H. Deconinck, P.L. Roe, and R. Struijs. A multidimensional generalisation of Roe's difference splitter for the Euler equations. *Computer and Fluids*, 22:215–222, 1993.
- H. Deconinck, R. Struijs, G. Bourgeois, and P.L. Roe. Compact advection schemes on unstructured meshes. VKI Lecture Series 1993–04, Computational Fluid Dynamics, 1993.
- M. Geiben, D. Kröner, and M. Rokyta. A Lax-Wendroff type theorem for cell-centered finite volume schemes in 2-D. *Report No. 278, Sonderforschungsbereich 256, Rheinische Friedrich-Wilhelms-Universität, Bonn*, 1993.
- A. Harten. On the symmetric form of conservation laws with entropy. *J. Comp. Phys.*, 49:151–164, 1983.
- A. Harten, P.D. Lax, and B. van Leer. On upstream differencing and Godunov type schemes for hyperbolic conservation laws. *Siam Rev.*, 25:35–61, 1983.
- T. Y. Hou and P. G. Le Floch. Why nonconservative schemes converge to wrong solutions : error analysis. *Math. Comp.*, 62(206):497–530, 1994.
- T. J. R. Hughes and M. Mallet. A new finite element formulation for CFD: III. the generalized streamline operator for multidimensional advective-diffusive systems. *Comp. Meth. Appl. Mech. Engrg.*, 58:305–328, 1986.
- M.Y. Hussaini, B. van Leer, and J. Van Rosendal, editors. *Upwind and High Resolution Schemes*. Springer Verlag, 1997.
- C. Johnson. *Numerical Solution of Partial Differential Equations by the Finite Element Method*. Cambridge University Press, Cambridge, 1987.
- C. Johnson and A. Szepessy. Convergence of the shock-capturing streamline diffusion finite element methods for hyperbolic conservation laws. *Math. Comp.*, 54:107–129, 1990.
- P. Lax and B. Wendroff. Systems of conservation laws. *Comm. Pure Appl. Math.*,

- 13:381–394, 1960.
- R. Abgrall. Toward the ultimate conservative scheme : Following the quest. *J. Comput. Phys.*, submitted.
- P. L. Roe. Approximate Riemann solvers, parameter vectors, and difference schemes. *J. Comput. Phys.*, 43, 1983.
- P. L. Roe and D. Sidilkover. Optimum positive linear schemes for advection in two and three dimensions. *Siam J. Numer. Anal.*, 29(6):1542–1588, 1992.
- R. Struijs, H. Deconinck, and P. L. Roe. Fluctuation Splitting Schemes for the 2D Euler equations. *VKI LS 1991-01, Computational Fluid Dynamics*, 1991.
- P. L. Roe. Linear advection schemes on triangular meshes. Technical Report CoA 8720, Cranfield Institute of Technology, 1987.
- P. L. Roe. “Optimum” upwind advection on a triangular mesh. Technical report, ICASE, NASA Langley R.C., 1990.
- E. Tadmor. The numerical viscosity of entropy stable schemes for systems of conservation laws. *Math. Comp.*, 49:91–103, 1987.
- E. van der Weide and H. Deconinck. Positive matrix distribution schemes for hyperbolic systems. In *Computational Fluid Dynamics '96*, pages 747–753. Wiley, 1996.
- B. van Leer. Towards the ultimate conservative difference schemes V. A second order sequel to Godunov's method. *J. Comp. Phys.*, 32, 1979.

SIMULATIONS OF RELATIVISTIC JETS WITH GENESIS

M. A. ALOY, J. M. IBANEZ, J. M. MARTI

Departamento de Astronomía y Astrofísica,

UVEG, 46100 Burjassot, Spain.

Emails: Miguel.A.Aloy@uv.es, Jose.M.Ibanez@uv.es,

Jose.M.Marti@uv.es

J. L. GOMEZ

Instituto de Astrofísica de Andalucía (CSIC)

Granada, Spain.

Email: jlgonz@iaa.es

AND

E. MUELLER

Max-Planck-Institut für Astrophysik,

85748 Garching, Germany.

Email: ewald@mpa-garching.mpg.de

Abstract. The multidimensional relativistic hydrodynamic code GENESIS has been used to obtain first results of 3D simulations of relativistic jets. We have studied the influence of a slight perturbation of the injection velocity field on the morphodynamics of otherwise axisymmetric relativistic jets.

1. Introduction

Astrophysical jets are continuous channels of plasma produced by some active galactic nuclei that are currently observed in radio frequencies. The relativistic nature of the plasma has been inferred from (essentially) two observational evidences: (i) the existence of superluminal motions of some radio components and, (ii) the high flux variability (even smaller than one day for some sources). Since several years the dynamical and morphological properties of axisymmetric relativistic jets are investigated by means of relativistic hydrodynamic simulations either steady fluid jets –e.g., (Wilson, 1987)– or dynamically evolving ones –e.g., (van Putten, 1993), (Duncan and

Hughes, 1994), (Martí et al., 1997), (Komissarov and Falle, 1998) –. In addition, relativistic MHD simulations have been performed in 2D – (Koide et al., 1996), (Koide, 1997) – and 3D – (Nishikawa et al., 1997), (Nishikawa et al., 1998)–. In their 3D simulations, (Nishikawa et al., 1997) and (Nishikawa et al., 1998), have studied mildly relativistic jets (Lorentz factor, $W = 4.56$) propagating both along and obliquely to an ambient magnetic field.

In this work we report on high-resolution 3D simulations of relativistic jets with the largest beam flow Lorentz factor performed up to now (7.09), the largest resolution (8 cells per beam radius), and covering the longest time evolution (75 normalized time units; a normalized time unit is defined as the time needed for the jet to cross a unit length). These facts together with the high performance of our hydrodynamic code allowed us to study the morphology and dynamics of 3D relativistic jets for the first time.

The calculations have been performed with the high-resolution 3D relativistic hydrodynamics code GENESIS (Aloy et al., 1999a), which is an upgraded version of the code developed by (Martí et al., 1997). GENESIS integrates the 3D relativistic hydrodynamic equations in conservation form in Cartesian coordinates including an additional conservation equation for the beam-to-external density fraction to distinguish between beam and external medium fluids. The code is based on a *method of lines* which first discretizes the spatial part of the relativistic Euler equations and solves the fluxes using the Marquina's flux formula (Donat et al., 1998). Then the semidiscrete system of ordinary differential equations is solved using a third order Runge-Kutta algorithm (Shu and Osher, 1989). High spatial accuracy is achieved by means of a PPM third order interpolation (Colella and Woodward, 1984). The computations were performed on a Cartesian domain (X,Y,Z) of size $15R_b \times 15R_b \times 75R_b$ ($120 \times 120 \times 600$ computational cells), where R_b is the beam radius. The jet is injected at $z = 0$ in the direction of the positive z -axis through a circular nozzle defined by $x^2 + y^2 \leq R_b^2$. Beam material is injected with a beam mass fraction $f = 1$, and the computational domain is initially filled with an external medium ($f = 0$).

We have considered a 3D model corresponding to model C2 of (Martí et al., 1997), which is characterized by a beam-to-external proper rest-mass density ratio $\eta = 0.01$, a beam Mach number $M_b = 6.0$, and a beam flow speed $v_b = 0.99c$ (c is the speed of light) or a beam Lorentz factor $W_b \approx 7.09$. An ideal gas equation of state (EOS) with an adiabatic exponent $\gamma = 5/3$ describes both the jet matter and the ambient gas. A simple γ -law EOS is valid considering that the collision mean free path of the particles forming the jet plasma is very large and, essentially, it may be considered as a collisionless gas. The effects of changing from $\gamma = 5/3$ to $\gamma = 4/3$ are only important for the morphology of the cocoon (Martí et al., 1997),

which is not directly observed (because within it there are not strong shocks able to emit in radio wavelengths). The beam is assumed to be in pressure equilibrium with the ambient medium.

The evolution of the jet was simulated up to $T \approx 150R_b/c$, when the head of the jet is about to leave the grid. The scaled final time $T \approx 4.6 \cdot 10^4 (R_b/100\text{ pc})$ yr is about two orders of magnitude smaller than the estimated ages of powerful jets. Hence, our simulations cannot describe the long term evolution of these sources.

Non-axisymmetry was imposed by means of a helical velocity perturbation at the nozzle given by

$$v_b^x = \zeta v_b \cos\left(\frac{2\pi t}{\tau}\right), \quad v_b^y = \zeta v_b \sin\left(\frac{2\pi t}{\tau}\right), \quad v_b^z = v_b \sqrt{1 - \zeta^2}, \quad (1)$$

where ζ is the ratio of the toroidal to total velocity and τ the perturbation period (i.e., $\tau = T/n$, n being the number of cycles completed during the whole simulation). This velocity field causes a differential rotation of the beam. The perturbation is chosen such that it does not change the velocity modulus, (i.e., mass, momentum and energy fluxes of the beam are preserved).

2. Morphodynamics of 3D relativistic jets

Here we consider two models: A, which has a 1% perturbation in helical velocity ($\zeta = 0.01$) and $n = 50$ and B, with $\zeta = 0.05$ and $n = 15$. Figure 1 shows various quantities of the model A in the plane $y = 0$ at the end of the simulation. Two values of the beam mass fraction are marked by white contour levels. The beam structure is dominated by the imposed helical pattern with amplitudes of $\approx 0.2R_b$ and $\approx 1.2R_b$ for A and B, respectively.

The overall jet's morphology is characterized by the presence of a highly turbulent, subsonic cocoon. The pressure distribution outside the beam is nearly homogeneous giving rise to a symmetric bow shock (Fig. 1b) in model A. Model B shows a very inhomogeneous pressure distribution in the cocoon. As in the classical case (Norman, 1996), the relativistic 3D simulation shows less ordered structures in the cocoon than the axisymmetric models. As seen from the beam mass fraction levels, the cocoon remains quite thin ($\sim 2R_b$) in A and widens ($\sim 4R_b$) in B.

The flow field outside the beam shows that the high velocity backflow is restricted to a small region in the vicinity of the hot spot (Fig. 1e), the largest backflow velocities ($\sim 0.5c$) being significantly smaller than in 2D models. The flow annulus with high Lorentz factor found in axisymmetric simulations is also present, but it is reduced to a thin layer around the beam and possesses sub-relativistic speeds ($\sim 0.25c$) in model A and mildly

relativistic (~ 0.7) in B. The size of the backflow velocities in the cocoon do not support relativistic beaming in case of small perturbations but such possibility is open in larger ones.

Within the beam the perturbation pattern is superimposed to the conical shocks at about 26 and $50 R_b$. The beam of A does not exhibit the strong perturbations (deflection, twisting, flattening or even filamentation) found by other authors (Norman, 1996) for 3D classical hydrodynamic jets; (Hardee, 1996) for 3D classical MHD jets). This can be taken as a sign of stability, although it can be argued that our simulation is not evolved far enough. For n15p01, the beam is about to be disrupted at the end of our simulation. Obviously, the beam cross section and the internal conical shock structure are correlated (Figure 1).

The helical pattern propagates along the jet at nearly the beam speed which could yield to superluminal components when viewed at appropriate angles. Besides this superluminal pattern, the presence of emitting fluid elements moving at different velocities and orientations could lead to local variations of the apparent superluminal motion within the jet. This is shown in Fig. 1f, where we have computed the mean (along each line of sight, and for a viewing angle of 40 degrees) local apparent speed. The distribution of apparent motions is inhomogeneous and resembles that of the observed individual features within knots in M87 (Biretta, Zhou and Owen, 1995).

The jet can be traced continuously up to the hot spot which propagates as a strong shock through the ambient medium. Beam material impinges on the hot spot at high Lorentz factors in A case, but the beam Lorentz factor strongly decreases for B. We could not identify a terminal Mach disk in the flow. We find flow speeds near (and in) the hot spot much larger than those inferred from the one dimensional estimate. This fact was already noticed for 2D models by (Komissarov and Falle, 1996) and suggested by them as a plausible explanation for an excess in hot spot beaming.

We find a layer of high specific internal energy (typically more than a tenfold larger than that of the gas in the beam core, see Fig. 1d) surrounding the beam like in previous axisymmetric models (Aloy et al., 1999a). The region filled by the shear layer is defined by $0.2 < f < 0.95$. It is mainly composed of forward moving beam material at a speed smaller than the beam speed (Fig. 1e). The intermediate speed of the layer material is due to shear in the beam/cocoon interface, which is also responsible for its high specific internal energy. The shear layer broadens with distance from $0.2R_b$ near the nozzle to $1.1R_b$ (in A) or $2.0R_b$ (in B) near the head of the jet. The diffusion of vorticity caused by numerical viscosity is responsible for the formation of the boundary layer. Although being caused by numerical effects (not by the physical mechanism of turbulent shear) the properties of PPM-based difference schemes are such that they can mimic turbulent flow

to a certain degree (Porter and Woodward, 1994). In our case, the PPM numerical viscosity, coming from the spatial interpolation procedure, might be even smaller than that of (Porter and Woodward, 1994) –because they make use of a simplified version of the PPM algorithm without the discontinuity detection and steepening procedures of (Colella and Woodward, 1984)–. However, the use of the Marquina’s flux formula provides a source of heat conduction and diffusion (Donat and Marquina, 1996) and therefore, we have a numerical description of the fluid close to that of (Porter and Woodward, 1994) –*i.e.*, inviscid and with a small amount of thermal conduction– but for a 3D-relativistic fluid. In addition, we have done a convergence study decreasing the resolution by a factor of two, and we have still found the above mentioned shear layer (Aloy *et al.*, 1999a).

The existence of such a boundary layer has been invoked by several authors (Komissarov, 1990), (Laing, 1996) to interpret a number of observational trends in FRI radio sources. Such a layer will produce a *top-bottom* asymmetry due to the light aberration (Aloy *et al.*, 1999b), and additionally, it can be used to explain the rails in polarization found by (Attridge, Roberts and Wardle, 1999). Other authors (Swain, Bridle and Baum, 1998) have found evidence for these boundary layers in FRII (3C353) radio sources.

The jet’s propagation proceeds in two distinct phases. First it propagates according to a linear of 1D phase, and then the behavior depends on the strength of the perturbation: it accelerates to a propagation speed which is $\approx 20\%$ larger than the corresponding 1D estimate in model A or it decelerates up to $0.37c$. The second result partially agrees with the one obtained by Nishikawa *et al.* (1997, 1998).

The axial component of the momentum of the beam particles (integrated across the beam) along the axis decreases by more than a 30% within the first $60 R_b$. Neglecting pressure and viscous effects, and assuming stationarity the axial momentum should be conserved, and hence the beam flow is decelerating. The momentum loss goes along with the growth of the boundary layer.

In model A, although the beam material decelerates, its terminal Lorentz factor is still large enough to produce a fast jet propagation. On the other hand, in 3D, the beam is prone to strong perturbations which can affect the jet’s head structure. In particular, a simple structure like a terminal Mach shock will probably not survive when significant 3D effects develop. It will be substituted by more complex structures in that case, *e.g.*, by a Mach shock which is no longer normal to the beam flow and which wobbles around the instantaneous flow direction. Another possibility is the generation of oblique shocks near the jet head due to off-axis oscillations of the beam. Both possibilities will cause a less efficient deceleration of the beam

flow at least during some epochs. At longer time scales the growth of 3D perturbations will cause the beam to spread its momentum over a much larger area than that it had initially, which will efficiently reduce the jet advance speed.

References

- Aloy, M.A., Ibáñez, J.M^a., Martí, J.M^a. and Müller, E. (1999a). A high-resolution code for 3D relativistic hydrodynamics. *ApJS*, **122**, pp 151 - 166.
- Aloy, M.A., Gómez, J.L., Martí, J.M^a., Ibáñez, J.M^a. and Müller, E. (1999b). Radio Emission from Three-dimensional Relativistic Hydrodynamic Jets: Observational Evidence of Jet Stratification . *ApJ*, **528**, pp L85 - L88.
- Attridge, J.M., Roberts, D.H., and Wardle, J.F.C. (1999). Radio Jet-Ambient Medium Interactions on Parsec Scales in the Blazar 1055+018. *ApJ*, **518**, pp L87 - L90.
- Biretta, J.A., Zhou, F., and Owen, F.N. (1995). Detection of Proper Motions in the M87 Jet. *ApJ*, **447**, pp 582 - 596.
- Colella, P., and Woodward, P.R. (1984). The Piecewise Parabolic Method (PPM) for Gas-Dynamical Simulations. *JCP*, **54**, pp 174 - 201.
- Donat, R., and Marquina, A. (1996). Capturing-Shock Reflections: An improved flux formula *JCP*, **125**, pp 42 - 58.
- Donat, R., Font, J.A., Ibáñez, J.M^a., and Marquina, A. (1998). A Flux-Split Algorithm Applied to Relativistic Flows. *JCP*, **146**, pp 58 - 81.
- Duncan, G.C., and Hughes, P.A. (1994). Simulations of relativistic extragalactic jets. *ApJ*, **436**, pp L119 - L122.
- Hardee, P.E. (1996). Entrainment and Structure of MHD Jets. Energy transport in radio galaxies and quasars, p 273. P.E. Hardee, A.H. Bridle J.A. Zensus (Editors). ASP Conference Series, Vol. 100.
- Koide, S. (1997). A Two-dimensional Simulation of a Relativistic Jet Bent by an Oblique Magnetic Field. *ApJ*, **478**, pp 66 - 69.
- Koide, S., Nishikawa, K.-I., and Mutel, R.L. (1996). A Two-dimensional Simulation of Relativistic Magnetized Jet. *ApJ*, **463**, pp L71 - L74.
- Komissarov, S.S. (1990). Emission by Relativistic Jets with Boundary Layers. *Sov. Astron. Lett.*, **16(4)**, pp 284 - 287.
- Komissarov, S.S., and Falle, S.A.E.G. (1996). Hot Spots in Relativistic Jets. Energy transport in radio galaxies and quasars; p 327. P.E. Hardee, A.H. Bridle J.A. Zensus (Editors). ASP Conference Series, Vol. 100.
- Komissarov, S.S., and Falle, S.A.E.G. (1998). The large-scale structure of FR-II radio sources. *MNRAS*, **297**, pp 1087 - 1108.
- Laing, R.A. (1996). Brightness and Polarization Structure of Decelerating Relativistic Jets. Energy transport in radio galaxies and quasars, p 241. P.E. Hardee, A.H. Bridle J.A. Zensus (Editors). ASP Conference Series, Vol. 100.
- Martí, J.M^a., Müller, E., Font, J.A., Ibáñez, J.M^a. and Marquina, A. (1997). Morphology and Dynamics of Relativistic Jets. *ApJ*, **479**, pp 151 - 163.
- Nishikawa, K.-I., Koide, S., Sakai, J., Christodoulou, D.M., Sol, H., and Mutel, R.L. (1997). Three-Dimensional Magnetohydrodynamic Simulations of Relativistic Jets Injected along a Magnetic Field. *ApJ*, **483**, pp L45 - L48.
- Nishikawa, K.-I., Koide, S., Sakai, J., Christodoulou, D.M., Sol, H., and Mutel, R.L. (1998). Three-dimensional Magnetohydrodynamic Simulations of Relativistic Jets Injected into an Oblique Magnetic Field. *ApJ*, **498**, pp 166 - 169.
- Norman, M.L. (1996). Structure and Dynamics of the 3D Supersonic Jet . Energy transport in radio galaxies and quasars, p 319. P.E. Hardee, A.H. Bridle J.A. Zensus (Editors). ASP Conference Series, Vol. 100.
- Porter, D.H. and Woodward, P.R. (1994). High-resolution simulations of compressible

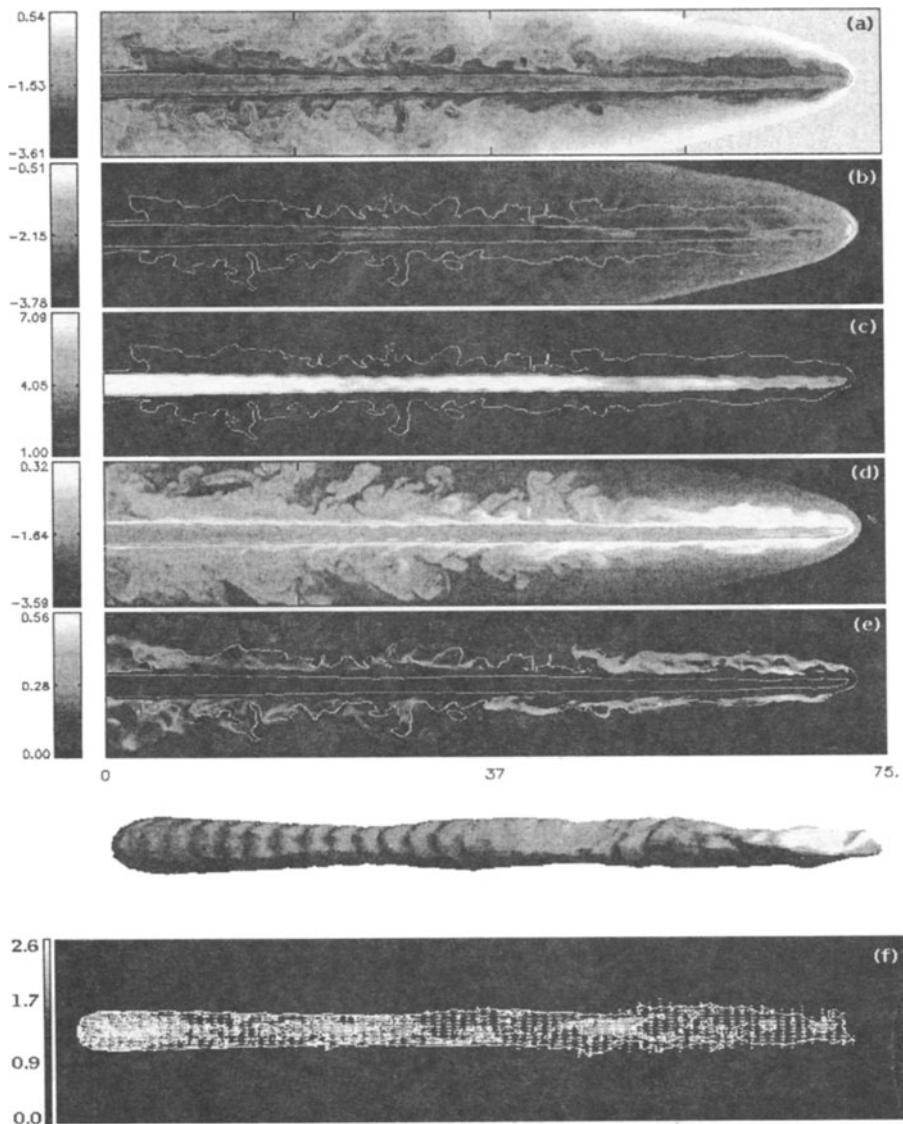


Figure 1. Rest-mass density, pressure, flow Lorentz factor, specific internal energy and backflow velocity distributions (from top to bottom) of the model discussed in the text in the plane $y = 0$ at the end of the simulation. White contour levels appearing in each frame correspond to values of f equal to 0.95 (inner contour; representative of the beam) and 0.05 (representative of the cocoon/shocked external medium interface). The panel under (e) displays the isosurface of $f = 0.95$. Panel (f): Mean local apparent speed observed at an angle of 40 degrees. Arrows show the projected direction and magnitude of the apparent motion the contours corresponding to values of $1.0c$, $1.6c$, and $2.2c$, respectively. Averages have been computed along each line of sight using the emission coefficient as a weight.

- convection using the piecewise-parabolic method. *ApJS*, **93**, pp 309 - 349.
- Shu, C.W., and Osher, S.J. (1989). Efficient Implementation of Essentially Non-Oscillatory shock-capturing schemes. 2. *JCP*, **83**, pp 32 - 78.
- Swain, M.R., Bridle, A.H., and Baum, S.A. (1998). Internal Structure of the Jets in 3C 353. *ApJ*, **507**, pp L29 - L33.
- van Putten, M.H.P.M. (1993). A two-dimensional relativistic ($\Gamma = 3.25$) jet simulation. *ApJ*, **408**, pp L21 - L24.
- Wilson, M.J. (1987). Steady relativistic fluid jets. *MNRAS*, **226**, pp 447 - 454.

RELATIVISTIC JETS FROM COLLAPSARS

M. A. ALOY

*Departamento de Astronomía y Astrofísica,
UVEG, 46100 Burjassot, Spain.*

Emails: Miguel.A.Aloy@uv.es

E. MUELLER

*Max-Planck-Institut für Astrophysik,
85748 Garching, Germany.*

Email: ewald@mpa-garching.mpg.de

J. M. IBANEZ, J. M. MARTI

*Departamento de Astronomía y Astrofísica,
UVEG, 46100 Burjassot, Spain.*

Emails: Jose.M.Ibanez@uv.es, Jose.M.Marti@uv.es

AND

A. MACFADYEN

*Astronomy Department,
University of California, Santa Cruz, CA 95064.
Email:* andrew@ucolick.org

Abstract. We have studied the relativistic beamed outflow proposed to occur in the collapsar model of gamma-ray bursts. A jet forms as a consequence of an assumed energy deposition of $\sim 10^{50} - 10^{51}$ erg/s within a 30° cone around the rotation axis of the progenitor star. The generated jet flow is strongly beamed (\lesssim few degrees) and reaches the surface of the stellar progenitor ($r \approx 3 \cdot 10^{10}$ cm) intact. At break-out the maximum Lorentz factor of the jet flow is about 33. Simulations have been performed with the GENESIS multi-dimensional relativistic hydrodynamic code.

1. Motivation and numerical setup

Various catastrophic collapse events have been proposed to explain the energies released in a gamma-ray burst (GRB) including compact binary

system mergers (Goodman, 1986), (Mochkovitch *et al.*, 1993), collapsars (Woosley, 1993) and hypernovae (Pacyński, 1998). These models all rely on a common engine, namely a stellar mass black hole (BH) which accretes several solar masses of matter from a disk (formed during a merger or by a non-spherical collapse). A fraction of the gravitational binding energy released by accretion is converted into a pair fireball. Provided the baryon load of the fireball is not too large, the baryons are accelerated together with the $e^+ e^-$ pairs to ultra-relativistic speeds (Lorentz factors $> 10^2$; Cavallo and Rees, 1978). The existence of such relativistic flows is supported by radio observations of GRB 980425 (Kulkarni *et al.*, 1998).

The dynamics of spherically symmetric relativistic fireballs has been studied by several authors by means of 1D Lagrangian hydrodynamic simulations –e.g., (Mèzárós, Laguna and Rees, 1993)–. It has been argued that the rapid temporal decay of several GRB afterglows is more consistent with the evolution of a relativistic jet after it slows down and spreads laterally than with a spherical blast wave (Kulkarni *et al.*, 1999a). The lack of a significant radio afterglow in GRB 990123 provides independent evidence for jet-like geometry (Kulkarni *et al.*, 1999b). Motivated by these observations and by the collapsar model of (MacFadyen and Woosley, 1999), we have simulated the propagation of jets from collapsars using relativistic hydrodynamics.

In (MacFadyen and Woosley, 1999) the continued evolution of rotating helium stars, whose iron core collapse does not produce a successful outgoing shock but instead forms a BH surrounded by a compact accretion disk, has been explored. Assuming that the efficiency of energy deposition by $\nu\bar{\nu}$ -annihilation or, e.g., magneto-hydrodynamic processes is higher in the polar regions, (MacFadyen and Woosley, 1999) obtained relativistic jets along the rotation axis, which remained highly focused, and capable of penetrating the star. However, as these simulations were performed with a Newtonian hydrodynamic code, appreciably superluminal speeds in the jet flow were obtained.

We have performed axisymmetric relativistic simulations of jets from collapsars starting from Model 14A of (MacFadyen and Woosley, 1999). The simulations have been performed with GENESIS a multidimensional relativistic hydrodynamic code (based on Godunov-type schemes) developed by (Aloy *et al.*, 1999) using 2D spherical coordinates (r, θ) . GENESIS employs a 3rd order explicit Runge–Kutta method (Shu and Osher, 1989) to advance in time the relativistic Euler equations written in conservation form. High spatial order is provided by a PPM reconstruction (Colella and Woodward, 1984) that sets up the values of the physical variables in order to solve linearized Riemann problems at every cell interface –using the Marquina’s flux formula (Donat *et al.*, 1998)–.

The innermost $2.03 M_{\odot}$ representing the iron core were removed from the helium star model by introducing an inner boundary at a radius of 200 km. When the central BH has acquired a mass of $3.762 M_{\odot}$, we mapped the model to our computational grid. In the r -direction the computational grid consists of 200 zones spaced logarithmically between the inner boundary and the surface of the helium star at $R_* = 2.98 \times 10^{10}$ cm. Assuming equatorial-plane symmetry we use four different zonings in the angular direction: 44, 90 and 180 uniform zones (*i.e.*, 2° , 1° and 0.5° angular resolution), and 100 nonuniform zones covering the polar region $0^\circ \leq \theta \leq 30^\circ$ with 60 equidistant zones (0.5° resolution) and the remaining 40 zones being logarithmically distributed between $30^\circ \leq \theta \leq 90^\circ$.

The gravitational field of the BH is described by the static Schwarzschild metric, neglecting the effects due to self-gravity of the star. We used the EOS of (Witti, Janka and Takahashi, 1994) which includes the contribution of non-relativistic nucleons treated as a mixture of Boltzmann gases, and radiation, as well as an approximate correction due to pairs e^+e^- . Full ionization and non-degeneracy of the electrons is assumed. We advect (*i.e.*, we do not solve additional Riemann problems for each component) nine non-reacting nuclear species which are present in the initial model.

In a consistent collapsar model the jet will be launched by any physical process which gives rise to a local deposition of energy and/or momentum. We mimic this process by depositing energy at a constant rate, \dot{E} , within a 30° cone around the rotation axis of the progenitor star. In radial direction the deposition region extends from the inner boundary to a radius of 6×10^7 cm. We consider two cases that bracket the expected \dot{E} of the collapsar models: 10^{50} erg/s, and 10^{51} erg/s.

2. Results

Low energy deposition rate (Model A). Using a constant $\dot{E} = 10^{50}$ erg/s a relativistic jet forms within a fraction of a second and starts to propagate along the rotation axis (Fig. 1). The jet exhibits all the typical morphological elements (Blandford and Rees, 1974): a terminal bow shock, a narrow cocoon, a contact discontinuity separating stellar and jet matter, and a hot spot. The propagation of the jet is unsteady, because of density inhomogeneities in the star. The Lorentz factor of the jet, W , increases non-monotonically with time, while the density drops to $\sim 10^{-6}$ gr/cm 3 . The density profile shows large variations (up to a factor of 100) due to internal shocks. The mean density in the jet is $\sim 10^{-2} - 1$ g/cm 3 . Some of the internal shocks are biconical and recollimate the beam. These shocks develop during the jet's propagation and may provide the “internal shocks” proposed to explain the observed gamma-ray emission (Katz, 1994).

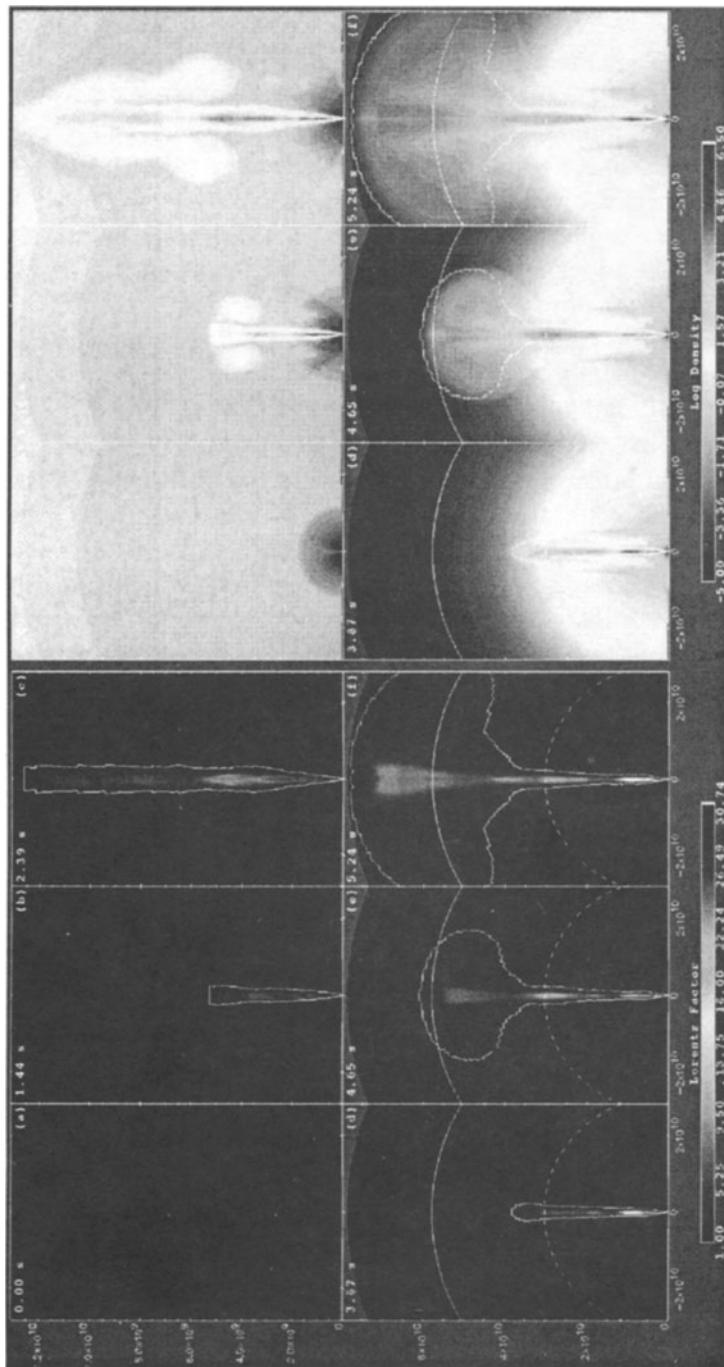


Figure 1. Contour maps of the logarithm of the rest-mass density (six top panels) and the Lorentz factor for model A at different evolution times. Note the change in the scale between left and right panels.

A particularly strong recollimation shock forms during the early stages of the evolution, followed by a strong rarefaction that causes the largest acceleration of the beam material giving rise to a maximum in W . When the jet encounters a region along the axis where the density gradient is positive the jet's head is decelerated, while a central channel in the beam is cleaned by outflow into the cocoon through the head. This leads to an acceleration of the beam. The combination of both effects (deceleration of the head and beam acceleration) increases the strength of the internal shocks.

The relativistic treatment of the hydrodynamics leads to an overall qualitatively similar evolution to (MacFadyen and Woosley, 1999) (formation of a jet), being, however, quantitatively very different. We find that the results strongly depend on the angular resolution, and the minimum acceptable one is 0.5° (at least near the axis). At this resolution we find $W_{\max} \sim 15 - 20$ (at shock break-out) at a radius $\sim 8 \times 10^9$ cm. Within the uncertainties of the jet mass determinations due to finite zoning and the lack of a precise numerical criterion to identify jet matter, the baryon load, η , seems to decrease with increasing resolution. In the highest resolution run we find $\eta \simeq 1.3 \pm 1.2$ at shock break-out (see also Sect. 4).

High energy deposition rate (Model B). Enhancing \dot{E} by a tenfold ($\dot{E} = 10^{51}$ erg/s), the jet flow reaches larger values of W_{\max} . We observe transients during which W_{\max} becomes as large as 40 ($W_{\max} = 33.3$ at shock breakout). The jet propagates faster than in model A. The time required to reach the surface of the star is 2.27 s instead of 3.35 s. The opening angle of the jet at shock breakout is $\sim 10^\circ$, i.e., the jet is less collimated than model A. The strong recollimation shock present in the model A is not so evident here. Instead, several biconical shocks are observed, and W near the head of the jet is larger (~ 22 in the final model) because, due to the larger \dot{E} , the central funnel is evacuated faster, and because the mean density in the jet is 5 times smaller than in model A (η being twice as large).

Evolution after shock breakout. After reaching the stellar surface the relativistic jet propagates through a medium of decreasing density continuously releasing energy into a medium whose pressure is negligible compared to that in the jet cavity, and whose density is (initially) of the same order as that of the jet. These are jump conditions that generate a strong blast wave. The external density gradient determines whether the shock will accelerate or decelerate with time (Shapiro, 1979). In order to satisfy the conditions for accelerating shocks (Shapiro, 1979), we have generated a Gaussian atmosphere matching an external uniform medium. We use models A and B to simulate the evolution after shock breakout. The computational domain is extended for this purpose to a radius of $R_t = 7.6 \times 10^{10}$ cm. The jet reaches R_t (from the stellar surface) after 1.8 s in both models,

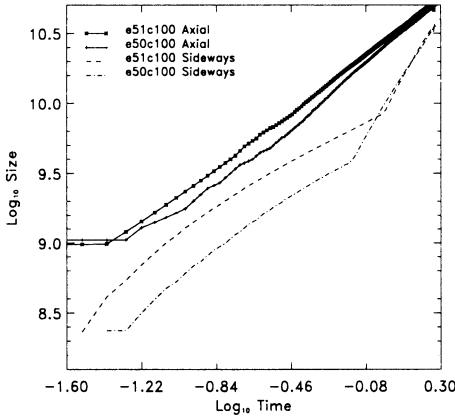


Figure 2. Evolution of the axial and lateral sizes of the jet cavity during the post-breakout epoch. Time is measured with respect to the breakout time for each model.

i.e., the mean propagation velocity is $\sim 0.85c$ (almost three times larger than that inside the star).

The evolution after shock breakout can be distinguished into three epochs (see Figs. 1 and 2), which are related with (i) the external thermodynamical gradients and (ii) the importance of the axial momentum flux relative to the pressure into the jet cavity. Both effects determine the shape of the expanding bubble –prolate– (see Figs. 1 and 2) during the post-breakout evolution. However, when the jet reaches the uniform part of the circumstellar environment, the shape changes appreciably, because the sideways expansion is faster. We have not followed the evolution long enough to see what happens when most of the bubble has reached the uniform part of the environment. Nevertheless, we can infer from Fig. 2 that the widening rate reduces with time in a way similar to what has happened to the axial expansion. At later times most of the bubble is inside the uniform medium, and the bubble will eventually be pressure driven. Hence an isotropic expansion is expected.

After shock breakout there are transients in which W_{\max} becomes almost 50 in some parts of the beam, W_{\max} is again obtained behind the strongest recollimation shock. The Lorentz factor near the boundary of the cavity blown by the jet grows from ~ 1 (at shock breakout) to ~ 3 in both models decreasing with latitude. At the end of the simulation W_{\max} is 29.35 (44.17) for model A (B), which is still smaller than the ones required for the fireball model (Caballo and Rees, 1978). However, our simulations have not been pushed far enough in time yet and, therefore, they can (at the present stage) neither account for the observational properties of GRBs nor

of their afterglows. Instead, our set of numerical models can be regarded as simulations of a proto-GRB, because the scales treated in the simulations are still by more than 100 times smaller than the typical distances at which the fireball eventually becomes optically thin ($\sim 10^{13}$ cm).

References

- Aloy, M.A., Ibáñez, J.M^a., Martí, J.M^a. and Müller, E. (1999). A high-resolution code for 3D relativistic hydrodynamics. *ApJS*, **122**, pp 151 - 166.
- Blandford, R.D., and Rees, M.J (1974). A “twin-exhaust” model for double radio sources. *MNRAS*, **169**, pp 395 - 415.
- Cavollo, G. and Rees, M.J. (1978). A qualitative study of cosmic fireballs and gamma-ray bursts. *MNRAS*, **183**, pp 359 - 365.
- Colella, P., and Woodward, P.R. (1984). The Piecewise Parabolic Method (PPM) for Gas-Dynamical Simulations. *JCP*, **54**, pp 174 - 201.
- Donat, R., Font, J.A., Ibáñez, J.M^a., and Marquina, A. (1998). A Flux-Split Algorithm Applied to Relativistic Flows. *JCP*, **146**, pp 58 - 81.
- Goodman, J. (1986). Are gamma-ray bursts optically thick?. *ApJ*, **308**, pp L47 - L50.
- Katz, J.I. (1994). Two populations and models of gamma-ray bursts. *ApJ*, **422**, pp 248 - 259.
- Kulkarni, S.R., et al. (1998). Radio emission from the unusual SN1998bw and its association with the gamma-ray burst of 25 April 1998. *Nature*, **395**, pp 663 - 669.
- Kulkarni, S.R., et al. (1999a). The afterglow, redshift and extreme energetics of the gamma-ray burst of 23 January 1999. *Nature*, **398**, pp 389 - 394.
- Kulkarni, S.R., et al. (1999b). Discovery of a Radio Flare from GRB 990123. *ApJ*, **522**, pp 97 - 100.
- MacFadyen, A. and Woosley, S.E. (1999). Collapsars - Gamma-Ray Bursts and Explosions in “Failed Supernovae”. *ApJ*, **524**, pp 262 - 289.
- Mézárros, P., Laguna, P., and Rees, M.J. (1993). Gasdynamics of relativistically expanding gamma-ray burst sources. *ApJ*, **415**, pp 181 - 190.
- Mochkovitch, R., Hernanz, M., Isern, J., and Martin, X. (1993). GRBs as collimated jets from neutron star/black hole mergers. *Nature*, **361**, pp 236 - 238.
- Pacyński, B. (1998). Are Gamma-Ray Bursts in Star-Forming Regions?. *ApJ*, **494**, pp L45 - L48.
- Shapiro, P.R. (1979). Relativistic blast waves in 2D. *ApJ*, **233**, pp 831 - 850.
- Shu, C.W., and Osher, S.J. (1989). Efficient Implementation of Essentially Non-Oscillatory shock-capturing schemes. 2. *JCP*, **83**, pp 32 - 78.
- Witti, J., Janka, H.-T., and Takahashi, K. (1994). Nucleosynthesis in neutrino-driven winds from protoneutron stars I. The alpha-process. *A&A*, **286**, pp 841 - 856.
- Woosley, S.E. (1993). Gamma-Ray burst from stellar mass accretion disks around black holes. *ApJ*, **405**, pp 273 - 277.

EXACT COMPUTATION IN NUMERICAL LINEAR ALGEBRA: THE DISCRETE FOURIER TRANSFORM

J.A.D.W. ANDERSON

*Department of Computer Science,
The University of Reading,
Reading, England, RG6 6AY
Email: J.A.D.W.Anderson@rdg.ac.uk*

AND

P.K. SWEBY

*Department of Mathematics, The University of Reading,
Reading, England, RG6 6AX
Email: P.K.Sweby@rdg.ac.uk*

Abstract.

An encoding of the rational rotations is introduced which allows all rational affine transformations to be computed exactly. This encoding is used in combination with transrational numbers to give total trigonometric functions. It is shown how these functions can be used to compute indefinitely many digits of the Discrete Fourier Transform.

1. Introduction

Traditionally numerical computation has been performed in floating point arithmetic, and is therefore prone to round-off error. Whilst many processes in CFD involve approximation in the discretisation of derivatives, there are several situations where a more precise computation would be beneficial, such as in the inversion of the large systems found in implicit calculations, and determining precisely in which cell a point lies in the computational grid. Rational arithmetic is one possible answer. Although it is still in its infancy we demonstrate its potential here on a simple problem involving the Discrete Fourier Transform (DFT).

Rational arithmetic is often implemented by having separate, variable-length bit strings represent the numerator and denominator of a rational number. In this scheme the result of a rational arithmetical operator is generally as long as the sum of the bit lengths of its arguments. This leads to linear growth in memory use and less than squared processing time on a sequential processor (Schönhage, 1986). Regardless of these costs, rational arithmetic has the very considerable advantage that it does not involve round-off error. This is generally useful in numerical algorithms and introduces the possibility of computing some numerical results exactly. Briefly, irrational results cannot be computed exactly in finite time, but rational results can be - provided there is enough computer memory and processing time available. By contrast, floating point numbers have a constant memory requirement and compute quickly, but they are equivalent to rational numbers of the form $a/2^n$ and so, in general, cannot support the exact computation achievable by rational numbers a/b with arbitrary b . Thus there are considerable advantages to be had by using rational, rather than floating point, arithmetic. Now, with the imminent arrival of Very Long Instruction Word (VLIW) computer processors, from both Sun Microsystems (MAJC, 1999) and a consortium of Intel and Hewlett-Packard (Chipanalyst, 1999), the practical role of rational computation is set to increase.

We have used rational arithmetic in the implementation of: the matrix inverse using both Gauss-Seidel iteration and LU Decomposition (Golub and van Loan, 1999; Jennings and McKeown, 1992); exact calculation of the rank of a matrix; generalized matrix inverses for rank deficient, including rectangular matrices (Ben-Israel and Greville, 1974; Nashed, 1976); and an iterative matrix inverse using the conjugate gradient method (Golub and van Loan, 1999; Jennings and McKeown, 1992). In each case, the implementation involves only elementary matrix transformations which are carried out exactly in rational arithmetic.

These rational algorithms are insensitive to ill-conditioning in the sense that the exact result is given for arbitrary data. But they are, of course, sensitive to ill-conditioning in the sense that alternative sets of close, but false, data might lead to very disparate results. Thus mathematical analysis of ill-conditioning need only be undertaken where it is intended to truncate a rational result, introduce a rational approximation of an irrational system, or where false data are expected.

Whilst exact arithmetic should be expected to lead to the insensitivity to ill-conditioning just described, it has a far more profound consequence: all convergent solutions in numerical analysis can be computed exactly. This is an immediate consequence of Turing's theory of computable numbers (Turing, 1936; Turing, 1937). According to this view of computability a number is computable if a Turing machine can generate indefinitely many

digits of the exact solution, each digit being generated in finite time. We show here how rational arithmetic can be used to compute indefinitely many digits of the DFT. In the longer version of this paper (Anderson and Sweby, 1999), this result is extended to the eigensystem of a symmetric matrix. As a very welcome side effect, we find that the rational approximation to the, generally, transcendental DFT, in the example computed, is far more accurate than the floating point approximation to the DFT computed to the same precision.

In the longer version of this paper (Anderson and Sweby, 1999) we define *transrational* numbers – rational numbers augmented by fractions with a zero denominator – and we show that transrational arithmetic is consistent with rational trigonometry. Transrational numbers are used in the total rational trigonometric functions: $\cos q$, $\sin q$, $\tan q$, $\sec q$, $\csc q$, $\cot q$; their inverses $\arccos q$, $\arcsin q$, $\arctan q$, $\operatorname{arcsec} q$, $\operatorname{arccsc} q$, $\operatorname{arccot} q$; and the rational exponential of a complex variable $\exp q$. The transrational trigonometric functions have the advantage that they give rise to exactly orthogonal and unitary numerical transformations. They are used to compute exactly orthogonal eigenvectors from a symmetric matrix. A number of practical applications of transrational arithmetic are discussed in the paper (Anderson and Sweby, 1999), but we report here only the computation of the DFT.

2. Fourier Transform

The Fourier Transform, and its discrete approximation the DFT, are unitary transformations. The DFT can be approximated naively by a truncated rational approximation or, more interestingly, the unitary transformation can be decomposed into a product of elementary unitary transformations (Gourlay and Watson, 1973), each of which is approximated arbitrarily closely by a rational elementary unitary transformation. The result is then a transformation which is exactly unitary, with inverse given exactly by the conjugate transpose, and which may be taken arbitrarily close to the Fourier Transform.

For simplicity consider, D , the smallest, non-trivial DFT, that is, the DFT in dimension three. This may be written as an active, pre-multiplying transformation of co-ordinates (Crownover, 1995) with $i = \sqrt{-1}$ and $\zeta = \exp(-2\pi i/3)$,

$$D = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 & 1 \\ 1 & \zeta & \zeta^2 \\ 1 & \zeta^2 & \zeta^4 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{2\sqrt{3}} - \frac{i}{2} & -\frac{1}{2\sqrt{3}} + \frac{i}{2} \\ \frac{1}{\sqrt{3}} & -\frac{1}{2\sqrt{3}} + \frac{i}{2} & -\frac{1}{2\sqrt{3}} - \frac{i}{2} \end{bmatrix}.$$

We define

$$\expq(ia) = \cosq(a) + i \sinq(a). \quad (1)$$

Substituting \expq for \exp in ζ gives a rational approximation Q to D when a rational a is chosen that gives a rotation close to a rotation of $2\pi/3$ radians.

It is advantageous to choose the rational parameter $a = n/d$ so that it gives a rotation which is excessively close to the true rotation. Consider the sequence

$$f\left(\frac{n_1}{1}\right), f\left(\frac{n_2}{2}\right), f\left(\frac{n_3}{3}\right), \dots, f\left(\frac{n_i}{i}\right),$$

where each $f(n_i/i)$ has no closer approximation, with the given denominator i , to some given $g(x)$, $x \in \mathbb{R}$. We call the least n_i/i the *first excessively close approximation* if and only if $f(n_i/i)$ is closer to the true value $g(x)$ than some $f(n_j/j)$ with $i < j$. The least n_j/j that is closer to the true value than the first excessively close approximation, $f(n_i/i)$, is called the *second excessively close approximation*, and so on. We have observed that the excessively close approximations to rotations which are a fraction of a full rotation spread out quickly, so that there is considerable advantage in using the excessively close approximations - because they give an accuracy considerably in excess of their precision.¹

There is some advantage in the above naive rational approximation to the DFT in that it is more accurate than its precision would suggest. But as each component of Q is computed independently, the components of Q are a truncation of the irrational components of D , and therefore Q is *not* unitary. A considerably better approximation results by decomposing the unitary transformation D into a sequence of elementary unitary transformations. This can be done by a modified Givens' orthogonalisation operating symbolically, as given below, or numerically on a rational truncated-series approximation to the DFT.²

Given D we find R_4 an elementary unitary rotoreflection and R_3, R_2, R_1 elementary unitary rotations, such that $R_4 R_3 R_2 R_1 D = I$. Then $R = R_4 R_3 R_2 R_1$ is D^{-1} , the Fourier synthesis, and the conjugate transpose R^* is D , the Fourier analysis. Introducing rational approximations Q_i to R_i then gives a rational unitary approximation Q to D . Note that R_4 transforms

¹We have implemented recurrence formulae that appear to generate in sequence all and only the excessively close approximations to some fractional rotations, but no generalisation of these formulae is immediately obvious. We would be extremely grateful to hear of any algorithm that is known to work for all fractional rotations.

²If a decomposition of the Fourier transform into specific elementary unitary transformations is known, then we would be very grateful to hear of it.

a pair of complex-conjugate elements on the major diagonal to real values and multiplies out $\det(D) = i$ by $-i$, giving $\det(RD) = \det(I) = 1$, as required. Specifically,

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & i & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{i\alpha_3} \cos \theta_3 & e^{i\beta_3} \sin \theta_3 \\ 0 & -e^{-i\beta_3} \sin \theta_3 & e^{-i\alpha_3} \cos \theta_3 \end{bmatrix} \times \\ \begin{bmatrix} e^{i\alpha_2} \cos \theta_2 & 0 & e^{i\beta_2} \sin \theta_2 \\ 0 & 1 & 0 \\ -e^{-i\beta_2} \sin \theta_2 & 0 & e^{-i\alpha_2} \cos \theta_2 \end{bmatrix} \begin{bmatrix} e^{i\alpha_1} \cos \theta_1 & e^{i\beta_1} \sin \theta_1 & 0 \\ -e^{-i\beta_1} \sin \theta_1 & e^{-i\alpha_1} \cos \theta_1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

and the Q_i are given by substituting the excessively accurate approximations to the Fourier components given in Table 1 into the corresponding R_i . The result is an exactly unitary Q that is more accurate than its precision suggests.

Substitute	R_1	R_2	R_3
$\cos \alpha$	1	1	$\text{cosq}(3329/2340)$
$\sin \alpha$	0	0	$\text{sinq}(3329/2340)$
$\cos \beta$	1	1	$\text{sinq}(-3329/2340)$
$\sin \beta$	0	0	$\text{cosq}(-3329/2340)$
$\cos \theta$	$\text{cosq}(985/2378)$	$\text{cosq}(629/1979)$	$\text{cosq}(-985/2378)$
$\sin \theta$	$\text{sinq}(985/2378)$	$\text{sinq}(629/1979)$	$\text{sinq}(-985/2378)$

TABLE 1. Excessively Accurate Fourier Components

The trigonometric values given in Table 1 have denominators of at most 23 bits, giving them a precision no better than a standard 32 bit floating point number with 23 bit mantissa. A standard 32 bit floating point approximation F to D was compared to Q as follows. Here P is a block-diagonal permutation matrix with unity in the first element of the major diagonal and unity on the antidiagonal of the remaining block, if any. Then, for all natural numbered dimensions n : $D_n^2 = P_n$, $D_n^4 = I_n$, and so on cyclically,

$$\|FF^* - I\|_2 / \|QQ^* - I\|_2 \approx 1.286 \times 10^{-6} / 0 = \infty.$$

In other words, Q , being exactly unitary, is infinitely closer to a unitary matrix than the 23 bit floating point approximation F . Further Q is, in

general, infinitely closer to unitary than any floating point approximation with a finite number of digits. This result holds regardless of how close Q is to Fourier. But we now see that Q is considerably closer to Fourier than F at the same, 23 bit, precision

$$\|F^2 - P\|_2 / \|Q^2 - P\|_2 \approx (2.262 \times 10^{-6}) / (4.417 \times 10^{-7}) \approx 5,$$

and the relative advantage of the rational approximation continues

$$\|F^4 - I\|_2 / \|Q^4 - I\|_2 \approx (4.338 \times 10^{-6}) / (4.773 \times 10^{-7}) \approx 9.$$

We expect that both higher precision and larger matrices will increase the accuracy of the rational approximations to the Fourier transform as compared to floating point approximations, in addition to the advantage that the rational approximations are exactly unitary and, in general, floating point approximations are not.

In addition to the practical utility of the rational approximation to the DFT there is a theoretical advantage to be gained. First, the computation of an exactly unitary approximation Q to the DFT of arbitrary dimension may be taken to any desired accuracy by using rational arithmetic. Second, we have a numerical measure $\epsilon = \|Q^2 - P\|_2$ of the departure of Q from Fourier. We may then arrange a computation to record in \hat{Q} the elements of Q taken to as many digits as have place value greater than ϵ . These digits are certainly correct. The computation may be repeated at higher precision *ad infinitum*, thus recording successive correct digits of Q without limit. Which is to say that the computation gives the DFT as a set \hat{Q} of computable numbers (Turing, 1936; Turing, 1937). Further, though this is of no practical interest, evaluating the denumerable infinity of digits gives the exact solution for both rational and irrational components of Q . (A machine which produced the first digit in 1/2 unit of time, the second in 1/4 unit of time, the third in 1/8 unit of time, and so on, would enumerate all of the digits after unit time.) Which is to say that this machine would compute the DFT exactly.

The existence of a measure ϵ which gives just the significant digits of a computation is valuable, because the rational solutions given by an algorithm may be truncated at this precision, if desired. Naturally, one would prefer to use a sharp bound on the significance of digits, so the mathematical derivation of sharp bounds in numerical analysis is extremely important when algorithms are redesigned for rational arithmetic.

The manoeuvre of re-running the entire computation just to deliver subsequent correct digits is somewhat artificial and is unnecessary in naturally iterative algorithms such as the computation of the eigensystem (Anderson and Sweby, 1999).

3. Discussion

In a longer version of this paper (Anderson and Sweby, 1999) we introduce transrational arithmetic as a faster method of implementing rational arithmetic on computer processors and use it to define total, rational, trigonometric functions. These are used to compute exactly orthogonal eigenvectors from a symmetric matrix and, as reported here, to compute the Discrete Fourier Transform as an exactly unitary transformation. Some scientific and engineering applications of transrational arithmetic are also discussed in (Anderson and Sweby, 1999).

References

- MAJC (1999). Microprocessor Architecture for Java Computing. <http://www.sun.com/microelectronics/MAJC>
- Chipanalyst (1999). The Merced/P7 processor. <http://www.chipanalyst.com>. Since the initial submission of this paper the trade name "Intellium" has been associated with this chip.
- Anderson J A D W and Sweby P K (1999). Exact computation in numerical linear algebra: The discrete fourier transform and symmetric eigensystem. Technical Report RUCS/1999/TR/010/A, Department of Computer Science, The University of Reading, England, RG6 6AY.
- Ben-Israel A and Greville T N E (1974). Generalized Inverses: Theory and Applications Wiley, New York, U.S.A.
- Crownover R M (1995). Introduction to Fractals and Chaos. Jones and Bartlett, Boston, U.S.A.
- Golub G H and van Loan C F (1986). Matrix Computations. North Oxford Academic, Oxford, England.
- Gourlay A R and Watson G A (1973). Computational Methods for Matrix Eigenproblems. John Wiley and Sons, London, England.
- Jennings A and McKeown J J (1992). Matrix Computation. Wiley, New York, U.S.A., second edition.
- Nashed M Z (1976). Generalized Inverses and Applications. Academic Press, New York, U.S.A.
- Schönhage A (1986). Equation solving in terms of complexity. In *Proc. of the International Congress of Mathematicians*, pp 131–153, Berkeley, California, U.S.A.
- Turing A M (1936). On computable numbers, with an application to the entscheidungsproblem. *Proceedings of the London Mathematical Society*, **42**, pp 230–265. Contains an error, corrected ibid. **43**, pp. 544-546.
- Turing A M (1937). On computable numbers, with an application to the entscheidungsproblem. a correction. *Proceedings of the London Mathematical Society*, **43**, pp 544–546, (1937). Correction of ibid. **42**, pp. 230-265.

COMPARATIVE STUDY OF HLL, HLLC AND HYBRID RIEMANN SOLVERS IN UNSTEADY COMPRESSIBLE FLOWS

A. BAGABIR AND D. DRIKAKIS

*Queen Mary and Westfield College,
University of London,
Department of Engineering,
London, E1 4NS, U.K.
Email: d.drikakis@qmw.ac.uk*

Abstract. In this paper we investigate the accuracy and efficiency of the HLL, HLLC and hybrid CBM-FVS (Characteristic-Based Method and Flux Vector Splitting) Riemann solvers for unsteady inviscid flows featuring shock-diffraction, multiple shock interactions and Richtmyer-Meshkov instability. The solvers have been used in conjunction with MUSCL interpolation for obtaining high resolution and a second-order accurate implicit-unfactored method. The latter provides time accurate solutions through Newton sub-iterations and Gauss-Seidel relaxation.

1. Introduction

Although many Riemann solvers and implicit schemes appear to provide high accuracy and efficiency in simple one-dimensional cases, one may find that this performance cannot be retained when the methods are applied in complex unsteady compressible flows which contain multiple shock reflections and/or diffractions as well as strong vorticity generation and flow instabilities. Most of the numerical investigations regarding the implementation of Riemann solvers in conjunction with implicit methods have focused primarily on steady flow problems, and more recently on unsteady compressible flows (Zóltak and Drikakis, 1998). Challenging problems for assessing the accuracy and efficiency of Riemann solvers are flows which contain complex gasdynamic phenomena such as shock-diffraction and multiple shock interactions as well as flow instabilities e.g. Richtmyer-Meshkov (RM) instability.

The aim of the present study is to assess in unsteady compressible flows three different Riemann solvers in conjunction with an implicit-unfactored method. The Riemann solvers are the HLL (Harten et al., 1983), Toro's HLLC (Toro et al., 1994), and the hybrid CBM-FVS (Zóltak and Drikakis, 1998; Eberle, 1987). The CBM-FVS has already been validated for unsteady shock diffraction around a cylinder (Zóltak and Drikakis, 1998). The validation has been performed against other numerical results obtained by unstructured-adaptive grid schemes and second-order Godunov-type discretisation (Drikakis et al., 1999). Results are presented here for shock diffraction around a cylinder and a wedge, as well as for cylindrical explosion in an enclosure featuring RM instability.

2. Numerical modelling

In the present study the governing equations are the two-dimensional Euler equations for a compressible fluid employed in curvilinear co-ordinates in conjunction with the ideal gas equation of state. The computational code which forms the basis for the present work is a finite volume program (Zóltak and Drikakis, 1998) in which the HLL (Harten et al., 1983), HLLC (Toro et al., 1994), and CBM-FVS (Zóltak and Drikakis, 1998; Eberle, 1987) Riemann solvers have been implemented. The CBM-FVS combines a Characteristic-Based Method (Zóltak and Drikakis, 1998; Eberle, 1987) and a modified Flux Vector Splitting scheme (Drikakis and Tsangaris, 1993). The latter provides additional numerical dissipation in hypersonic flows ($M > 3$). The CBM-FVS has also been used more recently in studying the physics of compressible flows with instabilities (Bagabir and Drikakis, 1999).

In the case of the HLL and HLLC solvers the acoustic wave speeds are calculated as proposed in (Einfeldt, 1988), while the contact wave speed for HLLC has been taken from (Batten et al., 1997). To provide high-resolution at the cell faces of the computational volumes, an improved version of the MUSCL interpolation (Thomas et al., 1985) has been employed. The Riemann solvers discussed above have been implemented in conjunction with second-order in time implicit-unfactored method (Zóltak and Drikakis, 1998), which employs Newton subiterations and point Gauss-Seidel relaxation. The Jacobian matrices arising from the implicit discretisation remain the same for all Riemann solvers employed here.

3. Results and discussion

3.1. SHOCK DIFFRACTION AROUND A CYLINDER

The first test case considered in this work is the unsteady shock wave diffraction around a cylinder at incident Mach number $M_S = 2.81$. The same problem has been studied in the past using the CBM-FVS solver (Zóltak and Drikakis, 1998) as well as using unstructured adaptive-grid methods (Drikakis et al., 1999; Ofengeim and Drikakis, 1997). Following (Zóltak and Drikakis, 1998), the computations were performed on a grid 241×161 . Ahead of the shock ambient conditions corresponding to $U_R = (\rho, u, w, e)^T = (1, 0, 0, 2.5)^T$ are employed. The states behind the shock are obtained by the Rankine-Hugoniot conditions. In Fig. 1, the iso-density contours at $t = 2$ are compared with the adaptive-grid results of (Drikakis et al., 1999). The adaptive-grid solution provides better resolution in shock waves and contact discontinuities (see also (Drikakis et al., 1999)), but similar flow patterns have also been obtained using a structured grid without any adaptation. As seen from Fig. 1, HLL, HLLC and CBM-FVS solvers give similar results. All schemes required about 200 time steps using $CFL = 2$.

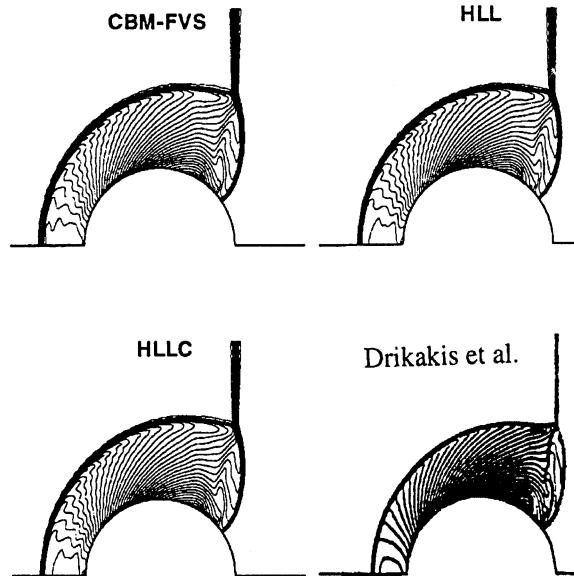


Figure 1. Iso-density contours ($t = 2$) for shock diffraction ($M_S = 2.81$) over a cylinder using the CBM-FVS, HLL, and HLLC solvers; Comparison with the results of (Drikakis et al., 1999).

3.2. SHOCK DIFFRACTION OVER A WEDGE

The second test case is the shock-diffraction ($M_S = 1.2$) over a wedge with angle 30° . The computations were performed on a curvilinear mesh with 421×241 points, similar to the one used by Toro (Toro, 1999). Comparison of the results obtained by the three Riemann solvers are shown in Fig. 2. These results correspond to single Mach reflection in which the three shocks, namely the incident, reflected and Mach stem, meet at the triple point. From the triple point a slip surface emerges that joins the wedge surface at a sharp angle. Although the HLL scheme considers a single intermediate region, its present implicit HLL version provides almost similar resolution for the slip surface, with the one obtained by the CBM-FVS and HLLC. This can be partly attributed to the use of the modified MUSCL scheme employed in this study. All schemes required about 200 time steps using $CFL = 5$.

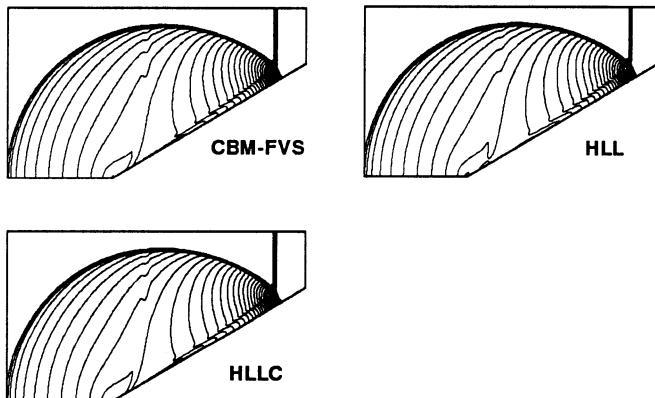


Figure 2. Iso-density contours for the shock-wedge diffraction problem ($M_S = 1.2$).

3.3. CYLINDRICAL EXPLOSION

The third case is the propagation of a blast wave in a square enclosure (Fig. 3). The initial conditions used in this study are similar to the ones used in (Bagabir and Drikakis, 1999) (see Fig. 3). There is a very hot region at the centre of the explosion with very low density and finite pressure, which slowly expands as time goes on. The computations were performed on a Cartesian mesh with 300×300 grid points. The flow should maintain quarter plane symmetry as the initial and boundary conditions are symmetric around the diagonals of the square. The interesting flow feature in this case is the development of the RM instability occurring after the cylindrical shock has reflected from the walls and interacted with the unstable low

density region around the centre of the enclosure (Bagabir and Drikakis, 1999). In the numerical simulations the instability appears as a symmetry-breaking i.e. asymmetric flow patterns. If, however, the numerical scheme exhibits large dissipation errors the instability will be suppressed, thus, obtaining a symmetric flow field.

Figure (4) shows the iso-Mach contours at different time instants for the case of $M_S = 2$. The time $t = 0$ ms corresponds to the time instant where the cylindrical blast comes for first time in contact with the walls. The first frame in Fig. (4) shows the shock segments reflecting from the walls of the enclosure moving towards the centre. The following two frames show the shock segments move back again towards the walls. The fourth and fifth frames show the development of four shock waves emerging from the middle of the walls and, subsequently, interacting at the centre of the enclosure. The RM instability is due to the multiple interactions with the low density region around the centre of the enclosure and this starts appearing more clearly at time $t = 4.5$ ms. Then, the flow field is excited by four corner shock waves (at $t = 6$ ms). As can also be seen from the last frame of Fig. 4, the flow becomes more asymmetric as time goes on (for more details see (Bagabir and Drikakis, 1999; Bagabir, 2000)). The CBM-FVS and HLLC capture the shock waves better than the HLL Riemann solver. Most importantly, it was found that the HLL scheme results in no symmetry-breaking even if the simulation is performed for long time intervals and on finer grids. The above behaviour is due to the dissipative properties of the HLL scheme. On the other hand, both the HLLC and CBM-FVS capture the RM instability and, additionally, provide grid-independent solutions for the wall pressure distributions without requiring a finer grid than the one employed here. The wall pressure distributions are of special interest in the case of explosions and fluid-structure interaction. All schemes required about 2200 time steps using $CFL = 10$.

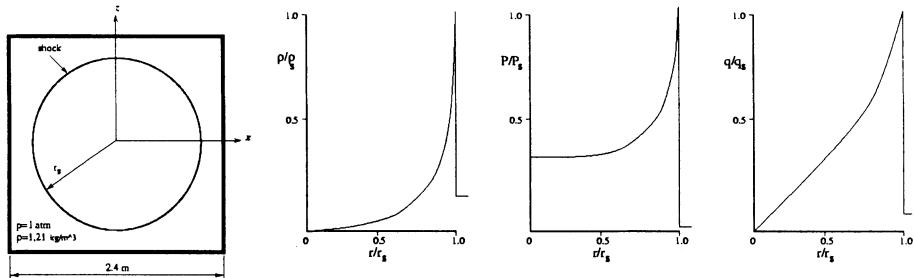


Figure 3. Problem configuration and initial condition profiles.

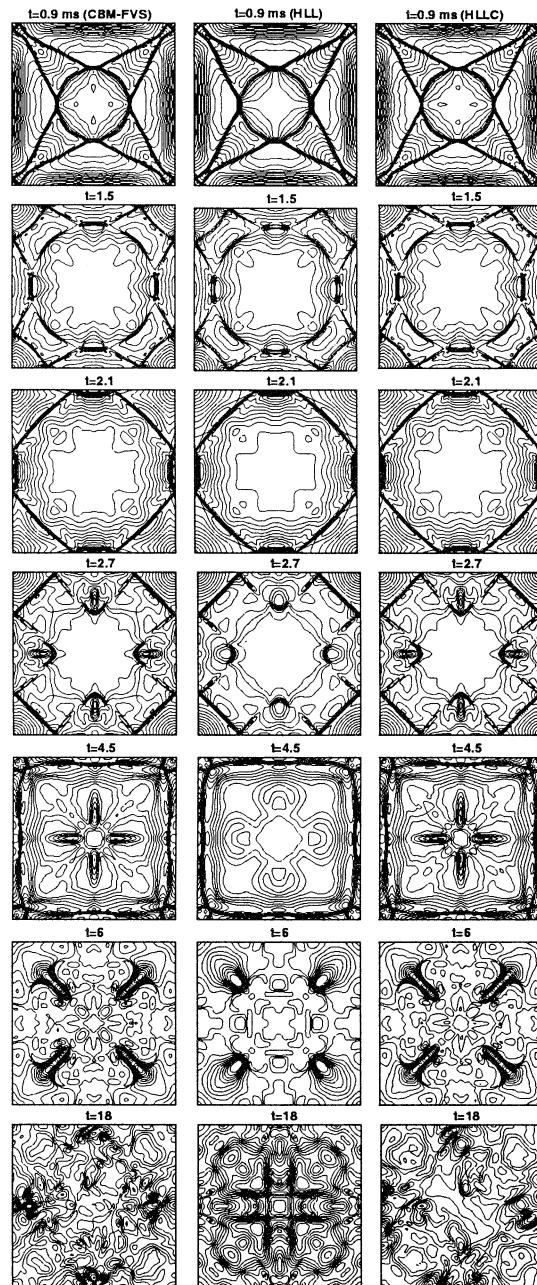


Figure 4. Iso-Mach contours for $M_S = 2$ as predicted by CBM-FVS (1^{st} column), HLL (2^{nd} column), and HLLC (3^{rd} column).

4. Conclusions

The accuracy and efficiency of an implicit-unfactored version of the HLL, HLLC and CBM-FVS Riemann solvers were investigated in three unsteady compressible flows, featuring shock-reflection, shock-diffraction, as well as RM instability. The CBM-FVS and HLLC were found to provide overall better accuracy than the HLL. The latter was not able to capture the development of the RM instability in the case of cylindrical explosion in an enclosure. Further numerical experiments (Bagabir, 2000) have shown that the original Roe's Riemann solver provides results similar to the CBM-FVS and HLLC methods, regarding the RM instability. On the other hand, it has been found (Bagabir, 2000) that the FVS scheme suppresses the instability development. The same was also found here for the HLL scheme. All schemes required the same number of time steps and almost similar number of work units.

References

- Bagabir A and Drikakis D (1999). On the Richtmyer-Meshkov instability during a blast wave propagation in an enclosure. *22nd Int. Symp. on Shock Waves*, London, UK.
- Bagabir A (2000). On the accuracy and efficiency of Godunov-type methods in various compressible flows. PhD dissertation, Queen Mary, University of London.
- Batten P, Lambert C, Causon D and Clarke N(1997). On the choice of wavespeeds for HLLC Riemann solvers. *SIAM J. Scientific Comput.*, **18**, pp 1553-1570.
- Drikakis D, Ofengeim D, Timofeev D and Voinovich P (1999). Computation of non-stationary shock-wave/cylinder interaction using adaptive-grid methods. *Journal of Fluids and Structures*, **11**, pp 665-691.
- Drikakis D and Tsangaris S (1993). On the solution of the compressible Navier-Stokes equations using improved flux vector splitting methods. *Applied Mathematical Modelling*, **17**, pp 282-297.
- Eberle A (1987). Characteristic Flux Averaging Approach to the Solution of Euler's Equations. VKI Lecture Series, Computational Fluid Dynamics, 1987-04.
- Einfeldt B (1988). On Godunov-type methods for gas dynamics. *SIAM J. Numer. Anal.*, **25**, pp 294-318.
- Harten A, Lax P D, and Van Leer B (1983). On upwind differencing and Godunov-Type schemes for hyperbolic conservation laws. *SIAM Review*, **25**, pp 35-61.
- Ofengeim D and Drikakis D (1997). Simulation of blast wave propagation over a cylinder. *Shock Waves Journal*, **7**, pp 305-317.
- Thomas J L, van Leer B and Walters R W (1985). Implicit flux split scheme for the Euler equations. AIAA-paper 85-1680.
- Toro E F (1999). Riemann Solvers and Numerical Methods for Fluid Dynamics. Second Edition, Springer-Verlag.
- Toro E T, Spruce M and Speares W (1994). Restoration of the contact surface in the HLL Riemann solver. *Shock Waves*, **4**, pp 25-34.
- Zóltak J and Drikakis D (1998). Hybrid upwind methods for the simulation of unsteady shock-wave diffraction over a cylinder. *Comput. Meth. in Appl. Mech. and Engrg.*, **162**, pp 165-185.

A NEW RECONSTRUCTION TECHNIQUE FOR THE EULER EQUATIONS OF GAS DYNAMICS WITH SOURCE TERMS

P. BARTSCH

Powertrain Systems,

Passenger Cars,

AVL List GmbH,

Hans-List-Platz 1,

A-8020 Graz, Austria.

Email: peter.bartsch@avl.com

AND

A. BORZI'

Institut für Mathematik,

Karl-Franzens-Universität Graz,

Heinrichstr. 36,

A-8010 Graz, Austria.

Email: alfio.borzi@kfunigraz.ac.at

Abstract. We present a new reconstruction technique for the Godunov scheme that accurately solves steady state gas dynamics problems in the presence of forcing terms. This method introduces modified conserved variables that can be used in any high-order reconstruction process and which contain information about the source terms. By using these variables, modified slopes are obtained and a robust and accurate reconstruction procedure is formulated. Applications are presented to show the accuracy and robustness of the method.

1. Introduction

We present a new reconstruction technique for the Godunov scheme that solves the quasi-one-dimensional equations of gas dynamics in a duct of variable cross section with thermal sources and wall friction. Our purpose is to obtain accurate non-oscillatory steady state solutions of these equations.

The Euler equations of gas dynamics can be represented by,

$$\frac{\partial}{\partial t}q + \frac{\partial}{\partial x}f(q) = -g(x, q), \quad (1)$$

where $q = (\rho, \rho u, e)$ is the set of conserved variables, $f : R^3 \rightarrow R^3$ is the flux function, and $g : R \times R^3 \rightarrow R^3$ is the source term.

In many applications it is important to compute steady states of (1) in which the non-zero flux gradient balance the source terms as,

$$A(q) \frac{\partial}{\partial x}q = -g(x, q), \quad (2)$$

where $A(q) \equiv \frac{\partial f}{\partial q}$ is the flux Jacobian. In such situations, splitting schemes may have difficulties to capture and maintain such states; see, e.g., LeVeque (LeVeque, 1998) and Toro (Toro, 1999). However, it is long recognized that a way to approach this problem is to try to improve the reconstruction technique; see (Liu, 1979) (Ben-Artzi and Falcovitz, 1984) (Glaz and Liu, 1984) (Glimm, Marshall, and Plohr, 1984) (Ben-Artzi and Falcovitz, 1995) (LeVeque, 1998) (LeVeque and Bale, 1998) (Bermúdez and Vázquez-Cendón, 1994) (Greenberg and LeRoux , 1996).

The new reconstruction method proposed in this paper is obtained as a modification of existing high-order reconstruction techniques and is based on the vector function,

$$\tilde{q}(x) := q(x) + \int_{x_0}^x A^{-1}(q(y))g(y, q)dy, \quad (3)$$

so that, whenever q satisfies (2), we have $A(q) \frac{\partial}{\partial x} \tilde{q} = 0$.

Therefore, we can apply accurate reconstruction methods to \tilde{q} that are effective for solving problems with zero source terms. Then, using (3) we are able to recover high-order conservative and non-oscillatory slopes for the original variables. This method is easy to implement and can be used in combination with any high-order reconstruction technique.

2. The Euler Equations of Gas Dynamics with Source Terms

Consider the Euler equations of gas dynamics with source terms given as follows:

$$\frac{\partial}{\partial t} \begin{pmatrix} \rho \\ \rho u \\ e \end{pmatrix} + \frac{\partial}{\partial x} \begin{pmatrix} \rho u \\ \rho u^2 + p \\ u(e + p) \end{pmatrix} = -\frac{1}{C} \frac{dC}{dx} \begin{pmatrix} \rho u \\ \rho u^2 \\ u(e + p) \end{pmatrix} + \frac{1}{V} \begin{pmatrix} 0 \\ -F_R \\ Q_Z \end{pmatrix}, \quad (4)$$

where ρ is the density, u is the velocity, ρu is the momentum, e is the total energy, and p is the pressure. The function F_R represents the wall friction and Q_Z the heat contribution to the energy. The effect of a variable cross section C are taken into account by the second term on the right-hand side of the vector equation (4) given above.

We also need a thermodynamic equation of state which relates the pressure p to the density ρ and the temperature T ,

$$p = \rho R T,$$

where R is the specific gas constant. The total energy is given by

$$e = \rho c_v T + \rho u^2 / 2,$$

where c_v is the specific heat at constant volume. It is convenient to write equation (4) in the quasi-linear form,

$$\frac{\partial}{\partial t} q + A(q) \frac{\partial}{\partial x} q = -g(x, q), \quad (5)$$

where $A(q)$ is the flux Jacobian defined by,

$$A(q) \equiv \frac{\partial f}{\partial q} = \begin{bmatrix} 0 & 1 & 0 \\ \frac{k-3}{2}u^2 & -(k-3)u & k-1 \\ \frac{k-2}{2}u^3 - \frac{uc^2}{k-1} & \frac{c^2}{k-1} - \frac{2k-3}{2}u^2 & ku \end{bmatrix}, \quad (6)$$

where $k = c_p/c_v$ is the ratio of specific heats and $c = (kp/\rho)^{1/2}$ is the sound speed.

The Euler system (5) can be solved by scalar techniques if *characteristic* variables are introduced; see, e.g., (Harten et al., 1987), (LeVeque, 1998), and (Toro, 1999). For this purpose, consider the Jacobian matrix $A(q)$ whose eigenvalues are $\lambda_1 = u - c$, $\lambda_2 = u$, and $\lambda_3 = u + c$, and the corresponding eigenvectors are denoted by r_1 , r_2 , and r_3 ; see (Harten et al., 1987). The characteristic variables w^k are given by $w^k = l_k q$, $k = 1, 2, 3$. Conversely, we have $q = \sum_{k=1}^3 w^k r_k$.

2.1. THE NEW RECONSTRUCTION METHOD AND THE GODUNOV SCHEME

In the following, $[0, L]$ is the spatial domain where L is the length of the pipe. The boundary conditions are assumed to be represented by the boundary functions $q_1(t)$ and $q_2(t)$. The initial condition is given by $q^0(x)$ at time $t = 0$. Assume that the spatial domain $[0, L]$ has been discretized in a finite number N_{cell} of cells of uniform mesh size Δx indexed by i .

In the Godunov approach, the discrete state vector function q_i^n denotes the cell average of the state vector variable $q(x, t)$ in the cell with index i at the n -th time level, $t = t^n$. With such averages, a piecewise constant vector function $q(x)$ is defined with value q_i^n in the i th cell and with jump discontinuities at each cell interface $x_{i+1/2}$, $i = 1, 2, \dots, N_{cell} - 1$. On the cell interface $x_{i+1/2}$ we denote with q_L and q_R the left- and right-hand side vector values of $q(x)$ at the interface.

To describe the entire algorithm let us first assume that the solution at the n -th time level has been computed. Correspondingly, we construct the variables $\tilde{q}^n = (\tilde{\rho}^n, \tilde{\rho}u^n, \tilde{e}^n)$. For this purpose, we need the inverse matrix of $A(q)$, that is:

$$A^{-1}(q) = \frac{1}{u(u^2 - c^2)} \begin{bmatrix} -c^2 + \frac{k+3}{2}u^2 & -ku & k-1 \\ u(u^2 - c^2) & 0 & 0 \\ -\frac{k-3}{k-1} \frac{u^2 c^2}{2} - \frac{k-3}{2} u^4 & \frac{c^2}{k-1} - \frac{2k-3}{2} u^2 & ku \end{bmatrix}. \quad (7)$$

This matrix is well defined if the determinant $\det(A(q_i))$ is non zero, that is, if $u \neq 0$ and $u \neq \pm c$. Therefore, we define the following:

$$\tilde{A}^{-1}(q_i) = \begin{cases} A^{-1}(q_i), & \text{if } 0 < |u| < c(1 - \epsilon), \\ \mathbf{0}, & \text{otherwise,} \end{cases} \quad (8)$$

and

$$\tilde{g}(x_i, q_i^n) = \begin{cases} \mathbf{0}, & \text{if } 0 < |u| < c(1 - \epsilon), \\ g(x_i, q_i^n), & \text{otherwise,} \end{cases} \quad (9)$$

where ϵ is some small parameter, e.g., $\epsilon = 0.05$ in our experiments.

We modify the standard Godunov scheme by introducing the following modified variables:

$$\begin{aligned} \tilde{q}_1^n &= q_1^n, \\ \tilde{q}_2^n &= q_2^n + \tilde{A}^{-1}(q_1^n)g(x_1, q_1^n)\frac{\Delta x}{2} + \tilde{A}^{-1}(q_2^n)g(x_2, q_2^n)\frac{\Delta x}{2}, \\ \tilde{q}_i^n &= q_i^n + \tilde{A}^{-1}(q_1^n)g(x_1, q_1^n)\frac{\Delta x}{2} + \sum_{j=2}^{i-1} \tilde{A}^{-1}(q_j^n)g(x_j, q_j^n)\Delta x \\ &\quad + \tilde{A}^{-1}(q_i^n)g(x_i, q_i^n)\frac{\Delta x}{2}, \quad i = 3, 4, \dots, N_{cell} \end{aligned} \quad (10)$$

In the same algorithm, we can also compute the vector of slopes $G_i^n = -\tilde{A}^{-1}(q_i^n)g(x_i, q_i)$ and evaluate the function $\tilde{g}(x_i, q_i)$. The variables \tilde{q} are then used to evaluate the corrected characteristic variables $\tilde{w}^k(\tilde{q})$ defined

as follows: $\tilde{w}_i^k = l_k^{(n)}(q_j^n) \tilde{q}_i^n$, for $i = j - m, \dots, j + m$, $k = 1, 2, 3$. Here, m is the desired order of reconstruction.

Next, we apply a scalar reconstruction scheme (we use the ENO method (Harten et al., 1987)) to each corrected characteristic variable and obtain the set of slopes σ_i^k with $k = 1, 2, 3$. Further, we estimate the vector of slopes of the modified conserved variables by the formula,

$$\tilde{S}_i^n = \sum_{k=1}^3 \sigma_i^k r_k. \quad (11)$$

Therefore, we can recover the slopes of the original conserved variables $(\rho, \rho u, e)$ by the algebraic relation,

$$S_i^n = \tilde{S}_i^n + G_i^n. \quad (12)$$

These values are used to compute the state function at $t^{n+1/2}$ given by,

$$q_i^{n+1/2} = q_i^n + \frac{\Delta t}{2} [-A(q_i^n) \tilde{S}_i^n - \tilde{g}(x_i, q_i^n)]. \quad (13)$$

The left- and right-hand side values $q_L^{n+1/2}$ and $q_R^{n+1/2}$ at each interface $x_{i+1/2}$, $i = 1, 2, \dots, N_{cell}$, are given by:

$$q_L^{n+1/2}(i + 1/2) = q_i^{n+1/2} + \frac{\Delta x}{2} S_i^n = q_i^{n+1/2} + \frac{\Delta x}{2} [\tilde{S}_i^n + G_i^n], \quad (14)$$

and

$$q_R^{n+1/2}(i + 1/2) = q_{i+1}^{n+1/2} - \frac{\Delta x}{2} S_{i+1}^n = q_{i+1}^{n+1/2} - \frac{\Delta x}{2} [\tilde{S}_{i+1}^n + G_{i+1}^n]. \quad (15)$$

The values q_L and q_R are then used to define the Riemann problem; see, e.g., (Toro, 1999):

$$\begin{aligned} \frac{\partial}{\partial t} q + \frac{\partial}{\partial x} f(q) &= 0, \\ q(x, 0) &= q_L, \quad x < x_{i+1/2}, \\ q(x, 0) &= q_R, \quad x > x_{i+1/2}. \end{aligned} \quad (16)$$

Here, the domain of interest is the $x-t$ plane with $t > 0$ and $-\infty < x < \infty$. In practice, for our purpose, x varies in a finite interval, say $[x_i, x_{i+1}]$.

This formulation is exploited because the solution $q^*(\frac{x-x_{i+1/2}}{t-t^n}, q_L^n, q_R^n)$ of the Riemann problem is known and is required to compute the fluxes at the

cell interfaces. In an $x - t$ plane, this solution consists of centered waves: two outer shock and/or rarefaction waves and an inner contact discontinuity wave. The solutions of two contiguous Riemann problems will not interact if the CFL condition is satisfied; see, e.g., (Griffiths, Stuart, and Yee, 1992). Therefore, under this condition an exact solution $q^*(0, q_L^n, q_R^n)$ to the approximate problem is found at the interface for the time $t \in [t^n, t^{n+1}]$. In practice, computationally less expensive but accurate solutions are obtained by an approximate Riemann solver. We shall use the approximate Riemann solver of Roe (Roe, 1986).

The fluxes through each cell interface are given by:

$$f_{i+1/2}^{n+1/2} = f(q^*(0, q_L^{n+1/2}(i + 1/2), q_R^{n+1/2}(i + 1/2))),$$

and

$$f_{i-1/2}^{n+1/2} = f(q^*(0, q_L^{n+1/2}(i - 1/2), q_R^{n+1/2}(i - 1/2))),$$

at the interfaces of the cell i for $t_{n+1/2}$. A second-order, conservative, finite-difference formula is then used to obtain q^{n+1} and advance in time:

$$q_i^{n+1} - q_i^n + \frac{\Delta t}{\Delta x} [f_{i+1/2}^{n+1/2} - f_{i-1/2}^{n+1/2}] + \Delta t g(x_i, q^{n+1/2}) = 0, \quad (17)$$

where $i = 1, 2, \dots, N_{cell}$.

2.2. A COMMENT ON THE CORRECTION PROCEDURE

The correction procedure presented in the previous sections requires, in case of systems of equations, to invert the flux Jacobian $A(q)$. This is a low order matrix and it is not a problem to determine the inverse $A(q)^{-1}$ analytically. However, for efficiency reasons it may be convenient to have equivalent formulations of $A(q)^{-1}g(x, q)$ at disposal. When only geometrical terms are present, one obtains (Glaz and Liu, 1984):

$$\frac{\partial}{\partial x} q = -\frac{1}{2C} \frac{dC}{dx} \rho u \left(\frac{r_1}{\lambda_1} + \frac{r_3}{\lambda_3} \right). \quad (18)$$

If only wall friction and heat contribution are given, we have,

$$\frac{\partial}{\partial x} \rho = \frac{(k-1)Q_Z + kuF_R}{(u^2 - c^2)u}, \quad (19)$$

whereas the contribution to the spatial derivative of the momentum is zero, and

$$\frac{\partial}{\partial x} e = -\frac{1}{k-1} (F_R + (k-3)u^2 \frac{\partial \rho}{\partial x}). \quad (20)$$

If geometrical sources are present together with wall friction and heat transfer, the component-wise sum of the above contributions are used. Notice that (18), (19), and (20) give the components of the slope G_i^n in (12), which in turn are used in the algorithm (10).

3. Steady State Solution in a Channel

For a reference solution, we consider steady state flow in a channel with variable cross sectional area. Assuming that the flow in a channel is essentially one dimensional, the conditions in each cross section are described by pressure, temperature, and gas velocity. Furthermore, it is assumed that there is no heat exchange with the channel walls, i.e., that the system is adiabatic, and that wall friction can be neglected. Under these assumptions, the gas undergoes isentropic state changes as long as no shocks are present in the channel. The three equations to describe the flow are:

Conservation of mass (Continuity equation): $\dot{m} = \rho u C = \text{const}$.

Conservation of energy: $c_p T + u^2/2 = \text{const}$.

The isentropic state equation: $(p/\rho)^k = \text{const}$ and $p = \rho RT$.

The boundary conditions are the channel inlet pressure p_0 , the channel inlet temperature T_0 , and the channel outlet pressure p_a . The energy equation is valid from the inlet to any point in the channel. As the fluid outside the channel is at rest, conservation of energy equation becomes,

$$c_p T_0 = c_p T + u^2/2. \quad (21)$$

Solving this equation for u , substituting the temperature difference with the pressure ratio from the isentropic state equation, and calculating the mass flow rate using the continuity equation gives:

$$\dot{m} = C p_0 \sqrt{\frac{2}{RT_0}} \sqrt{\frac{k}{k-1} \left[\left(\frac{p}{p_0}\right)^{\frac{2}{k}} - \left(\frac{p}{p_0}\right)^{\frac{k+1}{k}} \right]}. \quad (22)$$

This equation establishes a relationship between the static pressure in each cross section of the channel and the mass flow rate through the system. As the static pressure at the outlet is a known boundary condition, the mass flow rate can be calculated from (22) with $C = C_{\text{outlet}}$ and $p = p_a$. Afterwards, the pressure in each cross section is obtained by solving (22) for the static pressure p .

Care must be taken when equation (22) is applied to convergent-divergent channels (Laval nozzles). In such nozzles, the flow is accelerated up to the smallest cross section. If the maximum flow velocity in the minimum cross section is still smaller than the local speed of sound, the flow is decelerated

afterwards in the diverging part of the nozzle. The mass flow rate and the static pressure in each cross section are again obtained by the procedure outlined above. If the flow velocity in the smallest cross section is equal to the sound speed, the maximum physically possible mass flow rate through the system is reached. Then, the flow conditions in the converging part of the nozzle are not influenced anymore from the conditions downstream of the smallest cross section. The critical pressure ratio p_c/p_0 is associated with a local Mach number $M_a = 1$ in the throat. If this situation occurs, the mass flow rate through the system is obtained again from (22), but with $p/p_0 = p_c/p_0$ and $C = C_{min}$.

Afterwards, in the diverging part of the nozzle, the flow is accelerated to Mach numbers greater than one. In this flow regime, a shock decelerating the flow from supersonic to subsonic may occur. The conditions downstream of the shock are linked to those upstream of it by the Rankine-Hugoniot jump conditions. After the shock, the stagnation pressure is smaller than before it, which must be taken into account in equation (22). The location of the shock is determined by the condition that the subsonic deceleration in the remaining diverging part of the nozzle must be consistent with the outside pressure at the outlet of the nozzle. If no location in the system satisfies this condition, then no shock will occur in the channel and the flow will be supersonic throughout the diverging part of the nozzle. For a general reference on the subject of this section see the book of Shapiro (Shapiro, 1953).

4. Numerical Experiments

The proposed reconstruction process results in a Godunov scheme that is able to provide accurate non-oscillatory solutions to the Euler equations of gas dynamics with source terms even when severe flow conditions arise. We show this by means of experiments with converging channel and converging-diverging channel. To present these results, we consider three measurement points located on the channel axis of symmetry: at the inlet, at the middle of the channel length, and at the outlet. As initial condition for both examples, assume that the initial pressure within the pipe equals the *ambient* pressure $p = 1$ bar and the initial temperature is $T = 298$ K. Let us now consider the first example in detail.

4.1. EXAMPLE I: CONVERGING CHANNEL

Consider a converging channel 0.5 m long. Between the inlet (origin) and 0.1 m, the section diameter is 0.06 m. Then, between 0.1 m and 0.3 m it reduces linearly to a value of 0.03 m. The remaining part of the pipe has a constant section diameter of 0.03 m. At the inlet is $p_0 = 1.5$ bar

and the temperature $T_0 = 298$ K. In the numerical simulation, the pipe is subdivided into cells of 0.01 m length ($N_{cell} = 50$).

We can see in Table 1 that the steady state computed by the Godunov method with the modified reconstruction technique is very accurate. In fact, this solution is very close to that obtained by the algorithm described in section 3. As can be observed in Figure 1, the steady state solution is reached soon. Figure 1 also shows a magnification of results obtained at the second measurement point with a Godunov scheme using a ENO reconstruction (oscillating) and a ENO reconstruction based on the modified conserved variables.

TABLE 1. Values of pressure, temperature and Mach number at three measuring points along the converging channel.

Reference Solution			
$x(m)$	$p(PA)$	$T(K)$	M_a
0.00000E+00	0.14796E+06	0.29684E+03	0.14003E+00
0.25000E+00	0.13527E+06	0.28933E+03	0.38712E+00
0.50000E+00	0.10000E+06	0.26540E+03	0.78366E+00
modified Godunov method			
$x(m)$	$p(PA)$	$T(K)$	M_a
0.00000E+00	0.14800E+06	0.29680E+03	0.14000E+00
0.25000E+00	0.13510E+06	0.28930E+03	0.38850E+00
0.50000E+00	0.10000E+06	0.26570E+03	0.78380E+00

4.2. EXAMPLE II: CONVERGING-DIVERGING CHANNEL

Another severe test is given by a converging-diverging channel. This nozzle has an initial diameter of 0.06 m which contracts linearly to reach the value of 0.03 m after 0.1 m from the inlet. The constant diameter throat has a length of 0.05 m. Then the pipe diameter diverges linearly to reach the value of 0.06 m at 0.5 m from the inlet. The pressure at the inlet is $p_0 = 2.5$ bar and the temperature $T_0 = 298$ K. Under these conditions, $M_a = 1$ is reached and a shock occurs at $x = 0.415$ m. To show the ability of the algorithm to provide accurate solutions even with very coarse grids we take $N_{cell} = 10$, that is, each cell is 0.05 m long and the throat is modeled with a single cell.

Accurate steady state solutions are obtained for the converging-diverging channel problem. In table 2, we can see that a solution that is very close to the one which results by the algorithm described in section 3 is obtained.

The time evolution of p , T , and M_a is shown for three measurement points in Figure 2. The solution obtained without modification in the reconstruction procedure is affected by oscillations; see Figure 2.

TABLE 2. Values of pressure, temperature and Mach number at three measuring points along the converging-diverging channel.

Reference Solution			
$x(m)$	$p(PA)$	$T(K)$	M_a
0.00000E+00	0.24628E+06	0.29673E+03	0.14655E+00
0.25000E+00	0.33210E+05	0.16739E+03	0.19751E+01
0.50000E+00	0.10000E+06	0.29059E+03	0.35716E+00
modified Godunov method			
$x(m)$	$p(PA)$	$T(K)$	M_a
0.00000E+00	0.24640E+06	0.29680E+03	0.14360E+00
0.25000E+00	0.33380E+05	0.16690E+03	0.19980E+01
0.50000E+00	0.10000E+06	0.29510E+03	0.34738E+00

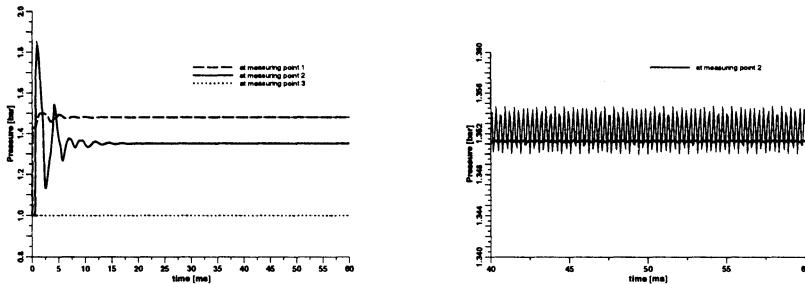


Figure 1. Converging channel solutions and solutions at the second measuring point (right).

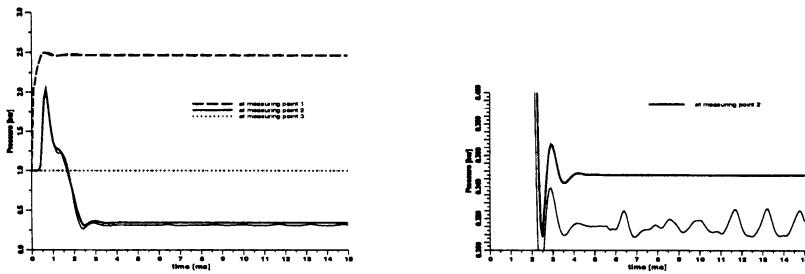


Figure 2. Converging-diverging channel solutions and solutions at the second measuring point (right).

5. Conclusions

We have presented a new reconstruction technique for Godunov method for the Euler equations of gas dynamics with source terms. This method solves the problem of balancing flux gradient with source terms. The proposed reconstruction technique is easy to implement in any high-order Godunov scheme. It has been tested on converging and converging-diverging channel problems under severe flow conditions. The accuracy of the solution method has been proven by comparison with the exact solution of the mass and energy conservation laws for adiabatic gas flows.

References

- Ben-Artzi M and Falcovitz J (1984). A Second-Order Godunov-Type Scheme for Compressible Fluid Dynamics. *Journal of Computational Physics* **55**, pp 1-30.
- Ben-Artzi M and Falcovitz J (1995). Recent Developments of the GRP Method. *JSME International Journal Series B* **38**, pp 497-517.
- Bermúdez A and Vázquez-Cendón M E (1994). Upwind Methods for Hyperbolic Conservation Laws with Source Terms. *Computers and Fluids* **23**, pp 8-20.
- Glaz H M and Liu T P (1984). The Asymptotic Analysis of Wave Interactions and Numerical Calculations of Transonic Nozzle Flow. *Advances in Applied Mathematics* **5**, pp 111-146.
- Glimm J, Marshall G, and Plohr B (1984). A Generalized Riemann Problem for Quasi-One-Dimensional Gas Flows. *Advances in Applied Mathematics* **5**, pp 1-30.
- Greenberg J M and LeRoux A (1996). A Well-balanced Scheme for the Numerical Processing of Source Terms in Hyperbolic Equations. *SIAM Journal on Numerical Analysis* **33**, pp 1-19.
- Griffiths D F, Stuart A M, and Yee H C (1992). Numerical Wave Propagation in an Advection Equation with a Nonlinear Source Term. *SIAM Journal on Numerical Analysis* **29**, pp 1244-1260.
- Harten A (1989). ENO Schemes with Subcell Resolution. *Journal of Computational*

- Physics* **83**, pp 148-184.
- Harten A, Engquist B, Osher S and Chakravarthy S R (1987). Uniformly High-order Accurate Essentially Non-oscillatory Schemes, III. *Journal of Computational Physics* **71**, pp 2-47.
- Hirsch C (1990). Numerical Computation of Internal and External Flows, Vol. 1 & 2. John Wiley & Sons.
- LeVeque R J (1998). Balancing Source Terms and Flux Gradients in High-resolution Godunov Methods: the Quasi-steady Wave-propagation Algorithm. *Journal of Computational Physics* **146**, pp 346-365.
- LeVeque R J and Bale D S (1998). Wave Propagation Methods for Conservation Laws with Source Terms. 7'th International Conference on Hyperbolic Problems, Zürich.
- LeVeque R J and Yee H C (1990). A Study of Numerical Methods for Hyperbolic Conservation Laws with Stiff Source Terms. *Journal of Computational Physics* **86**, pp 187-210.
- LeVeque R J (1998). Nonlinear Conservation Laws and Finite Volume Methods for Astrophysical Fluid Flow. Computational Methods for Astrophysical Flow, 27th Saas-Fee Advanced Course Lecture Notes, Steiner O and Gautschy A (Eds.), Springer.
- Liu T-P (1979). Quasilinear hyperbolic systems transonic nozzle flow. *Comm. Math. Phys.* **68**, pp 141-172.
- Roe P L (1986). Characteristic-Based Schemes for the Euler Equations. *Ann. Rev. Fluid Mech.* **18**, pp 337-365.
- Shapiro A (1953). The Dynamics and Thermodynamics of Compressible Fluid Flow, Vol. 1 & 2. John Wiley & Sons.
- Toro E F (1999). Riemann Solvers and Numerical Methods for Fluid Dynamics. Second Edition, Springer-Verlag.
- Woodward P and Colella P (1984). The numerical simulation of two-dimensional fluid flow with strong shocks. *Journal of Computational Physics* **54**, pp 115-173.

COLELLA-GLAZ SPLITTING SCHEME FOR THERMALLY PERFECT GASES

A. BECCANTINI

CEMIF, Université d'Evry-Val d'Essonne

40, rue du Pelvoux, 91020 EVRY CEDEX, FRANCE

Email: becc@semt2.smts.cea.fr

Abstract. In this paper we discuss the extension of the Colella-Glaz Splitting scheme, originally conceived for a calorically perfect gas, to non-reactive thermally perfect gases mixtures (gases with temperature-dependent specific heat capacities). This extension is very easy to perform, thanks to an original parameterization of the shock curves in the phase space, proposed by the author and here presented.

1. Introduction

An equilibrium gas is said perfect if intermolecular forces can be neglected (Callen, 1960). We can also distinguish between calorically perfect and thermally perfect gases mixture: in the former, the specific heat capacities are temperature independent, while in the latter the vibrational degrees of freedom of each species are taken into account via the dependence of specific heat capacities with respect to temperature.

The extension of the flux splitting schemes designed for calorically perfect gases to thermally perfect gases is not trivial. Note for instance that the Riemann invariants in the calorically perfect case can be expressed in a closed form; but in the thermally perfect case they can involve integral functions, which can be evaluated only numerically. However, as we will show, the shock curves can be expressed in a parametric form very easy to treat. For this reason, we have decided to extend the Flux Difference Splitting (FDS) scheme of Colella-Glaz (Colella, 1982) to the thermally perfect case. As for any FDS scheme, it computes the interface numerical flux by considering a field-by-field decomposition of the solution of the interface Riemann problem (RP). In this case intermediate states are evaluated by

imposing that both genuinely-non-linear (GNL) waves are shock waves, i.e. by intersecting “left” and “right” shock curves in the phase space (as opposed to the Osher scheme, where it is assumed that both non-linear waves are rarefaction waves). This weak solution of the Riemann problem is single-valued, even if it involves non-entropy respecting shock waves. For this reason it can be used to compute the numerical flux just as with an exact Riemann solver. But since the method is not entropy-respecting, an entropy fix must be required.

In the calorically perfect case, the Colella-Glaz scheme has not been widely implemented. The main reason is that the intersection of shock curves cannot be expressed in closed form even in the single-component gas case (as opposed to the Osher scheme). But in our opinion, this is one of the most interesting scheme in the thermally perfect case, thanks to the simplicity of the parameterization of the shock curves.

After a presentation of the 1D Euler equations for thermally perfect gases (section 2), with particular emphasis to the parameterization of the shock and the rarefaction curves in the phase space, we describe the implementation of the scheme with entropy-fix (section 3).

2. The 1D Euler equations for thermally perfect gases

The 1D Euler equations for a non-reactive thermally perfect gases mixture involving n species can be written in conservative form as

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}(\mathbf{U})}{\partial x} = \mathbf{0} \quad (1)$$

where $\mathbf{U} = (\rho, \rho w, \rho e_t, \rho Y_1, \dots, \rho Y_{n-1})^\top$ is the vector of the conserved variables, $\mathbf{F} = (\rho w, \rho w^2 + p, \rho w h_t, \rho Y_1 w, \dots, \rho Y_{n-1} w)^\top$ is the flux vector, $e_t = e + w^2/2$ is the total energy and $h_t = h + w^2/2$ is the total enthalpy. We also define $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ the vector of the mass fractions.

In the case of a thermally perfect gas mixture, the equations of state can be written as

$$p = \rho R T = \rho \left(\sum_{i=1}^n Y_i \frac{R_u}{W_i} \right) T, \quad e = \int_0^T \sum_{i=1}^n Y_i c_{v,i}(\xi) d\xi$$

R_u being the universal constant of perfect gases ($R_u = 8.314 \text{ J/mol/K}$) and W_i being the i -th species molar weight. We also recall the following relations (Callen, 1960):

$$c_{p,i}(T) = c_{v,i}(T) + \frac{R_u}{W_i}, \quad \gamma(T, \mathbf{Y}) = \frac{\sum_i Y_i c_{p,i}(T)}{\sum_i Y_i c_{v,i}(T)}, \quad c_s^2 = \left(\frac{\partial p}{\partial \rho} \right)_S = \gamma \frac{p}{\rho}$$

$c_{p,i}$ and $c_{v,i}$ being the constant pressure and constant volume specific heat capacities, c_s being the speed of sound.

First of all, we recall the following properties in order to specify the notations used in the parameterizations of the shock and the rarefaction curves.

- The Jacobian of the flux vector \mathbf{F} with respect to the conservative variables \mathbf{U} has three different eigenvalues: $\lambda^{(\pm)} = w \pm c_s$, with multiplicity one, relative to the GNL fields, and $\lambda^{(0)} = w$, with multiplicity n , relative to the linearly-degenerate (LD) field. Thus the system (1) is hyperbolic, even if not strictly.
- Under the hypothesis of speed of sound non-decreasing function of the temperature, the hyperbolic system (1) is convex (Beccantini, 2000).
- The family of states which can be connected to the state \mathbf{U}_j by a LD wave are given by imposing $w = w_j$ and $p = p_j$.
- By imposing the constancy of the Riemann invariants, we can determine the one-parameter family of states which can be connected to the left state \mathbf{U}_j by a $\lambda^{(-)}$ -rarefaction wave and to the right state \mathbf{U}_j by a $\lambda^{(+)}$ -rarefaction wave. These sets of states represent two curves in the phase space, the $\lambda^{(-)}$ and $\lambda^{(+)}$ rarefaction curves.
- By imposing the respecting of the Rankine-Hugoniot conditions, we can determine the one-parameter family of states which can be connected to the left state \mathbf{U}_j by a $\lambda^{(-)}$ -shock wave and to the right state \mathbf{U}_j by a $\lambda^{(+)}$ -shock waves. These sets of states represent two curves in the phase space, the $\lambda^{(-)}$ and $\lambda^{(+)}$ shock curves.

Let us now parameterize the $\lambda^{(\pm)}$ -rarefaction curves by choosing the temperature as parameter. After straightforward calculations, it can be shown that $\mathbf{Y}^{(\pm)} = \mathbf{Y}_j$ and

$$\begin{aligned} p^{(\pm)}(T; \mathbf{U}_j) &= p_j \exp \left(\int_{T_j}^T \frac{\gamma(\xi, \mathbf{Y}_j)}{\gamma(\xi, \mathbf{Y}_j) - 1} \frac{d\xi}{\xi} \right) \\ w^{(\pm)}(T; \mathbf{U}_j) &= w_j \pm \int_{T_j}^T \frac{\sqrt{\gamma(\xi, \mathbf{Y}_j) R_j}}{\gamma(\xi, \mathbf{Y}_j) - 1} \frac{d\xi}{\sqrt{\xi}} \\ \rho^{(\pm)}(T; \mathbf{U}_j) &= \rho_j \exp \left(\int_{T_j}^T \frac{1}{\gamma(\xi, \mathbf{Y}_j) - 1} \frac{d\xi}{\xi} \right) \end{aligned} \quad (2)$$

The involved $\lambda^{(\pm)}$ -rarefaction waves are admissible if and only if $T \leq T_j$. Moreover,

$$\begin{aligned} \lim_{T \rightarrow 0^+} \rho^{(\pm)}(T; \mathbf{U}_j) &= 0 \\ \lim_{T \rightarrow 0^+} p^{(\pm)}(T; \mathbf{U}_j) &= 0 \\ \lim_{T \rightarrow 0^+} w^{(\pm)}(T; \mathbf{U}_j) &= w_j \pm \int_{T_j}^0 \frac{\sqrt{\gamma(\xi, \mathbf{Y}_j) R_j}}{\gamma(\xi, \mathbf{Y}_j) - 1} \frac{d\xi}{\sqrt{\xi}} \end{aligned} \quad (3)$$

Let us now parameterize the $\lambda^{(\pm)}$ -shock curves by choosing the temperature as parameter. In that case, it can be shown (Beccantini, 2000) that $\mathbf{Y}^{(\pm)} = \mathbf{Y}_j$ and

$$\begin{aligned}\rho^{(\pm)}(T; \mathbf{U}_j) &= \zeta \rho_j \\ w^{(\pm)}(T; \mathbf{U}_j) &= w_j \pm \text{sign}(\zeta - 1) \sqrt{2(h - h_j) \frac{\zeta - 1}{\zeta + 1}} \\ p^{(\pm)}(T; \mathbf{U}_j) &= \rho(T, \mathbf{U}_j) R_j T\end{aligned}\quad (4)$$

where

$$\zeta = \sqrt{b^2 + c} + b, \quad b = \frac{(h - \frac{1}{2}R_j T) - (h_j - \frac{1}{2}R_j T_j)}{R_j T}, \quad c = \frac{T_j}{T}$$

The involved $\lambda^{(\pm)}$ -shock waves are entropy-respecting if and only if $T \geq T_j$. Moreover,

$$\begin{aligned}\lim_{T \rightarrow 0^+} \rho^{(\pm)}(T; \mathbf{U}_j) &= \rho_j^0 = \frac{R_j T_j}{2h_j - R_j T_j} \rho_j > 0 \\ \lim_{T \rightarrow 0^+} p^{(\pm)}(T; \mathbf{U}_j) &= 0 \\ \lim_{T \rightarrow 0^+} w^{(\pm)}(T; \mathbf{U}_j) &= w_j \mp \sqrt{2h_j \frac{1 - \zeta_j^0}{1 + \zeta_j^0}}\end{aligned}\quad (5)$$

with $\zeta_j^0 = \rho_j^0 / \rho_j < 1$.

Let us now consider the entropy-respecting solution of the RP, i.e. the one which considers only entropy-respecting shock waves and admissible rarefaction waves, between two states \mathbf{U}_L and \mathbf{U}_R . If these states are “not too distant”, i.e. if

$$w_L - \int_{T_L}^0 \frac{\sqrt{\gamma(\xi, \mathbf{Y}_L) R_L}}{\gamma(\xi, \mathbf{Y}_L) - 1} \frac{d\xi}{\sqrt{\xi}} \geq w_R + \int_{T_R}^0 \frac{\sqrt{\gamma(\xi, \mathbf{Y}_R) R_R}}{\gamma(\xi, \mathbf{Y}_R) - 1} \frac{d\xi}{\sqrt{\xi}} \quad (6)$$

it can be performed via the field-by-field decomposition

$$\mathbf{U}_L \xrightarrow{\lambda^{(-)}} \mathbf{U}_L^* \xrightarrow{\lambda^{(0)}} \mathbf{U}_R^* \xrightarrow{\lambda^{(+)}} \mathbf{U}_R$$

Thus we have to compute (T_L^*, T_R^*) such that $p_L^* = p_R^*$ and $w_L^* = w_R^*$. Conversely, if the condition (6) is not satisfied, the solution consists of two rarefaction waves separated by the vacuum state $\mathbf{U} = \mathbf{0}$.

3. The Colella-Glaz scheme: the entropy fix and the vacuum

We can perform the field-by-field decomposition of the RP by supposing that the GNL waves are both shock waves. By comparing (2) and (4), it is

obvious that this is computationally much easier than supposing that the GNL waves are both rarefaction waves. We emphasize that, in the particular case of specific heat capacities evaluated as polynomial functions of T , the evaluation of (4) and of their derivatives with respect to T , involve only elementary operations and square roots.

Actually, the “shock-shock” solution of the RP between two states \mathbf{U}_L and \mathbf{U}_R exists if and only if the left and right shock curves intersect each other, i.e.

$$w_L^0 = w_L + \sqrt{2h_L \frac{1 - \zeta_L^0}{1 + \zeta_L^0}} \geq w_R - \sqrt{2h_R \frac{1 - \zeta_R^0}{1 + \zeta_R^0}} = w_R^0 \quad (7)$$

where $\zeta_j^0 = R_j T_j / (2h_j - R_j T_j)$, $j = L, R$.

If condition (7) is satisfied, the solution thus obtained cannot be readily used to compute the numerical flux: an entropy fix must be implemented in order to avoid the capturing of non-entropy respecting shock waves. Since ρ, w, T are monotonic functions of x/t on a rarefaction waves, in the entropy fix here proposed we replace the non-entropy respecting shock waves with a linear approximation of ρ, w, T with respect to x/t , as represented in figure 1 (solid line); p is recovered from the values of ρ and T via the equation of state.

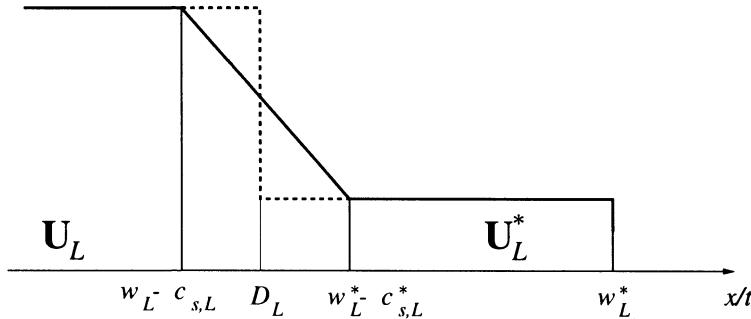


Figure 1. A non-entropy respecting $\lambda^{(-)}$ shock wave (dashed line), which violates the Lax entropy-condition $w_L - c_{s,L} > D_L > w_L^* - c_{s,L}^*$, D_L being the shock speed.

If condition (7) is not satisfied we cannot construct a vacuum solution, as in the case of the entropy-respecting field-by-field decomposition. In fact, from (5), we deduce that

$$\mathbf{0} \neq \lim_{T \rightarrow 0^+} \mathbf{U}^{(-)}(T; \mathbf{U}_L) \neq \lim_{T \rightarrow 0^+} \mathbf{U}^{(+)}(T; \mathbf{U}_R) \neq \mathbf{0}$$

By defining w_L^0 and w_R^0 as in (7), we can proceed as follows:

- if $w_L^0 \geq 0$ or $w_R^0 \leq 0$, we can compute the numerical flux by observing

that $\mathbf{U}_L \xrightarrow{\lambda^{(-)}} \mathbf{U}_L^0$ and $\mathbf{U}_R^0 \xrightarrow{\lambda^{(+)}} \mathbf{U}_R$ are non-entropy respecting shock waves;
• conversely, if $w_L^0 \leq 0 \leq w_R^0$, we impose that the numerical flux is $\mathbf{0}$.

From a theoretical point of view, it can be shown that we do not violate the continuity condition of the numerical flux with respect to \mathbf{U}_L and \mathbf{U}_R (Beccantini, 2000). From a practical point of view, this way of proceeding has shown to be robust.

4. Numerical experiments

In (Beccantini, 2000), we have compared the numerical scheme thus obtained with other classical flux splitting schemes in computing classical 1D and 2D test-cases. Summarizing the 1D numerical results, this scheme is accurate (exactly capturing of stationary contact discontinuities and stationary shock waves, as shown in figure 2) and robust in dealing with strong shock and strong rarefaction waves (figure 3). As far as 2D computations are concerned, it suffers of the typical deficiencies which affect many flux splitting solvers capable of capturing stationary contact discontinuities (for instance, as the Godunov scheme, it fails in computing the “odd-even perturbation” test-case proposed by Quirk (Quirk, 1994)).

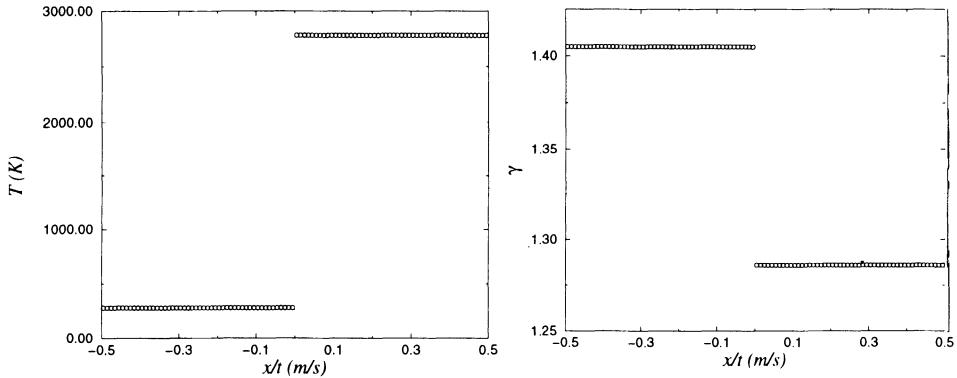


Figure 2. Behavior of the Colella-Glaz scheme in computing a $\lambda^{(-)}$ entropy-respecting stationary shock wave. Mixture of N₂ and O₂ (specific heat capacities evaluated as 4-th order polynomials of T). $Y_{N_2,L} = Y_{N_2,R} = 0.77785$, $p_L = 0.1013$ MPa, $T_L = 278$ K, $T_R = 10 T_L$; $p_R \approx 6.655$ MPa, $w_L \approx 2474$ m/s, $w_R \approx 376.6$ m/s.

5. Conclusion

In this paper we have presented the extension of Colella-Glaz FDS scheme to perfect gases with temperature dependent specific heat capacities. This extension is really easy to perform, thanks to the original parameterization

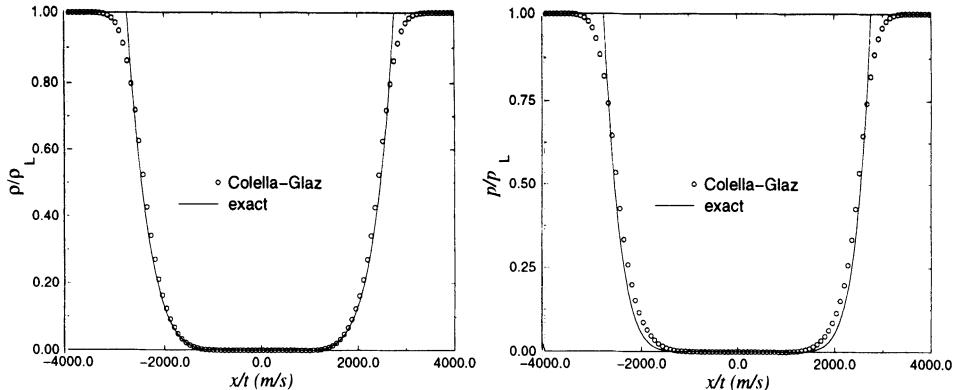


Figure 3. Behavior of the Colella-Glaz scheme in computing the vacuum formation. Mixture of N₂ and O₂ (specific heat capacities evaluated as 4-th order polynomials of T). $Y_{N_2,L} = Y_{N_2,R} = 0.77785$, $\rho_L = \rho_R = 1.0 \text{ kg/m}^3$, $p_L = p_R = 0.1013 \text{ MPa}$, $w_R = -w_L = 7.5\sqrt{p_L/\rho_L}$. CFL = 0.35, $t_{\text{fin}} = 1.25 \cdot 10^{-4}$, $N_x = 100$, $D_x = 0.01$.

of the shock curves (4). Moreover the entropy-fix and the way of treating the vacuum here proposed make the scheme robust in dealing with strong rarefaction waves.

Acknowledgments

This work has been funded by the French Atomic Energy Commission (CEA).

References

- Callen H (1960). Thermodynamics. John Wiley & Sons, New York.
- Colella P (1982). Glimm's Method for Gas Dynamics. *SIAM J. Sci. Stat. Comp.*, **3**.
- Beccantini A (2000). Upwind Splitting Schemes for Ideal Gases with Temperature Dependent Specific Heat Capacities. *PhD Thesis, Université d'Evry val d'Essonne* (in preparation).
- Quirk JJ (1994). A Contribution to The Great Riemann Solver Debate, *International Journal for Numerical Methods In Fluids*, **18**

MESHLESS PARTICLE METHODS: RECENT DEVELOPMENTS FOR NONLINEAR CONSERVATION LAWS IN BOUNDED DOMAIN

B. BEN MOUSSA

*Département de Génie mathématique et modélisation,
Unité mixte du CNRS, MIP
INSA, Toulouse, France.
Email : benmou@gmm.insa-tlse.fr*

Abstract. This paper is devoted to provide a suitable model of particle approximation for nonlinear conservation laws in bounded domains by performing an original approach of boundary forces introduced by Monaghan. Afterwards we give a new convergence analysis result in the scalar case by revisiting a concept of measure-valued solutions initially developed by Diperna. To achieve this convergence result, we have been led to give a new definition of such a concept, which is weaker than the one introduced by Szepessy. This in the sense that our definition has fewer requirements concerning the definition of the boundary term's trace of measure-valued solutions.

1. Introduction

In the last few years, many important works and numerical applications around of crack propagation in elasticity, high velocity impact, free surface flow, or multiphase flow in fluid dynamics, showed the ability of Meshless particle methods, as opposed to classical methods like finite elements method, to handle complex situations. Such methods have experienced a significant improvement, through various versions ; we first quote the most popular of them, SPH method (Smooth Particle Hydrodynamics) which was introduced by Lucy in (Lucy, 1977) and Monaghan in (Gingold and Monaghan, 1977), we also refer to [(Gingold and Monaghan, 1983),(Benz, 1989),(Bicknell, 1991), (Randles and Libertsky, 1996),(Hernquist and Katz, 1989) and also references therein] for some improvements and applications.

Furthermore, three variants of SPH method, have been introduced to provide more accurate results : normalized smoothing functions for SPH methods, see (Johnson and Beissel, 1996), MLSPH method (least squares particle hydrodynamics) (Dilts, 1999) and, finally, MLSRK (moving least square reproducing kernel method) (Liu Li and Belytschko, 1997). Despite those developments, their convergence properties and their links with conservativity were not well understood. This motivated our first interest, with Vila, for the study of this method in [(Ben Moussa, 1998), (Ben Moussa and Vila, 2000), (Vila, 1999), (Ben Moussa et al., 1999)] by remedying these problems successively by :

- considering the weak formulation of the problem, in order to derive a new form of the particle scheme which accounts for physical properties related to shock waves,
- connecting the classical particle approximation with finite volume technique, and on that account, we introduce a new hybrid particle approximation, which is well suited in the use of the first point,
- using approximate Riemann solvers instead of the current technique of artificial viscosity used in the context of SPH method, in order to stabilize the scheme, as soon as we proceed to Euler scheme in time,
- introducing a new discrete linear differential operator which approximates the first derivative of any regular function in a more accurate sense than the one of which the SPH method takes advantage. This enables us to introduce a new version of particle methods, the so-called Renormalized Meshless Derivative (RMD)(Ben Moussa et al., 1999).

On the other side, curiously, we witnessed a belated emergence of industrial codes using those methods, when dealing with industrial problems that obviously happen in surrounding walls ; this arises particularly from the deficiency of conceptualization in the numerical treatment of boundary conditions, as opposed to the other techniques of approximation like finite elements method. Nevertheless, there were some attempts to solve this difficulty. We quote the work of Benz in (Benz, 1993) for a semi-analytical approach, which uses an approximation of the integrals of the shape function and its derivatives. There is another approach which consists in replacing the boundaries by boundary particles introduced by Monaghan in (Monaghan, 1995), in the simulation of gravity currents.

Our aim in this paper is to perform the last one without any compromising of the flexibility of the SPH which widely contributes until now in its popularity, by introducing a new and general tool to perform an approach of boundary forces in the context of wall boundaries. Such a strategy takes advantage of the equilibrium condition for a uniform field to give a nice volumic particle approximation of the boundary integral term provided by

the weak formulation of problem (1). As a matter of fact, we particularly interpret it as a suitable approximation of an additional source term.

An outline of this paper is as follows. In the next section, we recall the basic principle of RMD method. In section (3), we review our model of the particle scheme (Ben Moussa et al., 1999) derived from the weak formulation of the conservation laws and we describe our new model of boundary forces. Section (3.2) is devoted to state the main convergence theorem and also to give an idea of its proof.

2. Position of the problem and RMD method

Let v be a regular vector field in \mathbb{R}^d . As we have mentioned above, we are interested in the numerical analysis of the particle method for the following model of PDE's in conservation form with boundary conditions.

$$\begin{cases} L_v(u) + \operatorname{div} F(u, x, t) = S(u, x, t), & (x, t) \in \Omega \times \mathbb{R}^+, u \in \mathbb{R} \\ u(x, 0) = u_0(x), & u_0 \in L^\infty(\Omega), \\ b(s, t), & b(s, t) \in L^\infty(\partial\Omega \times \mathbb{R}^+), \end{cases} \quad (1)$$

where $x = (x^1, \dots, x^d)$, $F = (F^1, \dots, F^d)$, $v = (v^1, \dots, v^d)$, and L_v is the transport operator given by :

$$u \longrightarrow L_v(u) = \frac{\partial u}{\partial t} + \sum_{\alpha=1,d} \frac{\partial}{\partial x^\alpha} (v^\alpha u).$$

2.1. RENORMALIZED MESHLESS DERIVATIVE

The particle method SPH, which is a Lagrangian method, has been introduced as an alternative to grid-based codes to solve the equations of motion of compressible fluids. From the methodological point of view, to implement such methods, we need to take a set of initial distribution of particles and their associate weights together provided by some quadrature formula $(x_k^0, w_k^0)_{k \in K}$, such that :

$$\Omega = \cup_{k \in P} B_k^0, \quad w_k^0 = \operatorname{meas}(B_k^0), \quad C_1 h^d \leq w_k^0 \leq C_2 h^d \quad \forall k \in P.$$

Afterwards, we classically move such particles along the characteristic curves of the field a while modifying naturally the weights in order to take into account the deformations due to the field v , to obtain at each time t a quadrature formula $(x_k(t), w_k(t))_{k \in K}$, where $x_k(t)$ is the position of the

in the sequel of this paper we omit the time t dependence when there is no ambiguity

particle and $w_k(t)$ its weight. Such an evolution is governed by

$$\begin{aligned} (i) \quad & \frac{d}{dt}x_k = v(x_k, t), & x_k(0) = \xi_k \\ (ii) \quad & \frac{d}{dt}w_k = \operatorname{div}v(x_k, t)) w_k & w_k(0) = \omega_k^0. \end{aligned} \quad (2)$$

We also denote by B_k the image at time t of the cell B_k^0 by the vector field v , so thanks to its regularity, we get :

$$\Omega = \cup_{k \in P} B_k, \quad w_k = \operatorname{meas}(B_k), \quad C_3 h^d \leq w_k \leq C_4 h^d \quad \forall k \in P.$$

Thus in these conditions, the accuracy of the particle approximation is connected with the quadrature formula $(x_k, w_k)_{k \in P}$ as follows :

$$\int_{\Omega} g(x) dx \approx \sum_{k \in P} w_k g(x_k).$$

It is proved in (Raviart, 1985) that this approximation is accurate for any $t > 0$ as soon as it is accurate initially. On the other side, by construction, the particle method as opposed to the classical methods, takes advantage of the particles distribution to perform a specific rule to provide an approximation of the first derivative operator appearing in (1) by means of a classical mollification.

$$D_{h,\varepsilon}g(x) := \sum_{k \in P} w_k g(x_k) \nabla \zeta^{\varepsilon}(x - x_k), \quad \zeta_{\varepsilon}(x) = \frac{1}{\varepsilon^d} \zeta(\frac{x}{\varepsilon}), \quad (3)$$

where the shape function $\zeta \in W^{2,\infty}(\Omega)$ is supposed to have a compact support, to be nonnegative, normalized by unity mass and finally has a radiad symmetry property.

Raviart in (Raviart, 1985) has studied the convergence of such an approximation. The result of his study is that

$$\|D_{h,\varepsilon}\varphi(x) - \nabla\varphi(x)\|_{\infty} \rightarrow 0 \text{ as } h = o(\varepsilon^2), \quad h \rightarrow 0, \quad \varepsilon \rightarrow 0.$$

In practice, the ratio h/ε^2 characterizes the mean interaction among particles for instance in 2D ; it is well known that each particle must interact with 27 neighbour particles. To perform the accuracy of such an approximation, we have introduced in (Ben Moussa et al., 1999) a new version of particle methods, the so-called Renormalized meshless derivative which takes advantage of the symmetric form of a previous discrete operator D_h as follows :

$$D_{h,\varepsilon}^s \varphi(x) := \mathcal{N}(x)(D_{h,\varepsilon}\varphi(x) - \varphi(x).D_{h,\varepsilon}1(x)), \quad (4)$$

where the additional matrix $\mathcal{N}(x)$ is introduced to provide a more accurate result in the sense that

$$D_{h,\varepsilon}^s f(x) = \nabla f(x) \text{ for any } f \in P1, \quad (5)$$

where $P1$ is the space of a polynomial of degree less or equal to one. Therefore, it is easy to see that this last requirement is true if and only if the matrix $\mathcal{N}(x)$ is computed by :

$$\mathcal{N}(x) = [E(x)]^{-1} \quad E^{ij}(x)_{(1 \leq i,j \leq d)} = \sum_{l \in P} w_l (x_l^j - x^j) \partial^i \zeta^\varepsilon(x - x_l), \quad (6)$$

which makes sense if and only if the matrix $E(x)$ is invertible. In (Ben Moussa, 2000), we prove that this condition is surprisingly satisfied under less restricting conditions related to the initial particle distribution. Besides, as opposed to the standard SPH method which needs that $h = o(\varepsilon^2)$, here, we only need that the ratio $\frac{h}{\varepsilon}$ is $\mathcal{O}(1)$, and therefore for any $\varphi \in W^{2,\infty}(\Omega)$:

$$\|D_{h,\varepsilon}^s \varphi(x) - D\varphi(x)\|_\infty \rightarrow 0 \text{ as } \varepsilon \rightarrow 0. \quad (7)$$

We refer to (Ben Moussa et al., 1999) for numerical applications and comparisons results in 2D between RMD and SPH methods, in the frame of high velocity impact, where, on the one hand, we have clearly seen a substantial difference related to the accuracy confirming the above analysis result, and, on the other hand, for RMD approach, the interaction among particles only requires that each particle interacts with 6 neighbour particles.

3. The numerical particle scheme and boundary forces

To account for the physical phenomena of chock waves and Rankine Hugo-niot conditions, we have developed in unbounded domain (i.e. $\Omega = \mathbb{R}^d$), a new strategy to perform the suitable model of particle scheme by considering the weak form of the equation (1) [see (Ben Moussa and Vila, 2000), (Ben Moussa, 1998)]. This leads us to introduce a new hybrid particle approximation (see (Ben Moussa, 1998)), which is allow to this last one to make sense in the sense of distributions, namely :

$$\bar{u}_h(x, t) = \sum_{k \in P} u_k \chi_{B_k}(x), \quad u_k(0) = \frac{1}{\text{meas}(B_k)} \int_{B_k} u_0(x) dx, \quad (8)$$

where u_k stands for an approximation of the exact solution $u(x_k, t)$ of equation (1). We also introduce the adjoint operator of $D_{h,\varepsilon}^s$ by means of a quadrature formulae and a discrete scalar product as :

$$(D_{h,\varepsilon}^s \varphi, \Psi)_h = - (\varphi, D_{h,\varepsilon}^{s,*} \Psi)_h \quad (\varphi, \Psi)_h := \sum_{k \in P} w_k \varphi_k \cdot \Psi_k. \quad (9)$$

3.1. WEAK FORM OF THE SCHEME

We begin by denoting by L_a^* the adjoint operator of L_a and we also consider a regularization $\chi^\kappa(x)$ of the characteristic function of domain Ω , in a first time to neglect the boundary integral term, then the weak formulation of the scheme could be written as :

$$\int_{[0,T]} [(\bar{u}_h, L_a^*(\varphi))_h + (F(\bar{u}_h, x, t), \chi^\kappa D_{h,s} \varphi)_h + (S(\bar{u}_h, x, t), \varphi)_h] dt = 0. \quad (10)$$

Making an integration by part with respect to t , we easily get that (10) is true if and only if :

$$\frac{d}{dt} (w_k u_k) + w_k \sum_{l \in K} w_l^n (\chi^\kappa(x_k) F_k \cdot \mathcal{N}(x_k) A_{kl} - \chi^\kappa(x_l) F_k \cdot \mathcal{N}(x_l) A_{kl}) = w_k S_k, \quad (11)$$

where the notations used above are as follows :

$$A_{kl} := \nabla_{(x=x_k)} \zeta^\epsilon(x_k - x_l), \quad \text{and} \quad G_k = G(u_k, x_k, t).$$

Within sight of the convergence analysis, we symmetrize all the other quantities appearing in the scheme (11), namely the matrix $\mathcal{N}(x)$ and the function $\chi^\kappa(x)$ by evaluating them at point $x_{kl} = \frac{x_k+x_l}{2}$. Therefore, using the notations $\chi_{kl}^\kappa = \chi^\kappa(x_{kl})$, $\mathcal{A}_{kl} = \mathcal{N}(x_{kl}) A_{kl}$ and $n_{kl} = \frac{\mathcal{A}_{kl}}{\|\mathcal{A}_{kl}\|}$, then equation (11) reads

$$\frac{d}{dt} (w_k u_k) + w_k \sum_{l \in K} w_l \chi_{kl}^\kappa (F_k + F_l) \cdot n_{kl} \|\mathcal{A}_{kl}\| = w_k S_k.$$

This approximation could fail to converge if we use an explicit in time discretization as Euler forward (it is sufficient to consider a 1-dimensional situation with $v \equiv 0$ and $F(u, x, t) \equiv cu$ with c a constant velocity, to see that we get the unconditionally unstable explicit centered scheme). Thus, we need to add some upwinding in the discretization to stabilize the scheme, which could be achieved classically by adding some amount of artificial viscosity. This approach is effectively used in SPH algorithms for Euler compressible equations developed in (Gingold and Monaghan, 1983). Here, we propose a somewhat different approach which uses nonlinear upwinding and Riemann approximate solvers well known in the field of finite difference schemes for nonlinear hyperbolic equations. Yes indeed, this form of the scheme introduces the Riemann problem related with any couple (x_k, x_l) of neighbouring particles :

$$\begin{cases} \frac{\partial}{\partial t}(v) + \frac{\partial}{\partial x}(F(v, x_{kl}, t) \cdot n_{kl}) = 0 \\ v(x, 0) = \begin{cases} u_k & \text{if } x < 0 \\ u_l & \text{if } x > 0, \end{cases} \end{cases}$$

which suggests naturally of its nice approximation to replace the centered approximation $(F_k + F_l).n_{kl}$ by the numerical flux of a finite difference scheme $2g(n_{kl}, u_k, u_l)$, which is required to satisfy :

$$\begin{aligned} (i) \quad g(n(x), u, u) &= F(x, t, u).n(x) \\ (ii) \quad g(n, u, v) &= -g(-n, v, u). \end{aligned}$$

The numerical viscosity $Q(n, u, v)$ of such a numerical scheme, is classically defined in the scalar case (i.e. $u \in \mathbb{R}$) as :

$$Q(n(x), u, v) = \frac{F(u, x, t).n(x) - 2g(n(x), u, v) + F(v, x, t).n(x)}{v - u}.$$

There are a lot of numerical fluxes well suited for such an upwinding ; among them, we can quote the Lax Friedrichs and the Godunov schemes which belong to the widest class of E schemes - see (Osher, 1984) and (Vila et al, 1995) for detailed dissipation properties.

Our numerical scheme, which consists in finding functions $t \in \mathbb{R}^+ \rightarrow u_k(t) \in \mathbb{R}$, $i \in P$ which are solutions of the differential system :

$$\frac{d}{dt}(w_k u_k) + w_k \sum_{j \in P} w_l 2\chi_{kl}^\kappa g(n_{kl}, u_k, u_l) \|\mathcal{A}_{kl}\| = w_k S_k,$$

also reads by introducing the numerical viscosity as :

$$\begin{aligned} \frac{d}{dt}(w_k u_k) + \\ w_k \sum_{l \in P} w_l \chi_{kl}^\kappa [(F(u_k, x_{kl}, t) + F(u_l, x_{kl}, t)).n_{kl} + \Pi_{kl}] \|\mathcal{A}_{kl}\| = w_k S_k \quad (12) \end{aligned}$$

where $\Pi_{kl} = Q(n_{kl}, u_k, u_l)(u_k - u_l)$. This form is very closed to the one classically used in the SPH literature (Gingold and Monaghan, 1983).

3.2. APPROXIMATION OF THE BOUNDARY TERM

As we have seen in the previous section, the algorithm of our particle method is completely defined by the resolution of differential equations (2),(12). Thus, we easily see that by the system (12), the interaction among particles is controlled only by the size ϵ of the shape function ζ^ϵ , whereas the new position of particles and their associate weights at any time step are given by (2), as opposed to the other classical methods in the most complex applications, where the adaptativity of the Eulerian Grid is required. Our

second aim in this paper is to provide a suitable particle approximation of boundary integral term without any compromising of this last flexibility which characterize our particle scheme. To this purpose, the basic idea is first to give an equivalent volumic approximation of boundary conditions, by means of a local coordinate system in the neighbourhood of $\partial\Omega$ such that $x = \bar{x} - yn(\bar{x})$ with $(\bar{x}, y) \in \partial\Omega \times [0, 3\kappa]$, for some $\kappa > 0$, as :

$$b(x, t) = b(\bar{x}, t)\beta^\kappa(y),$$

where the mollifier β^κ is devoted to characterize the desired profil of boundary forces in a volumic sense. We refer to (Monaghan, 1995) for concret exemples as Leonard-Jones inter-molecular force with a repulsive core and attractive well or central force. Secondly, we would like to introduce an additional Riemann problem associated to each particle of the fluid k located at the position $x_k \in \Omega$ such that $\beta^\kappa(y) \neq 0$ as follows :

$$\begin{cases} \frac{\partial}{\partial t}(v) + \frac{\partial}{\partial x}(F(v, x_k, t).\tilde{n}_k) = 0 \\ v(x, 0) = \begin{cases} u_k & \text{if } x < 0 \\ b_k = b(x_k) & \text{if } x > 0, \end{cases} \end{cases}$$

where the main unknown here is the direction \tilde{n}_k . So, as for volumic terms, the introduction of a numerical flux g in order to approximate a solution of this Riemann problem enables us to write our global numerical scheme including an associate volumic approximation of boundary term as follows :

$$\frac{d}{dt}(w_k u_k) + w_k \sum_{l \in K} w_l \chi_{kl}^\kappa 2g(n_{kl}, u_k, u_l) \|\mathcal{A}_{kl}\| + \theta_k g(\tilde{n}_k, u_k, b_k) = w_k S_i,$$

where the function $\theta(x)$ and the direction \tilde{n}_k , which will be determined, are devoted successively :

- to guarantee the flexibility of the approximation,
- to provide also a weak consistency property with the boundary integral term,
- to achieve the determination of our model of boundary forces of equation (1).

This model is well posed if particularly it is still valid in the case of pure convection phenomena (i,e where $F(u, x, t) = cte$ and $S(u, x, t) = 0$), namely

$$\frac{d}{dt}(w_k u_k) = 0 \iff \theta(x_k) \tilde{n}_k = \sum_{l \in K} w_l \chi^\kappa(x_{kl}) \mathcal{A}_{kl} := \partial_x^{h,\epsilon} \chi^\kappa(x_k).$$

In such a case, a suitable choice is as follows :

$$\tilde{n}_k = \frac{-\partial_x^{h,\epsilon} \chi^\kappa(x_k)}{\|\partial_x^{h,\epsilon} \chi^\kappa(x_k)\|}, \quad \theta(x_k) = \|\partial_x^{h,\epsilon} \chi^\kappa(x_k)\|.$$

In order to show the weak consistency of this last approximation with the boundary integral term, let us provide a concret exemple by choosing as a function χ^κ by :

$$\chi^\kappa(x(\bar{x}, y)) = \begin{cases} 0 & 0 \leq y < \kappa \\ 1/2 + 3(y - 2\kappa)/(4\kappa) - (1/4)((y - 2\kappa)/\kappa)^3 & \kappa \leq y < 3\kappa \\ 1 & y \geq 3\kappa, \end{cases} \quad (13)$$

and (\bar{x}, y) is a local coordinate system in the neighbourhood of $\partial\Omega$ such that $x = \bar{x} - yn(\bar{x})$ with $(\bar{x}, y) \in \partial\Omega \times [0, 3\kappa]$, for some $\kappa > 0$. Thus a tedious computation using essentially a Taylor expansion applied to the funtion χ^κ together with the result (7), proves that $\partial_x^{h,\epsilon}\chi^\kappa(x_k)$ is an approximation of $\nabla\chi^\kappa = -\nabla(1 - \chi^\kappa)$, in which case, we can prove that our new model of boundary forces is consistent with boundary integral term. Yes indeed by means of a change of variables, we have :

$$\int_{\Omega} \varphi \operatorname{div}(1 - \chi^\kappa) dx = \int_{\partial\Omega} \int_1^3 \varphi(\bar{x}, \kappa y) \left(3/4 - \frac{3}{4}(y - 2)^2 \right) n(\bar{x}) J(x(\bar{x}, \kappa y)) d\bar{x} dy$$

Therefore, we get :

$$\lim_{\kappa \rightarrow 0} \int_{\Omega} \varphi \operatorname{div}(1 - \chi^\kappa) dx = \int_{\partial\Omega} \varphi(\bar{x}) n(\bar{x}) J(x(\bar{x})) d\bar{x} = \int_{\partial\Omega} \varphi n d\sigma(x).$$

The time discretization is performed by using a forward Euler scheme as :

$$\begin{cases} x_k^{n+1} - x_k^n = \tau^n v(x_k^n, t^n), & w_k^{n+1} - w_k^n = \tau^n \operatorname{div}(v(x_k^n, t)) w_k^n, \\ w_k^{n+1} u_k^{n+1} = w_k^n \left(u_k^n - \Delta t \sum_{l \in P} w_l^n \chi_{kl}^\kappa 2g(n_{kl}, u_k^n, u_l^n) \| \mathcal{A}_{kl}^n \| + \right. \\ \left. \Delta t (S_k^n - \theta_k^n g(\tilde{n}_k, u_k^n, b_k^n)) \right). \end{cases} \quad (14)$$

4. Convergence analysis

We can then prove the following theorem in the scalar nonlinear case :

Theorem: Let $u \in L^\infty(\Omega \times \mathbb{R}^+)$ be the unique entropy solution of (1) recently established by Otto in (Otto, 1996) with initial and boundary conditions $(u_0, b) \in L^\infty(\Omega) \times L^\infty(\partial\Omega \times \mathbb{R}^+)$. Thus we suppose that the numerical flux g belongs to the class of E-flux, then by means of a suitable CFL condition, the approximate solutions \bar{u}_h given by (8) and the scheme (2),(14) converges towards u in $L^\infty(\mathbb{R}^+; L^1(\Omega))$ when successively $\varepsilon \rightarrow 0$ and $\kappa \rightarrow 0$.

Idea of the proof. The proof of this theorem is more technical [see (Ben Moussa, 1998),(Ben Moussa, 2000)] and can allow three levels of difficulties ; such levels are as follows :

- we first establish the L^∞ stability of \bar{u}_h by proving in a similar way for unbounded domain [see (Ben Moussa and Vila, 2000)] that the particle scheme is a convex combination of 1D finite difference schemes for which we add a controlled term due to the source term and the explicit dependence in x of the flux F . we point out that this result constitutes the minimum requirement to use the framework of Young measures. Afterwards, we also show that the total variation of the scheme (or the so-called weak BV estimate) associated to \bar{u}_h blows up like $\varepsilon^{-1/2}$ as ε goes to zero. This result is indispensable in order to controll the additional term due to the numerical viscosity.
- thanks to the first level, we first deduce that \bar{u}_h converges weakly for the L^∞ weak star topology. Thus by using a Young theorem we have that there is a Young measure $\nu_{x,t} : \Omega \times \mathbb{R}_+ \rightarrow \text{Prob}(\mathbb{R})$ such that :

$$\text{supp}(\nu_{x,t}) \subset \{\lambda : |\lambda| \leq K\} \quad \forall (x,t) \in \Omega \times \mathbb{R}_+$$

and for all g in $\mathcal{C}(\mathbb{R})$ the $L^\infty(\Omega \times \mathbb{R}_+)$ weak star limit as $h \rightarrow 0$

$$g(\bar{u}_h(.)) \rightarrow <\nu_{x,t}, g(\lambda, x, t)> = \int_{\mathbb{R}} g(\lambda, x, t) d\nu_{x,t}(\lambda) \quad \forall (x,t) \in \Omega \times \mathbb{R}_+$$

exists. Secondly, we start from the global entropy dissipation of the scheme, to derive the existence of a Young measure defining in the scalar case a measure valued solution of problem (1). In fact, this step necessitate the introduction of new definition of such concept in more appropriate sense, namely

Definition: A Young measure $\nu_{x,t}$ obtained in the above Thanks to Young theorem, is said a measure-valued solutions of problem (1), if and only if :

At level of volumic terms :

$\forall \varphi \geq 0 \in \mathcal{C}_c^1(\Omega \times \mathbb{R}^+)$ and $\forall \eta$ entropy (convex and regular) function associated with an entropy flux $H(u, x, t)$ such that $\partial_u H^i(u, x, t) = \eta'(u) \partial_u F^i(u, x, t)$ such that :

$$\begin{aligned} \mathcal{M}^\eta(\varphi) := & \int_{\Omega \times \mathbb{R}} \{ <\nu_{x,t}, \eta(\lambda)> \partial_t \varphi + <\nu_{x,t}, H(\lambda, x, t)> \nabla_x \varphi \} dx dt \\ & + \int_{\Omega \times \mathbb{R}} <\nu_{x,t}, \sum_{i=1,d} (\partial_{x^i}(H^i(\lambda, x, t)) - \eta'(\lambda) \partial_{x^i} F^i(\lambda, x, t))> \varphi dx dt \\ & + \int_{\Omega} \eta(u_0(x)) \varphi(x, 0) dx + \int_{\Omega \times \mathbb{R}} <\nu_{x,t}, \eta'(\lambda) S(\lambda, x, t)> \varphi dx dt \geq 0. \end{aligned} \tag{15}$$

At level of boundary term :

There is a Radon measure $\vartheta_{s,t} \in \mathcal{M}_b(\partial\Omega \times \mathbb{R}_+)$, such that :

$$\begin{aligned} \forall \varphi \geq 0 \in \mathcal{C}_c^1(\bar{\Omega} \times \mathbb{R}^+), \quad & \forall c \in \mathbb{R} \\ \lim_{\delta \rightarrow 0} \mathcal{M}^{\eta_c^\delta}(\varphi) - \int_{\partial\Omega \times \mathbb{R}} \operatorname{sgn}(b-c)\varphi(s,t) d\vartheta_{s,t} \\ + \int_{\partial\Omega \times \mathbb{R}} <\gamma\nu_{s,t}, F(c,s,t)n> \operatorname{sgn}(b-c) \varphi d\sigma(x) dt \geq 0, \end{aligned} \tag{16}$$

where η_c^δ represent some regularization of Kruskov entropy function η_c and $\gamma\nu(\cdot)$ is a "trace" of ν on $\partial\Omega \times \mathbb{R}_+$ in the sense of Szepessy (Szepessy, 1989).

- next, we prove that the last definition of measure-valued solutions is in fact equivalent to the one used by Szepessy in (Szepessy, 1989) and finally, we have to adapt his uniqueness result of such measure valued solutions, precisely, we show that :

$$\nu_{x,t} = \delta_{u(x,t)}, \text{ for almost every } (x,t) \in \Omega \times \mathbb{R}^+,$$

in our present case of more robust setting of bounded and measurable functions $u \in L^\infty(\Omega \times \mathbb{R}^+)$ like Kruskov in unbounded domains instead of the space of function with bounded variation which ensures the existence of the trace. \square

References

- Benharbit S, Chalabi A and Vila J P (1995). Numerical viscosity, and convergence of finite volume methods for conservation laws with boundary conditions. *SIAM Journal On Numerical Analysis* vol **32** number **3**, pp 775-796.
- Ben Moussa B and Vila J P (2000). Convergence of SPH method for scalar nonlinear conservation laws. *SIAM Journal On Numerical Analysis* vol **37** number **3**, pp 863-887.
- Ben Moussa B, Lanson N and Vila J P (1999). Convergence of Meshless Methods for Conservation Laws : application to Euler equations. *International Series of Numerical Mathematics Birkhauser Verlag Basel/Switzerland* vol **129**, pp 31-40.
- Ben Moussa B (1998). Analyse numérique de méthodes particulières de type SPH pour les lois de conservation. PhD-Thesis INSAT département de génie mathématique et modélisation, defended on 20 January 1998.
- Ben Moussa B (2000). Measure-valued solutions for conservation laws with boundary conditions : application to convergence analysis of SPH method. Submitted to *SIAM Journal On Numerical Analysis*.
- Ben Moussa B (2000). Renormalized particle method for conservation laws : methodology, convergence analysis and application to Euler equations. *In preparation*.
- Benz W (1989). Smooth Particle Hydrodynamics : a Review. *Harvard-Smithsonian Center for Astrophysics, Preprint* vol **2884**.
- Benz W and Asphaug A (1993). Impact Simulations with Fracture : I. Methods and Tests. *Icarus* vol **107**, pp 98-116.

- Bicknell G V (1991). The equations of motion of particles in smoothed particle hydrodynamics. *SIAM J. Sci. Stat. Comput.* **vol 12 number 5**, pp 1198-1206.
- Dilts (1999). Moving-least-square-Particle Hydrodynamics I : Consistency and stability. *Inter. Jour. Numer. Meths. Engineer.* **vol 44**, pp 1115-1155.
- Gingold R A and Monaghan J J (1977). Smoothed particle hydrodynamics, theory and application to non-spherical stars. *Mon. Not. Roy. Astr. Soc.* **vol 181**, pp 375-389.
- Gingold R A and Monaghan J J (1983). Shock simulation by the particle method S.P.H. *J. Comput. Phys.* **vol 52**, pp 374-389.
- Godunov S K (1959). A Finite Difference Method for the Computation of Discontinuous Solutions of the Equations of Fluid Dynamics. *Mat. Sb.* **47**, pp 357-393.
- Hernquist L and Katz N (1989). TREESPH : a unification of SPH with the hierarchical tree method. *The Astronomy J. S. S.* **vol 70**, pp 419-446.
- Johnson G R and Beissel S R (1996). Normalized Smoothing functions for impact computations. *Int. Jour. Num. Methods Eng.* **vol 39**, pp 2725-2741.
- Liu W K, Li S and Belytschko T (1997) Moving least square reproducing kernel method, part I : Methodology and convergence. *Computer Methods in Applied Mechanics and Engineering*, **vol 143**, pp 422-433.
- Lucy L (1977) A numerical approach to the testing of the fission hypothesis. *Astronomy Journal* **vol 82**, pp 1013- .
- Monaghan J J (1995). Simulating Gravity Currents with SPH. III Boundary Forces. *Report 28*.
- Osher S (1984). Riemann solvers, the entropy condition and difference approximations. *Siam Jour. Num. Anal* **vol 21 number 2**, pp 217-235.
- Otto F (1996). Initial-boundary value problem for a scalar conservation law. *C. R. Acad. Sci. Paris* **vol 322 number I**, pp 729-734.
- Randles R W and Libertsky L D (1996). Smoothed Particle Hydrodynamics, Some recent improvements and Applications. *Comp. Meth. Appli. Mech. Eng.* **vol 139**, pp 375-408.
- Raviart P A (1985). An analysis of particle methods. *Num. method in fluid dynamics*, F. Brezzi ed. *Lecture Notes in Math.* **vol 1127**, Berlin, Springer.
- Special issue Gridless methods. *Comput. methods Appl. Mech. Engrg.* **vol 139**.
- Szepessy A (1989). Measure-valued solutions to conservation laws with boundary conditions. *Math. Arch. Rat. Mech. Anal.*, pp 181-193.
- Vila J P (1999). On particle weighted methods and Smooth Particle Hydrodynamics. *Math. Models and Methods in Applied Sciences* **vol 9 number 2**, pp 161-209.

APPLICATION OF WAVE-PROPAGATION ALGORITHM TO TWO-DIMENSIONAL THERMOELASTIC WAVE PROPAGATION IN INHOMOGENEOUS MEDIA

A. BEREZOVSKI

*Department of Mechanics and Applied Mathematics,
Institute of Cybernetics at Tallinn Technical University,
Akadeemia tee 21, 12618, Tallinn, Estonia
Email: Arkadi.Berezovski@cs.ioc.ee*

AND

G. A. MAUGIN

*Université Pierre et Marie Curie,
Laboratoire de Modélisation en Mécanique,
UMR 7607, Tour 66, 4 Place Jussieu, Case 162,
75252, Paris Cédex 05, France
Email: GAM@ccr.jussieu.fr*

Abstract.

The system of equations for thermoelastic wave propagation in an inhomogeneous medium is not in the conservation form. Nevertheless, a modification of the wave propagation algorithm for conservation laws is successfully used for the numerical simulation. The modification is made in two ways. First, the algorithm is represented in terms of contact quantities to provide the satisfaction of the thermodynamic consistency conditions between adjacent cells. As usual, the finite volume Godunov scheme is improved by introducing correction terms to obtain high resolution results. Secondly, a composite scheme is obtained by application of the Godunov step after each three second-order Lax-Wendroff steps. The multidimensional motion is accomplished by including into consideration the transverse fluctuations. At last, the elimination of source terms is made following the method of balancing source terms after independent solution of the heat conduction equation. Results of computation for certain test problems show the efficiency and physical consistency of the algorithm.

1. Introduction

The classical theory of thermoelasticity is governed by the balance law of linear momentum, which can be presented in the small strain approximation in the absence of body forces as (Germain, 1973)

$$\rho_0 \frac{\partial v_i}{\partial t} - \frac{\partial \sigma_{ij}}{\partial x_j} = 0, \quad (1)$$

and the Duhamel-Neumann relation between the Cauchy stress tensor, σ_{ij} , and the strain tensor, ε_{ij} , which uses two elastic parameters, λ, μ , (or Young's modulus, E , and Poisson's ratio, ν) and linear expansion coefficient, α , (Nowacki, 1986)

$$\sigma_{ij} = [\lambda \varepsilon_{kk} - \alpha(3\lambda + 2\mu)(T - T_0)]\delta_{ij} + 2\mu\varepsilon_{ij}. \quad (2)$$

Here ρ_0 is the density of material, v_i are components of velocity vector, t is time, x_i are spatial coordinates, T_0 is the reference temperature, δ_{ij} is the Kronecker delta. Deformation rate equation follows from the time derivative of the Duhamel-Neumann relation (2)

$$\frac{\partial \sigma_{ij}}{\partial t} = \lambda \frac{\partial v_i}{\partial x_i} \delta_{ij} + \mu \left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right) - \alpha(3\lambda + 2\mu) \frac{\partial T}{\partial t} \delta_{ij}. \quad (3)$$

In the thermal stress approximation, the system of equations (1), (3) is complemented by the heat conduction equation (Nowacki, 1986)

$$\frac{\partial T}{\partial t} = k \nabla^2 T, \quad (4)$$

where k is the heat conductivity.

From the mathematical point of view, the problem is to find a solution of the system of equations (1), (3)-(4) with corresponding initial and boundary conditions. In general, the initial and boundary conditions are too complicated to obtain an analytical solution of the considered problem. Therefore, an efficient and robust numerical method is needed.

It should be noted that the heat conduction equation (4) can be solved independently. Having this solution, one can represent the system of equations (1), (3) in a general matrix form of hyperbolic conservation laws with given source terms

$$\frac{\partial q}{\partial t} + A \frac{\partial q}{\partial x} + B \frac{\partial q}{\partial y} = \psi. \quad (5)$$

The construction, analysis, and implementation of approximate solutions to conservation laws were the major focus of an enormous amount of

activity in recent decades (LeVeque, 1990), (Godlewski and Raviart, 1996). Modern algorithms were developed that achieve high resolution, stability and efficiency of numerical schemes. Nevertheless, it is well known that monotone (or positive) approximations are at most first-order accurate (Godunov, 1959). The lack of monotonicity for higher order methods is reflected by spurious oscillations in the vicinity of jump discontinuities. The common approach to suppress these oscillations is the introduction of certain limiter functions. To obtain a more physically consistent algorithm, the governing equations are formulated in the second part of the paper directly in terms of parameters of the discrete elements, which represent a continuum in numerical schemes. It appears that the recently proposed wave propagation algorithm (LeVeque, 1997), briefly remembered in the third part of the paper, can be easily reformulated in terms of contact quantities, and, in essence, is thermodynamically consistent. However, the originally advised limiters do not allow controlling the satisfaction of this consistency. Instead, the recently proposed idea of composite schemes (Liska and Wendroff, 1998) is used. A variant of the composite scheme is applied to calculations of thermoelastic wave propagation in inhomogeneous media. The obtained results are presented in the fourth part of the paper.

2. Thermomechanics of discrete systems

Since numerical schemes deal with discrete elements representing a medium, we will try to formulate the governing equations in terms of parameters of the discrete elements. This will be done by means of the thermodynamics of discrete systems.

The main idea of the thermodynamics of discrete systems is the extension of thermodynamic state space by virtue of so-called contact quantities for the description of non-equilibrium states of a system (Muschik, 1993). By definition a discrete system is a domain $G \in \mathbf{R}^3$ separated by a partition ∂G from its environment G^* . The interaction between G and G^* is described by exchange quantities. In a Schottky system *per se*, this interaction consists of heat, work and mass exchanges. For instance, considering heat exchange \dot{Q} , the so-called *contact temperature* Θ is defined by the following inequality:

$$\dot{Q} \left(\frac{1}{\Theta} - \frac{1}{T^*} \right) \geq 0. \quad (6)$$

for vanishing work and mass exchange rates. Here T^* is the thermostatic temperature of the equilibrium environment.

The contact temperature is the dynamical analogue to the thermostatic temperature. The interpretation of the contact temperature is as follows: From (6) follows that the heat exchange \dot{Q} and the bracket have always the

same sign. If we now presuppose that there exists exactly one equilibrium environment for each arbitrary discrete system for which the net heat exchange between them vanishes, then the defining inequality (6) determines the contact temperature Θ of the system as that thermostatic temperature T^* of the system's environment for which this net heat exchange vanishes.

The contact quantities provide the complete thermodynamic description of non-equilibrium states of a separated discrete system. It should be noted that the values of the defined contact quantities differ from the values of usual bulk parameters even in the case of local equilibrium. For interacting elements, the values of bulk and contact quantities of adjacent elements are connected additionally by thermodynamic consistency conditions (Berezovski, 1997), which will be specified later.

To extend the concepts of the thermodynamics of discrete systems onto the thermoelastic case, we divide the body in a finite number of identical elements. The state of each element is identified with the thermodynamic state of a discrete system associated with the element. It is assumed that each element is in local equilibrium. In the linear thermoelastic case, the state space includes the strain tensor ε_{ij} , which allows us to define additionally the contact dynamic stress tensor Σ_{ij}

$$\frac{\partial \varepsilon_{ij}}{\partial t}(\Sigma_{ij} - \sigma_{ij}^*) \geq 0, \quad (\dot{Q} = 0), \quad (7)$$

where σ_{ij}^* is the Cauchy stress tensor in the environment.

The connection between the bulk and contact quantities is established by means of the integral balance law of the linear momentum

$$\frac{\partial}{\partial t} \int_V \rho_0 v_i dV = \int_{\partial V} \Sigma_{ij} n_j dA, \quad (8)$$

the integral form of the deformation rate equation

$$\frac{\partial}{\partial t} \int_V \sigma_{ij} dV = \int_{\partial V} (2\mu H_{ijk} n_k + \lambda \delta_{ij} V_k n_k) dA - \int_V \alpha(3\lambda + 2\mu) \delta_{ij} \frac{\partial T}{\partial t} dV, \quad (9)$$

and the integral form of the heat conduction equation

$$\frac{\partial}{\partial t} \int_V T dV = \int_{\partial V} k n_j \nabla_j T dA. \quad (10)$$

Here $H_{ijk} = 1/2(\delta_{ik} V_j + \delta_{jk} V_i)$, V_i are components of the contact deformation velocity, which is associated with the contact stress tensor, n_j are components of outward normal vector.

At the first look, the integral balance laws could be obtained by direct integration of equations (1), (3)-(4) over the volume of the element. It

should be noted, however, that the contact quantities could not appear in the case of the formal integration. As it was noted, the bulk and contact quantities are additionally connected by the thermodynamic consistency conditions (Berezovski, 1997)

$$\begin{aligned} \sigma_{ij}^{(1)} + \Sigma_{ij}^{(1)} - T^{(1)} \left(\frac{\partial \sigma_{ij}^{(1)}}{\partial T} \right)_{\varepsilon_{ij}} - \Theta^{(1)} \left(\frac{\partial \Sigma_{ij}^{(1)}}{\partial T} \right)_{\varepsilon_{ij}} &= \\ = \sigma_{ij}^{(2)} + \Sigma_{ij}^{(2)} - T^{(2)} \left(\frac{\partial \sigma_{ij}^{(2)}}{\partial T} \right)_{\varepsilon_{ij}} - \Theta^{(2)} \left(\frac{\partial \Sigma_{ij}^{(2)}}{\partial T} \right)_{\varepsilon_{ij}}, \end{aligned} \quad (11)$$

for each pair of adjacent cells numbered 1 and 2 (superscripts within parentheses). The thermodynamic consistency conditions express the continuity of derivatives of the internal energy at constant temperature.

The integral balance laws (8)–(10) lead to the finite volume approximation in a natural way, and the main problem is to determine the values of contact quantities. These values will be determined by means of a modified wave-propagation algorithm.

3. Two-dimensional wave-propagation algorithm

The modification of the wave-propagation algorithm (LeVeque, 1997) will be presented on the example of linear thermoelasticity in an inhomogeneous medium in two space dimensions. First, the homogeneous system of equations (5) is considered, which corresponds to the pure elastic case

$$\frac{\partial q}{\partial t} + A(x, y) \frac{\partial q}{\partial x} + B(x, y) \frac{\partial q}{\partial y} = 0. \quad (12)$$

The algorithm for the solution of these equations can be represented in terms of contact quantities

$$q_{mn}^{k+1} = q_{mn}^k - \frac{\Delta t}{\Delta x} (AQ_{mn}^+ - AQ_{mn}^-) - \frac{\Delta t}{\Delta y} (BQ_{mn}^+ + BQ_{mn}^-), \quad (13)$$

where subscripts mn indicate the placement of discrete elements in the 2D grid and superscript k denotes time step. Contact quantities AQ_{mn}^\pm and BQ_{mn}^\pm include the fluctuations arising from Riemann problems in the x - and y -directions, respectively, and second-order correction terms and transverse fluctuations in the same way as in (LeVeque, 1997). It should be noted that these definitions of the contact stresses satisfy the thermodynamic consistency conditions (11) in their isothermal form.

It is well known that the Lax-Wendroff scheme produces oscillations behind discontinuities (Liska and Wendroff, 1998). The usual way to reduce spurious oscillations is to introduce limiter functions in order to modify the second-order corrections near discontinuities. However, the application of limiters still seems to be more of an art than a science. Moreover, the fulfilling of the thermodynamic consistency conditions cannot be controlled in this case. It seems that the recently proposed composite schemes (Liska and Wendroff, 1998) are more convenient, because of using filters that are consistent with differential equations. Here the composite scheme is obtained by application of the Godunov step after each three second-order Lax-Wendroff steps. Obviously, the thermodynamic consistency conditions remain satisfied at each step.

In the thermoelastic case, the temperature equation (10) in two dimensions is solved by means of a simple algorithm in an inhomogeneous medium (Berezovski and Rosenblum, 1996). According to the method of balancing source terms (LeVeque, 1998), the modified bulk velocities both in x - and y -directions

$$(v_1)_{mn}^{\pm} = (v_1)_{mn} \pm \frac{\alpha_{mn}(3\lambda_{mn} + 2\mu_{mn})}{2(\lambda_{mn} + 2\mu_{mn})} \left(\frac{\partial T}{\partial t} \right)_{mn} \Delta x, \quad (14)$$

$$(v_2)_{mn}^{\pm} = (v_2)_{mn} \pm \frac{\alpha_{mn}(3\lambda_{mn} + 2\mu_{mn})}{2(\lambda_{mn} + 2\mu_{mn})} \left(\frac{\partial T}{\partial t} \right)_{mn} \Delta y, \quad (15)$$

are used for the calculation of normal dynamic stresses and corresponding contact velocities, which allows to eliminate the source terms. The first-order Godunov method as well as transverse propagation, second-order correction, and composition are applied then as above.

4. Numerical results

Figure 1 shows a snapshot of the mechanical trace (normal stress) of a thermoelastic wave inside a medium (2D computations) with laterally continuously varying properties. The properties of the medium, corresponding to pure aluminium at the left boundary ($c_p = 6420 \text{ m/s}$, $c_s = 3040 \text{ m/s}$, $\rho = 2700 \text{ kg/m}^3$), are varied linearly to those of copper at the right boundary ($c_p = 4560 \text{ m/s}$, $c_s = 2600 \text{ m/s}$, $\rho = 8960 \text{ kg/m}^3$) (Fig.2). The wave was excited by a purely thermal shock ($\Delta T = 100K$) at a part of the bottom boundary ($20 < x < 80$), as it is shown by narrow black rectangle in Figure 2, all other boundaries being stress free. The wave front curves as it propagates, due to lateral inhomogeneity.

The proposed method for numerical simulation of thermoelastic wave propagation is based on recent developments in high-resolution schemes for conservation laws. The combination of wave propagation method with

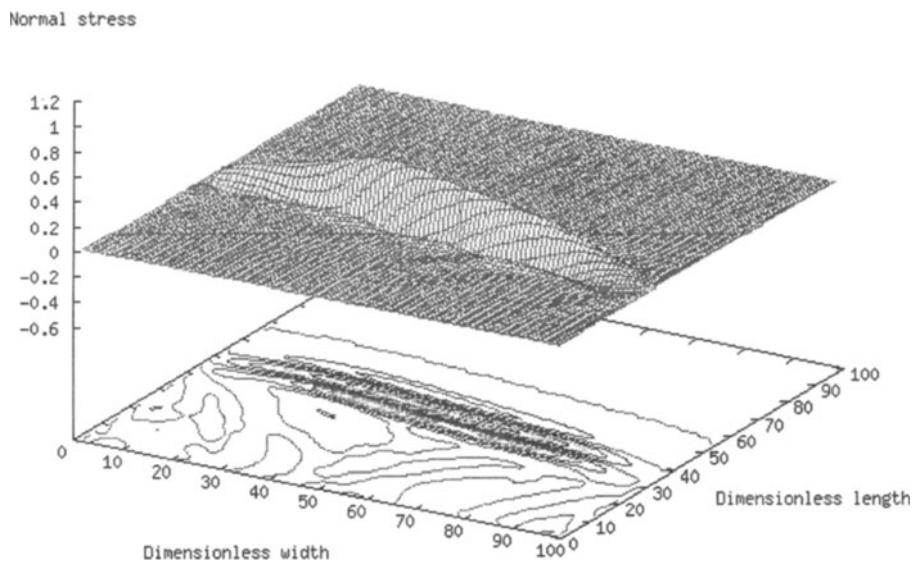


Figure 1. Thermoelastic wave propagation inside a medium with laterally varying properties, 80 time steps, Courant number 0.9.

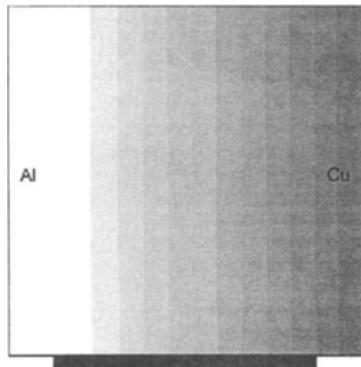


Figure 2. Continuous variation of material properties inside the computational domain

filtering of the second order accurate scheme by the procedure that is consistent with the differential equations allows us to use the advantages of both approaches. In particular, the method is successfully applied to the inhomogeneous thermoelastic case, where differential equations are in a non-conservative form with source terms. The distinctive feature of the al-

gorithm is the satisfaction of the thermodynamic consistency conditions at each step.

Acknowledgements

Support of the Estonian Science Foundation under contract No.3203 and of BMBF (Germany) under contract 03N9005_2 (A.B.) and of the European Network TMR. 98-0229 on "Phase transitions in crystalline substances" (G.A.M.) is gratefully acknowledged.

References

- Germain P (1973). Cours de Mécanique des Milieux Continus. v. 1, Masson.
 Nowacki W (1986). Thermoelasticity. Pergamon-PWN.
 LeVeque R J (1990). Numerical Methods for Conservation Laws. Birkhäuser Verlag.
 Godlewski E and Raviart P A (1996). Numerical Approximation of Hyperbolic Systems of Conservation Laws. Springer.
 Godunov S K (1959). A Finite Difference Method for the Computation of Discontinuous Solutions of the Equations of Fluid Dynamics. *Mat. Sb.* **47**, pp 271–306.
 LeVeque R J (1997). Wave Propagation Algorithms for Multidimensional Hyperbolic Systems. *J. Comput. Phys.* **131**, pp 327–353.
 Liska R and Wendroff B (1998). Composite Schemes for Conservation Laws. *SIAM J. Numer. Anal.* **35**, pp 2250–2271.
 Muschik W (1993). Fundamentals of Non-Equilibrium Thermodynamics. Non-Equilibrium Thermodynamics with Application to Solids, pp 1–63. Muschik W (Editor). Springer.
 Berezovski A (1997). Continuous Cellular Automata for Simulation of Thermoelasticity, *Proc. Estonian Acad. Sci. Phys. Mat.* **46**, pp 5–13.
 Berezovski A and Rosenblum V (1996). Thermodynamic Modelling of Heat Conduction, *Proc. Estonian Acad. Sci. Engin.* **2**, pp 196–208.
 LeVeque R J (1998). Balancing Source Terms and Flux Gradients in High-resolution Godunov Methods: the Quasi-steady Wave-propagation Algorithm, *J. Comput. Phys.* **148**, pp 346–365.

UNSTRUCTURED MESH SOLVERS FOR HYPERBOLIC PDES WITH SOURCE TERMS: ERROR ESTIMATES AND MESH QUALITY

M. BERZINS

*School of Computer Studies,
University of Leeds,
Leeds LS2 9JT, U.K.
Email: martin@scs.leeds.ac.uk*

AND

L.J.K. DURBECK
*Department of Computer Science,
University of Utah, USA.
Email: ldurbeck@cs.utah.edu*

Abstract. The solution of hyperbolic systems with stiff source terms is of great importance in areas such as atmospheric dispersion. The finite-volume approach used here for such problems employs Godunov-type methods, a sophisticated splitting approach for efficiency and adaptive tetrahedral meshes to provide the necessary resolution for physically meaningful solutions. This raises the issues of how to estimate the error for Godunov type methods and what is an appropriate mesh for such applications. A new mesh visualization and haptic-interface tool will be shown to help clarify this and its use illustrated for a model problem in three space dimensions.

1. Introduction

Unstructured triangular and tetrahedral meshes are widely used in engineering and scientific computing for solving problems via finite element and finite volume methods. At the same time Godunov methods are widely used in the solution of problems with hyperbolic parts (Godlewski and Raviart, 1996; Kröner, 1997; Toro, 1999). The intention here is to consider some of the issues that arise from combining these approaches when solving

problems such as the 3D advection reaction problem, taken from a model of atmospheric dispersion from a power station plume - a concentrated source of NOx emissions, (Tomlin, 1999). The photo-chemical reaction of this NOx with polluted air leads to the generation of ozone at large distances downwind from the source. An accurate description of the distribution of pollutant concentrations is needed over large spatial regions in order to compare with field measurement calculations. The complex chemical kinetics in the atmospheric model gives rise to sudden changes in the concentration of the chemical species in both space and time. These changes must be matched by changes in the spatial mesh and the timesteps if high resolution is required, (Tomlin, 1999). The effects of the plume interestingly causes levels of ozone to rise above the background levels at quite large distances downwind from the source of NOx. This application is modelled by the atmospheric diffusion equation in three space dimensions given by:

$$\frac{\partial c_s}{\partial t} + \frac{\partial u c_s}{\partial x} + \frac{\partial v c_s}{\partial y} + \frac{\partial w c_s}{\partial z} = D + R_s + E_s - \kappa_s c_s, \quad (1)$$

where c_s is the concentration of the s 'th compound, u, v and w , are wind velocities and κ_s is the sum of the wet and dry deposition velocities. E_s describes the distribution of emission sources for the s 'th compound and R_s is the chemical reaction term which may contain nonlinear terms in c_s . D is the diffusion term. For n chemical species a set of n coupled partial differential equations (p.d.e's) is formed.

The solution techniques employed consist of time integration methods specially designed for explicit convection and implicit source terms handled by using a very efficient Gauss-Seidel iteration. Finite volume cell-vertex and cell-centred Godunov-type schemes (Godunov, 1959; Van Leer, 1984) are both used for space discretization. For this atmospheric diffusion model, the meshes and means of obtaining them are described in (Johnson, 1998; Speares, 1997). The advantage of the Godunov-type methods based on upwinding and approximate Riemann problems is that it is possible to preserve positivity of the solution - a key requirement for reacting flow problems. Mesh adaptation using h refinement, even based on simple gradient information gives dramatically improved solutions, see (Tomlin, 1999) but raises the issue of whether or not the mesh is appropriate for all the species.

The only sure way of knowing whether or not the mesh is appropriate is to use error indicators and to understand how the error depends on both the solution and on element shape, preferably by visualization. It is hard to visualize all the mesh elements in a full 3D mesh display and it is difficult to comprehend fully the myriad of element shapes and sizes, see Figure 1. The combined haptic and visual interface of (Durbeck, 1999) has been designed to overcome the daunting task of finding "bad" tetrahedra in a

visually complex mesh. In the remainder of this paper an error indication approach will be outlined and used in combination with the visual interface to find bad tetrahedra in a 3D adaptive unstructured mesh.

2. Adaptive Numerical Solution Techniques.

In order to illustrate the approaches consider the simple 3D advection equation

$$U_t + aU_x + bU_y + cU_z = 0 \quad (2)$$

The numerical method employed is a first order accurate, conservative cell-centred finite volume scheme. The numerical solution in element i at time t_n is denoted by u_i^n , and is an approximation to the exact element averaged volume integral of the solution, (Speares, 1997), over V_i the volume of element i , and is usually regarded as being valued at the element centroid for cell centred schemes. The numerical solution at the next time level t^{n+1} may be written as:

$$u_i^{n+1} = u_i^n - \delta t F_i(t_n, \underline{U}) \quad \text{where} \quad F_i(t_n, \underline{U}) = \frac{1}{V_i} \sum_{k=0}^3 A_k \mathbf{F}_k \cdot \mathbf{n}_k \quad (3)$$

and where the sum is over the k faces of the element i . The \mathbf{n}_k are the outward face unit normal vectors and A_k the face areas. The fluxes \mathbf{F}_k represent the numerical flux function for each element face, termed the element face fluxes, and are determined by the scheme. In the case of the Godunov scheme these element face numerical fluxes are constructed from the solution of the local element Riemann Problem (RP) at each element face, see (Godunov, 1959; Speares, 1997). In the calculations described here both first and second order schemes are used, (Van Leer, 1984).

A standard method for choosing the timestep in the numerical solution of p.d.e.s is to use a CFL condition. Although such a condition may ensure stability it may be imprecise as an accuracy control, particularly when complex chemistry source terms are present in the p.d.e. problem. It is important to use an error control which reflects the spatial and temporal contributions to the error incurred.

The global error in the numerical solution can be expressed as the sum of the spatial discretization error, and the global time error, Efficient time integration requires that the spatial and temporal are roughly the same order of magnitude. The need for spatial error estimates to be unpolluted by temporal error requires the spatial error to be the larger of the two errors. One approach for achieving this is described by, (Berzins, 1995), who balances the spatial and temporal errors by controlling the local time error to be a fraction of the local growth in the spatial discretization error.

The local-in-time spatial error, $\hat{e}(t_{n+1})$, for the timestep from t_n to t_{n+1} is defined as the spatial error at time t_{n+1} given the assumption that the spatial error, $e(t_n)$, is zero. The error $\hat{e}(t_{n+1})$ is estimated by the difference between the computed solution and the first-order solution which satisfies an o.d.e. system given by

$$\dot{v}_{n+1}(t) = \underline{F}_N^*(t, v_{n+1}(t)), \quad (4)$$

where $v_{n+1}(t_n) = \underline{V}(t_n)$ and where $\underline{F}_N^*(.,.)$ is obtained by using the limiter function $\Phi(.)$ in the spatial discretization method, (zero for a first order scheme), to be that for a second order scheme. The local-in-time space error is estimated by

$$\hat{e}(t_{n+1}) = \underline{V}(t_{n+1}) - v_{n+1}(t_{n+1}) \quad (5)$$

and is computed by applying, say, the forward Euler method method to equation (4), thus giving (with one evaluation of $\underline{F}_N^*(.,.)$ per timestep):

$$\hat{e}(t_{n+1}) = \delta t [\underline{F}_N(t_n, \underline{V}(t_n)) - \underline{F}_N^*(t_n, \underline{V}(t_n))]. \quad (6)$$

While reliable error estimators for finite volume unstructured mesh solvers exist for simple problems, e.g. (Kroner, 2000), there are no such estimators for problems with complex source terms. Consequently, we are forced to rely instead on local error indicators such as those described above. For problems without source terms the estimate of Kroner and Ohlberger may be adapted to estimate this local in time space error. Let $\hat{e}(t)$ be the local in time spatial error computed on a timestep then combining the estimates of Corollary(2.14) of (Kroner, 2000) and the ideas of (Berzins, 1995) gives

$$\int \int \int_V \hat{e}(t_{n+1}) d\tau = a \delta t h^2 Q + 2\sqrt{b c \delta t h^2 Q} \quad (7)$$

where a, b, c are constants, see (Kroner, 2000) and for an evenly spaced mesh with spacing h and timestep δt the value of Q is given by

$$Q = \sum_{j \in NT} h |u_j^{n+1} - u_j^n| + L \sum_{E \in NT} (\delta t + h) |u_j^n - u_l^n|$$

where L is a constant, u_j^n is the solution value associated with the j th tetrahedron out of a mesh of NT tetrahedra with edges $E \in NT$ at time t_n . The important feature of this error estimator is that, apart from the constants, the only solution information used consists of solution jumps across faces i.e. $u_j^n - u_l^n$ and solution changes in time $u_j^{n+1} - u_j^n$ on a particular tetrahedron. However the estimate does not reflect the fact that

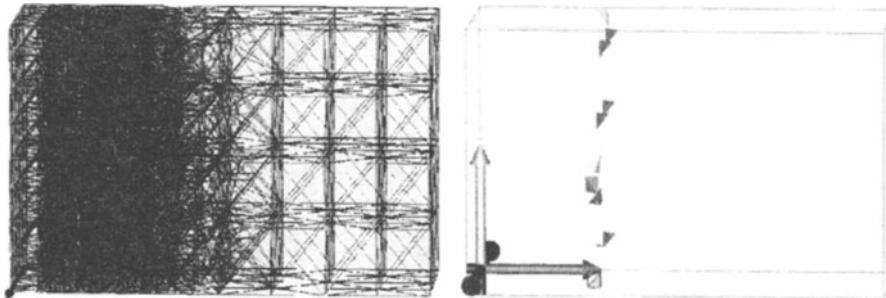


Figure 1. (a) Wire frame mesh and (b) visualization of poor elements

face orientation in flow problems is critical as error may not be convected through faces aligned with the flow.

2.1. A SIMPLE 1D ADVECTION EQUATION EXAMPLE

Consider the advection of a simple one-dimensional discontinuity moving from left to right in a 3D channel, as defined by equation (2) with $a = 1, b = c = 0$. A typical example of a 3D unstructured mesh with 16,000 elements. is shown in Figure 1a. The mesh is shown in wire frame, with all the nodes and their attachments shown, and has been refined about the discontinuity.

It is of interest to evaluate the error estimation approaches on a similar simple 1D version of Problem 4 (linear advection) in (Berzins, 1995). The local in time error being measured about halfway across the domain. Figure 2 shows the spatial distribution of the error $\hat{e}(t)$ with the solid line being the true error and the values * showing the error estimate defined by equation (6) and the values + showing the time local error. The peaks in the error graph occur where the scheme smooths the top and bottom of the discontinuity. The figure shows that the error estimator does a good job of estimating the structure and the magnitude of the local-in-time spatial error, particularly as the cfl number is reduced, (Berzins, 1995). of array Table 1 shows the values of the error indicators for different values of the CFL number. The results show that both error estimators do a good job of estimating the L1 norm of the error growth over a single timestep.

3. Visual Mesh Quality Analysis

Error indicators for the simple advection equation example were investigated visually with a user interface developed by (Durbeck, 1999). Durbeck's interface both serves as a visual debugger for the advection mesh and

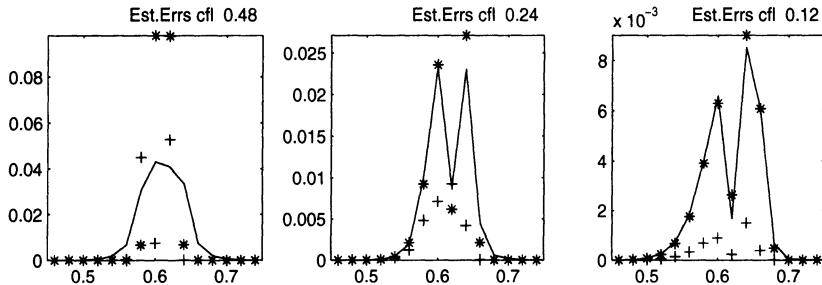


Figure 2. Graphs of local space and time errors

TABLE 1. Comparison of L1 error norm for error indicators

CFL Number	0.96	0.48	0.24	0.12	0.06	0.03
True \hat{e}	1.17e-2	3.35e-2	1.46e-3	6.12e-4	2.81e-4	1.33e-4
Berzins eqn(6)	4.53e-2	4.18e-2	1.42e-3	6.23e-4	2.73e-4	1.26e-4
Kroner eqn(7)	1.15e-1	8.13e-2	2.55e-3	8.42e-4	2.85e-4	9.90e-5

presents analytic information about the mesh geometry and error indicators so that the user can deduce the potential causes for poor quality elements. A view of the mesh, reduced via the debugger to its poorest elements, is shown in Figure 1b. The elements are displayed as solids, with lighting and shading effects. The color assigned to each tetrahedron corresponds with its relative error indicator value. Comparison of Figure 1a with 1b indicates that the the poorest elements are roughly aligned and occur near the leading edge of the area refined to represent the discontinuity.

The visual debugger also provides closeups used for analysis of a specific error indicator. The worst element depicted in Figure 1b is shown in closeup view in Figure 3, along with all neighbouring elements which may contribute to its error value. The information presented in this view is intended to correspond closely with the error indicator: in our case, an element's poor quality can be a combination of its shape, orientation and precise vertex locations within the mesh. The same inquiry continues outward to its neighbours and, to a lesser extent, the next level outward as well, as they contribute to the element in question. The worst element and its direct neighbours are displayed as shaded solids and the (less important) next level outward in wire frame. Graphical representations of each element are annotated with the element number, error indicator value, and solution value. Color also provides relative error indicator values. As shown in Figure 3b and 3c, the closeup can be rotated about, and exploded outward from, the central element in order to better view all the tetrahedra.

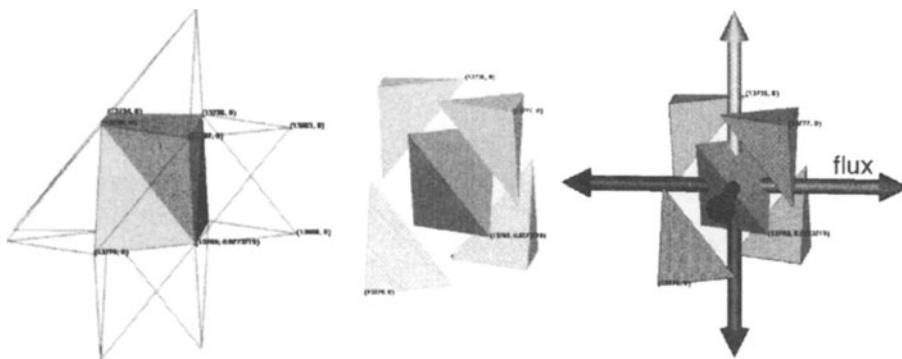


Figure 3. (a) In place (b) exploded (c) rotated closeup
views of worst element and its neighbours

As seen in Figure 3, The two main contributors to the central element's high error value appear to be its orientation, which causes two faces to be close to perpendicular to the flux, and the wedge shape of the element, which causes these two faces to be relatively wide. Thus the user has been able to easily identify the cause of poor mesh quality in a complex three dimensional meshes of the type described in Section 1.

References

- Berzins M (1995) Temporal Error Control in the Method of Lines for Convection Dominated Equations. *SIAM J. on Sci. Comput.* **16**, pp.558-580.
- Durbeck L J K (1999) Contrast Displays: A Haptic and Visual Interface Designed Specifically for Mesh Quality Analysis. M.Sc. Thesis Univ. of Utah.
- Godunov S K (1959). A Finite Difference Method for the Computation of Discontinuous Solutions of the Equations of Fluid Dynamics. *Mat. Sb.* **47**, pp 357-393.
- Godlewski E and Raviart P A (1996). Numerical Approximation of Hyperbolic Systems of Conservation Laws. Springer.
- Johnson C R, Berzins M, Zhukov L, and Coffey R (1998) SCIRun: Applications to Atmospheric Diffusion Using Unstructured Meshes. *Numerical Methods for Fluid Dynamics VI*. Editor M. J. Baines. ICFD, Oxford Univ. pp111-122.
- Kröner D (1997). Numerical Schemes for Conservation Laws. Wiley Teubner.
- Kröner D and Ohlberger M (2000) A posteriori error estimates for upwind finite volume schemes for nonlinear conservation laws in multi-dimensions." *Mathematics of Computation*, **69**, pp25-39.
- Speares W and Berzins M (1997) A 3D Unstructured Mesh Adaptation Algorithm for Time-Dependent Shock-dominated Problems. *International Journal for Numerical Methods in Fluids* **25** pp81-104.
- Tomlin A S, Ghorai S, Hart G and Berzins M (1999) 3-D Adaptive Unstructured Meshes for Air Pollution Modelling. *Environmental Management and Health* **10/4** 267-274.
- Toro E F (1999). Riemann Solvers and Numerical Methods for Fluid Dynamics. Second Edition, Springer-Verlag.
- van Leer B (1984). On the Relation Between the Upwind-Differencing Schemes of Godunov, Engquist-Osher and Roe. *SIAM J. Sci. Stat. Comput.* **5**, pp 1-20.

CONSTANCY PRESERVING, CONSERVATIVE METHODS FOR FREE-SURFACE MODELS

L. BONAVENTURA

*Dipartimento di Ingegneria Civile ed Ambientale
Università di Trento
Mesiano di Povo, 38050 (TN), Italy
e-mail: bonavent@ing.unitn.it*

AND

E.S. GROSS

*Dipartimento di Ingegneria Civile ed Ambientale
Università di Trento
Mesiano di Povo, 38050 (TN), Italy
email: e.s.gross@home.com[†]*

Abstract

A constancy preserving formulation of conservative scalar transport schemes is presented in the framework of the semi-implicit discretization of free-surface flow equations. Consistency between the discretized free-surface equation and the discretized scalar transport equation is shown to be necessary for constancy preservation. This property is assured to hold for a wide range of advection schemes, given a specific and widely applied discretization of the free-surface equation. The practical relevance of the consistency with continuity condition is then demonstrated by various numerical tests.

1. Introduction

The monotonicity properties of conservative numerical schemes for the basic advection equation

[†]Permanent address:
1777 Spruce Street
Berkeley, CA 94709 USA
email: e.s.gross@home.com

$$\frac{\partial S}{\partial t} + \frac{\partial(US)}{\partial x} + \frac{\partial(VS)}{\partial y} = 0 \quad (1)$$

have been widely investigated. By Godunov's theorem, monotonicity preservation requires numerical schemes to be either first-order or nonlinear. Sufficient conditions for monotonicity preservation, such as the property of being total variation diminishing or l_1 contracting, have also been studied, see e.g. (R.LeVeque, 1990) , (E.F.Toro, 1997). However, the TVD property only extends to first-order, fully multidimensional schemes, see e.g. (J.B.Goodman and R.LeVeque, 1985). An even weaker property, which is however desirable of any numerical scheme for equation (1), is that an initially uniform scalar field remains uniform in the absence of sources and sinks. Although trivially true for conservative schemes on domains with cell volumes that are constant in time, this property does not always follow for conservative discretizations of the analogs of (1) in the case of free-surface flows. In fact, for free-surface flows, this property holds only if the conservative scheme employed is consistent with the discretization of the free-surface continuity equation, in a sense to be specified later. The purpose of this communication is to show that large errors may occur if this consistency is violated: even when monotone or TVD schemes are applied, initially constant scalar fields may develop unphysical maxima, especially when the free-surface changes abruptly or collapses to the bottom of the computational domain. For definiteness, only the discretization of the shallow water equations is considered here, but analogous behaviour can also be displayed in three-dimensional free-surface flows. Three-dimensional flows will be considered in an extended version of this paper. In the case of a specific semi-implicit discretization of the continuity equation, which has been widely applied to shallow water problems in hydraulics and in estuarine simulations, see e.g. (V.Casulli, 1990), (V.Casulli and E.Cattani, 1994), (E.S.Gross, V.Casulli, L.Bonaventura and J.R.Koseff, 1998), the corresponding consistent discretization of the scalar transport equation is shown. As a result, it appears that this property is essential to grant effective accuracy of a given numerical algorithm for free-surface dynamical simulation of transport phenomena. The relevance of consistency with continuity has also been pointed out in (S.Lin and R.B.Rood, 1996), (E.S.Gross, J.R.Koseff and S.G.Monismith, 1999), (R. LeVeque, 1996). It is also interesting to remark that the consistency with the discretized continuity equation is essential when linking conservative schemes to their advective analogues, see e.g. (G.S.Stelling, H.W.J.Kernkamp and M.M.Laguzzi, 1998), and that a reformulation of this property also plays a key role in the monotonicity proof presented in (X.D.Liu, 1993).

2. Consistency with continuity and constancy preservation for the shallow water equations

The continuity equation and the basic mass conservation law for advected solute can be written in the case of the two-dimensional shallow water equations as

$$\frac{\partial H}{\partial t} + \frac{\partial(UH)}{\partial x} + \frac{\partial(VH)}{\partial y} = 0 \quad (2)$$

$$\frac{\partial(SH)}{\partial t} + \frac{\partial(USH)}{\partial x} + \frac{\partial(VSH)}{\partial y} = 0. \quad (3)$$

Here, it is assumed that $H(x, y, t) = h(x, y, t) + d(x, y)$ is the total fluid depth, where h represents the height of the free-surface above a given reference level and d is the depth between this level and the lower boundary of the considered domain. $U(x, y, t)$ and $V(x, y, t)$ are the vertically averaged horizontal velocity components, $S(x, y, t)$ is the concentration of some passively transported solute. These equations are to be coupled to the corresponding momentum equations, either in conservative or advective form.

The condition of consistency with continuity (CWC) will be defined along the lines of (S.Lin and R.B.Rood, 1996) as follows: *a discretization of the advection equation (3) is consistent with continuity if, given a general flow field and a constant initial datum S_0 , it degenerates to the discretization of the continuity equation (2).* This definition applies whether or not the cell volumes vary in time and space. Clearly, a scheme satisfying the CWC condition will leave constant initial fields constant.

For definiteness, a specific discretization of (2) is now introduced, but all the considerations that follow will also apply if any other discretization is chosen. A widely applied semi-implicit discretization of (2) on an Arakawa C-type staggered grid, see e.g. (V.Casulli, 1990), (V.Casulli and E.Cattani, 1994), is given by

$$H_{i,j}^{n+1} = H_{i,j}^n - \frac{\Delta t}{\Delta x} \left(U_{i+\frac{1}{2},j}^{n+\frac{1}{2}} H_{i+\frac{1}{2},j}^n - U_{i-\frac{1}{2},j}^{n+\frac{1}{2}} H_{i-\frac{1}{2},j}^n \right) - \frac{\Delta t}{\Delta y} \left(V_{i,j+\frac{1}{2}}^{n+\frac{1}{2}} H_{i,j+\frac{1}{2}}^n - V_{i,j-\frac{1}{2}}^{n+\frac{1}{2}} H_{i,j-\frac{1}{2}}^n \right). \quad (4)$$

In the context of this discretization approach, the depths $d_{i\pm\frac{1}{2},j}$, $d_{i,j\pm\frac{1}{2}}$ are defined at the sides of each computational cell, while the free-surface height $h_{i,j}$ is defined at the cell center. This discretization is coupled in (V.Casulli, 1990) to the semi-implicit, semi-lagrangian discretization of the

momentum equations, thus resulting in an extremely efficient and widely applied numerical method. The values of water depth at the cell sides are computed here as

$$H_{i+\frac{1}{2},j} = h_{i+\frac{I+1}{2},j} + d_{i+\frac{1}{2},j} \quad H_{i,j+\frac{1}{2}} = h_{i,j+\frac{J+1}{2}} + d_{i,j+\frac{1}{2}},$$

where $I = -\text{sign}(u_{i+\frac{1}{2},j}^{n+\frac{1}{2}})$, $J = -\text{sign}(v_{i,j+\frac{1}{2}}^{n+\frac{1}{2}})$. This is similar to what has been proposed in (G.S.Stelling, H.W.J.Kernkamp and M.M.Laguzzi, 1998), where it is shown how this kind of definition can be used to achieve positivity of the total water depth under appropriate conditions on the time step. Furthermore, as it will be shown in a more extended version of this paper, the maximum principle can be proven for a multidimensional upwind scheme in the free-surface framework, provided that an analogous condition holds.

The discretization (4) allows the semi-implicit coupling to the momentum equation to yield an algebraic problem that can be very efficiently solved. The general form of the discretizations of equation (3) consistent with the discrete continuity equation (4) is given by

$$\begin{aligned} S_{i,j}^{n+1} H_{i,j}^{n+1} &= S_{i,j}^n H_{i,j}^n \\ &\quad - \frac{\Delta t}{\Delta x} \left(U_{i+\frac{1}{2},j}^{n+\frac{1}{2}} S_e H_{i+\frac{1}{2},j}^n - U_{i-\frac{1}{2},j}^{n+\frac{1}{2}} S_w H_{i-\frac{1}{2},j}^n \right) \\ &\quad - \frac{\Delta t}{\Delta y} \left(V_{i,j+\frac{1}{2}}^{n+\frac{1}{2}} S_n H_{i,j+\frac{1}{2}}^n - V_{i,j-\frac{1}{2}}^{n+\frac{1}{2}} S_s H_{i,j-\frac{1}{2}}^n \right). \end{aligned} \quad (5)$$

Here, S_w , S_e , S_n , S_s denote values of the concentration S at the sides of the computational cells. The specific reconstruction procedures for these values determine the discretization scheme chosen. However, it is immediate that formulating the discretization of the advection equation as in (5) is sufficient to satisfy the consistency with continuity condition as previously defined, as long as the reconstruction procedure for the face values correctly interpolates constants.

In order to show how the CWC condition can be violated, one may consider a generic finite volume discretization of equation (3):

$$\begin{aligned} S_{i,j}^{n+1} H_{i,j}^{n+1} &= S_{i,j}^n H_{i,j}^n \\ &\quad - \frac{\Delta t}{\Delta x} \left[U_e S_e H_e - U_w S_w H_w \right] - \frac{\Delta t}{\Delta y} \left[V_n S_n H_n - V_s S_s H_s \right]. \end{aligned} \quad (6)$$

Whenever the values of U, V, H used in the computation of the mass fluxes in (6) do not coincide with the values at the cell sides which appear in

(4), the discrete continuity equation will not in general be recovered for the case of a constant initial datum. It is to be remarked here that employing a different discretization of the continuity equation, so as to yield CWC for the general scheme (6), is not advisable, given the great computational advantages of the semi-implicit and mass conservative discretization (4). For definiteness, a special case of CWC violation will be considered in the following numerical comparisons, which results of the discretization

$$\begin{aligned} S_{i,j}^{n+1} H_{i,j}^{n+1} &= S_{i,j}^n H_{i,j}^n \\ &- \frac{\Delta t}{\Delta x} \left(U_{i+\frac{1}{2},j}^{n+\frac{1}{2}} (SH)_e - U_{i-\frac{1}{2},j}^{n+\frac{1}{2}} (SH)_w \right) \\ &- \frac{\Delta t}{\Delta y} \left(V_{i,j+\frac{1}{2}}^{n+\frac{1}{2}} (SH)_n - V_{i,j-\frac{1}{2}}^{n+\frac{1}{2}} (SH)_s \right). \end{aligned} \quad (7)$$

Here, $(SH)_w$, $(SH)_e$, $(SH)_n$, $(SH)_s$ denote again values of the conserved quantity SH interpolated onto the sides of the computational cells. It is to be remarked that this is, *a priori*, a perfectly reasonable discretization approach and that far worse ways to violate CWC could be devised. For example, the velocity values u^{n+1}, v^{n+1} could be used instead of the time averaged values employed in the equations.

In the numerical tests described in the next section, the results obtained by (5) and (7) will be compared using exactly the same interpolation procedure in both discretization schemes, so as to single out the effects of CWC violation. However, a simple example of the problems resulting from the use of (7) can be given on a one-dimensional domain that consists of only two grid cells with equal scalar concentration ($S_1^0 = S_2^0$) and arbitrary depths. Flux of volume and scalar mass is only allowed between the two grid cells. Using first-order upwind to interpolate the scalar concentration to the flux faces and assuming positive velocity, the discretization (7) applied to this simple case yields for cell 1

$$(SH)_1^{n+1} = (SH)_1^n - \frac{\Delta t}{\Delta x} U_{\frac{3}{2}}^{n+\frac{1}{2}} (SH)_1^n. \quad (8)$$

Substitution of (4) for cell 1 results in

$$S_1^{n+1} \left(H_1^n - \frac{\Delta t}{\Delta x} U_{\frac{3}{2}}^{n+\frac{1}{2}} H_{\frac{3}{2}}^n \right) = S_1^n \left(H_1^n - \frac{\Delta t}{\Delta x} U_{\frac{3}{2}}^{n+\frac{1}{2}} H_1^n \right). \quad (9)$$

Thus, $S_1^{n+1} \neq S_1^n$ when $H_1^n \neq H_{\frac{3}{2}}^n$, so that, even in this very simple application, constant initial data are not preserved in general.

3. Numerical results

Several numerical simulations have been carried out in order to estimate the effects of violating the CWC condition. In all the simulations, in order to avoid sources of numerical error other than the chosen advection scheme, the initial datum for H, U was chosen as the appropriate value at the initial time for the analytical solution of the nonlinear shallow water equations considered in each case. The velocity field was then given at each time step by the analytical solution. The total fluid depth was computed by the discrete continuity equation (4) and the scalar concentration field was finally updated with a conservative scheme using the previously computed values of H, U . In all the tests, a comparison has been carried out between the results obtained using schemes (5) and (7), respectively. Exactly the same procedures for reconstruction of the face values of S, SH were used in both discretization schemes, so as to single out the effects of CWC violation.

One-dimensional tests were performed with the analytical solution of the Riemann problem in the case of constant bottom depth, see e.g. (E.F.Toro, 1992). The initial value for H was taken to be a positive constant for $x \leq 0$ and 0 for positive values of x . The initial datum for S was assumed to be constant and equal to 1. In these simulations, cells with total depth less than 10^{-4} m. were considered as dry cells. The results obtained with a Lax-Wendroff slope limited TVD second-order scheme for reconstruction of the concentration values at the sides of the cell are displayed in table 1, where relative errors in l_1, l_2 norm and maximum and minimum values of the computed solution are shown after 50 time steps. Furthermore, a one-dimensional test with an initial datum for S given by a cosine pulse was also run. No new maxima are produced by the CWC slope limited second-order scheme, while the corresponding non CWC scheme yields new maxima, as shown at various times in table 2. It is to be remarked that, in this test with flat channel bottom, the simple upwind scheme would yield correct results in both the CWC and non CWC form. However, if CWC were violated by incorrect specification of the velocity field, larger errors would be produced. This is shown in table 3, where the results of the same test as above are shown for the case of simple upwind discretization in which CWC was violated by the choice of the velocity field at time $n + 1$. As expected, the largest errors are observed to occur at the edge of the wet area in all the above tests.

For two-dimensional tests, the analytical solutions given in (W.C.Thacker, 1981) for oscillations in a parabolic basin were employed. Again, the initial datum for S was chosen to be constant and equal to 1. The Lax-Wendroff slope limited TVD second-order scheme reconstruction procedure was employed for the concentration values at the sides of the cell. In these sim-

TABLE 1. Results of one-dimensional simulation with constant initial datum, second-order TVD slope limited scheme.

	l_1	l_2	maximum	minimum
CWC	1.04e-8	5.84e-8	1.00	1.00
non CWC	6.25e-3	4.97e-2	1.07	3.16e-2

TABLE 2. Maximum values in one-dimensional simulation with non constant initial datum, second-order TVD slope limited scheme.

	10 time steps	20 time steps	30 time steps	40 time steps
CWC	0.99	0.99	0.99	0.99
non CWC	1.07	1.06	1.06	1.05

ulations, cells with total depth less than 10^{-4} m. were considered as dry cells. The results obtained are displayed in table 4, where relative errors in l_1 , l_2 norm and maximum and minimum values of the computed solution are shown after 200 time steps, which correspond to half the period of the free-surface oscillations. Again, the largest errors occur in the areas where wetting and drying occur. Furthermore, a two-dimensional test with an initial datum for S given by a cosine hill was also run. No new maxima are produced by the CWC second-order scheme, while the corresponding non CWC scheme yields large new maxima, as shown at various simulation times in table 5.

4. Conclusion

In the framework of free-surface flows, the link between consistency of the advection scheme with the discretized continuity equation and constancy preservation has been emphasized. The general form of a CWC scheme for a specific discretization of the continuity equation has been presented. Several

TABLE 3. Results of one-dimensional simulation with constant initial datum, upwind scheme.

	l_1	l_2	maximum	minimum
CWC	8.77e-9	3.73e-8	1.00	1.00
non CWC	6.37e-3	0.23	2.25	0.00

TABLE 4. Results of two-dimensional simulation with constant initial datum, second-order scheme.

	l_1	l_2	maximum	minimum
CWC	5.44e-7	7.21e-7	1.00	1.00
non CWC	8.55e-3	4.13e-2	2.68	-5.48e-2

TABLE 5. Maximum values in two-dimensional simulation with non constant initial datum, second-order scheme.

	30 time steps	60 time steps	90 time steps	120 time steps
CWC	1.00	1.00	1.00	1.00
non CWC	2.35	1.94	1.54	1.36

numerical tests display the errors which can be induced by the violation of these conditions in the case of a conservation law coupled to the shallow water equations.

Acknowledgements

The authors are indebted to Vincenzo Casulli and Guus Stelling for many suggestions and useful discussions. Thanks are also due to Luigi Fraccarollo and Giorgio Rosatti for useful comments and advice.

References

- S.Lin and R.B.Rood (1996). Multidimensional Flux-Form Semi-Lagrangian Transport Schemes, *Monthly Weather Review*, **119**, pp. 2046-2070.
- V.Casulli (1990). Semi-Implicit Finite Difference Methods for the Two-Dimensional Shallow Water Equations, *Journal of Computational Physics*, **86**, n.1, pp. 56-74.
- V.Casulli and E.Cattani (1994). Stability, Accuracy and Efficiency of a Semi-Implicit Method for Three-Dimensional Shallow Water Flow, *Computers and Mathematics with Applications*, **27**, pp. 99-112.
- J.B.Goodman and R.LeVeque (1985). On the Accuracy of Stable Schemes for 2D Scalar Conservation Laws, *Mathematics of Computation*, **45**, pp. 15-21.
- E.S.Gross, J.R.Koseff and S.G.Monismith (1999). Evaluation of Advection Schemes for Estuarine Salinity Simulations, *ASCE Journal of Hydraulic Engineering*, **125**, pp. 32-46.
- E.S.Gross, V.Casulli, L.Bonaventura and J.R.Koseff (1998). A Semi-Implicit Method for Vertical Transport in Multidimensional Models, *International Journal for NUMERICAL METHODS IN FLUIDS*, **28**, pp. 157-186.
- R.LeVeque (1990). *Numerical Methods for Conservation Laws*, Birkhäuser, Basel .
- R. LeVeque (1996). High Resolution Conservative Algorithms for Advection in Incompressible Flow, *SIAM Journal of Numerical Analysis*, **33**, pp. 627-665.

- G.S.Stelling, H.W.J.Kernkamp and M.M.Laguzzi (1998). Delft Flooding System: A powerful tool for inundation assesment based upon a positive flow simulation, in: Hydroinformatics 98, Babovic and Larsen (eds.), Balkema.
- W.C.Thacker (1981). Exact Solutions to the Nonlinear Shallow-Water Wave Equations, *Journal of Fluid Mechanics*, **107** , pp. 499-508.
- E.F.Toro (1997). *Riemann Solvers and Numerical Methods for Fluid Dynamics*, Springer.
- E.F.Toro (1992). Riemann Problems and the WAF Method for the Two-Dimensional Shallow Water Equations, *Philosophical Transactions of the Royal Society of London*, **A341**, pp. 499-530.
- X.D.Liu (1993). A Maximum Principle Satisfying Modification of Triangle Based Adaptive Stencils for the Solution of Scalar Hyperbolic Conservation Laws, *SIAM Journal of Numerical Analysis*, **30**, N.3, pp.701-716.

HYPERBOLIC-ELLIPTIC SPLITTING FOR THE PSEUDO-COMPRESSIBLE EULER EQUATIONS

A. BONFIGLIOLI

Department of Environmental Engineering and Physics,

University of Basilicata,

Contrada Macchia Romana, 85100 Potenza, Italy

Email: bonfiglioli@unibas.it[†]

Abstract

Following previous work on the *canonical* decomposition of the subsonic, compressible Euler equations into their steady hyperbolic and elliptic components, a similar decomposition for the incompressible equations is proposed. The artificial compressibility approach is used make the incompressible Euler equations hyperbolic in time. The canonical form of this pseudo-compressible system consists in an hyperbolic component corresponding to the convection of total pressure along the streamlines and a Cauchy-Riemann system corresponding to the omni-directional propagation of the (artificial) acoustic waves.

The discretization of the pseudo-unsteady system is accomplished using Fluctuation Splitting schemes and unstructured meshes.

1. Introduction

Over the last few years a number of authors (Ta'asan, 1993; Ta'asan, 1994; Mesaros and Roe, 1995; Paillère and Deconinck, 1995; Sidilkover, 1999) tried to exploit the mixed hyperbolic-elliptic nature of the steady incompressible and subsonic compressible Euler equations to develop more accurate discretization schemes. The general approach consists in bringing the Euler system into a form which isolates the hyperbolic and elliptic components. This canonical form (Ta'asan, 1993), in which the two components are mutually decoupled at the discrete linearized level, suggests the use of

[†]present address: Dept. of Engineering, Queen Mary and Westfield College, London E1 4NS, UK, Email: a.bonfiglioli@qmw.ac.uk

upwind schemes for the discretization of the hyperbolic part and directionally unbiased schemes for the elliptic part (Ta'asan, 1994).

In this paper we extend to the incompressible Euler equations the algorithm used in (Mesaros and Roe, 1995; Paillère and Deconinck, 1995) for the compressible equations. In doing so, the steady equations are solved by time marching a preconditioned pseudo-unsteady system. The preconditioning matrix is chosen such that it brings the equations into their canonical form. Using the analysis outlined in (Turkel and Roe, 1996), a four-parameters family of preconditioning matrices that preserves the canonical decomposition is derived. The simplest choice of these free parameters leads to a preconditioning matrix which is identical to the one proposed in (Turkel, 1993). Numerical experiments show that improved accuracy is obtained when solving the preconditioned equations.

2. Hyperbolic-elliptic splitting

We consider the two-dimensional, steady, incompressible, inviscid equations in primitive variables ($\mathbf{U} = (p, \mathbf{u})^t$). Written in a Cartesian reference frame, these read:

$$\begin{aligned} u_x + v_y &= 0 \\ uu_x + vu_y + p_x &= 0 \\ uv_x + vv_y + p_y &= 0 \end{aligned} \tag{1}$$

If we consider an orthogonal reference frame (ξ, η) with the unit vector \mathbf{e}_ξ aligned with the local streamline, the system (1) simplifies to:

$$\underbrace{\begin{pmatrix} 0 & 1 & 0 \\ 1 & q & 0 \\ 0 & 0 & q \end{pmatrix}}_{A_\xi} \begin{pmatrix} p \\ q \\ r \end{pmatrix}_\xi + \underbrace{\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}}_{A_\eta} \begin{pmatrix} p \\ q \\ r \end{pmatrix}_\eta = 0. \tag{2}$$

In this stream-aligned reference frame, the differential of the velocity vector is denoted by $\partial \mathbf{u} = \partial q \mathbf{e}_\xi + \partial r \mathbf{e}_\eta$.

We now look at the characteristics of the system (2). The matrix:

$$A_\xi^{-1} A_\eta = \begin{pmatrix} 0 & 0 & -q \\ 0 & 0 & 1 \\ 1/q & 0 & 0 \end{pmatrix}$$

has eigenvalues $(\pm i, 0)$ and can be factored, using its eigenvector decomposition, as:

$$A_{\xi}^{-1} A_{\eta} = \underbrace{\begin{pmatrix} i/2 & -i/2 & 0 \\ -i/(2q) & i/(2q) & 1/q \\ 1/(2q) & 1/(2q) & 0 \end{pmatrix}}_R \underbrace{\begin{pmatrix} i & 0 & 0 \\ 0 & -i & 0 \\ 0 & 0 & 0 \end{pmatrix}}_{\Lambda} \underbrace{\begin{pmatrix} -i & 0 & q \\ i & 0 & q \\ 1 & q & 0 \end{pmatrix}}_L.$$

Following (Turkel and Roe, 1996), we rewrite the factorization as:

$$A_{\xi}^{-1} A_{\eta} = (R J^t) \left((J^t)^{-1} \Lambda (J)^{-1} \right) (J L) = R^{\#} \Lambda^{\#} L^{\#},$$

where:

$$R^{\#} = \begin{pmatrix} 0 & -\sqrt{2}/2 & 0 \\ 0 & \sqrt{2}/(2q) & 1/q \\ \sqrt{2}/(2q) & 0 & 0 \end{pmatrix}, \quad \Lambda^{\#} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

and the matrix J is given in (Turkel and Roe, 1996). Introducing the change of variables:

$$\partial \mathbf{V} = L^{\#} \partial \mathbf{U} = \begin{pmatrix} \sqrt{2}q \partial r \\ \sqrt{2} \partial p \\ \partial p + q \partial q \end{pmatrix} \quad L^{\#} = \begin{pmatrix} 0 & 0 & \sqrt{2}q \\ \sqrt{2} & 0 & 0 \\ 1 & q & 0 \end{pmatrix} \quad (3)$$

equation (2) can be rewritten as:

$$(R^{\#})^{-1} (L^{\#})^{-1} \mathbf{V}_{\xi} + \Lambda^{\#} \mathbf{V}_{\eta} = 0,$$

or, using matrix notation:

$$\left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{array} \right) \left(\begin{array}{c} v_1 \\ v_2 \\ v_3 \end{array} \right)_{\xi} + \left(\begin{array}{cc|c} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{array} \right) \left(\begin{array}{c} v_1 \\ v_2 \\ v_3 \end{array} \right)_{\eta} = 0. \quad (4)$$

Equation (4) gives the desired block diagonal form: the third equation expresses convection of total pressure along the streamline (Bernoulli's theorem), while the 2×2 block represents the Cauchy-Riemann equations in the (v_1, v_2) variables.

Since we are interested in solving the steady equations using a pseudo-transient approach, we shall consider an unsteady problem of the form:

$$P^{-1} \mathbf{U}_t + A_{\xi} \mathbf{U}_{\xi} + A_{\eta} \mathbf{U}_{\eta} = 0 \quad (5)$$

where P is a preconditioning matrix, still to be determined.

A general approach for deriving a family of preconditioning matrices that preserve the canonical decomposition is proposed in (Turkel and Roe, 1996) where it is applied to the compressible equations. The same analysis is applied here to the incompressible equations.

As in the steady case, performing the change of variables (3) in (5) gives:

$$\underbrace{(R^\#)^{-1} A_\xi^{-1} P^{-1} (L^\#)^{-1}}_{D^\#} \mathbf{V}_t + (R^\#)^{-1} (L^\#)^{-1} \mathbf{V}_\xi + \Lambda^\# \mathbf{V}_\eta = 0. \quad (6)$$

To preserve the canonical decomposition, it is required that $D^\#$ is a block diagonal matrix having the same structure as the matrices that multiply the space derivatives in (6). We take:

$$D^\# = \begin{pmatrix} P & Q & 0 \\ R & S & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

For well posedness the determinant $\Delta = PS - QR$ of $D^\#$ must be positive. Solving for $P^{-1} = A(R^\#)^{-1} D^\# (L^\#)$ and its inverse, gives:

$$P^{-1} = \begin{pmatrix} (S+1)/q & 1 & R \\ 1 & q & 0 \\ Q & 0 & qP \end{pmatrix}, \quad P = \frac{1}{\Delta} \begin{pmatrix} qP & -P & -R \\ -P & (\Delta+P)/q & R/q \\ -Q & Q/q & S/q \end{pmatrix}.$$

The simplest choice obviously consists in choosing $D^\#$ equal to the identity matrix. This gives a preconditioning matrix:

$$P = \frac{1}{q} \begin{pmatrix} q^2 & -q & 0 \\ -q & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \frac{1}{q} \begin{pmatrix} q^2 & -\mathbf{u}^t & 0 \\ -\mathbf{u} & \mathbf{I} + \mathbf{uu}/q^2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (7)$$

which is identical to the one proposed in (Turkel, 1993), eqn. 19.

3. The numerical algorithm

A numerical algorithm for discretizing the incompressible Euler equations using the artificial compressibility approach (Chorin, 1967) and Fluctuation Splitting schemes (van der Weide et al., 1999) has been presented in (Bonfiglioli, 1998); see also (Michelsen, 1996).

The conservation equations for mass and momentum, written in quasi-linear form, read:

$$\mathbf{U}_t + \mathbf{F}_\mathbf{U} \cdot \nabla \mathbf{U} = 0, \quad \mathbf{F} = (a^2 \mathbf{u}, \mathbf{u}\mathbf{u} + p\mathbf{I}). \quad (8)$$

In Eq. 8 \mathbf{F}_U represents the jacobian matrix of the flux function \mathbf{F} and the constant a denotes the artificial sound speed (Chorin, 1967).

While in (Bonfiglioli, 1998) the discrete schemes were applied directly to the conservation form of the governing equations (Eq. 8), which is equivalent to use the classical artificial compressibility approach, in the following we shall alter the time dependent behaviour of the equations using the preconditioning matrix P given by Eq. 7. The pseudo-unsteady system to be discretized takes the form:

$$\mathbf{U}_t + \left(P T^{-1} \right) \mathbf{F}_U \cdot \nabla \mathbf{U} = 0, \quad (9)$$

were the preconditioning matrix P is given in Eq. 7 and the matrix:

$$T = \begin{pmatrix} a^2 & 0 \\ \mathbf{u} & \mathbf{I} \end{pmatrix}$$

has to be introduced to account for the fact that the analysis presented in Sect. 2 was based on the non-conservative form of the Euler system, (Eq. 1), while the discretization is applied to the conservation form (Eq. 8).

We now introduce a change of variables, similar to the one given by Eq. 3 and such that the pseudo-unsteady system (9) takes the canonical, block diagonal, form:

$$\partial_t \mathbf{V} + \left(\begin{array}{c|ccc} \mathbf{e}_\xi & 0 & 0 & 0 \\ \hline 0 & \mathbf{e}_\xi & \mathbf{e}_\eta & 0 \\ 0 & \mathbf{e}_\eta & -\mathbf{e}_\xi & \mathbf{e}_\zeta \\ 0 & 0 & \mathbf{e}_\zeta & \mathbf{e}_\xi \end{array} \right) \cdot \nabla \mathbf{V} = 0, \quad \partial \mathbf{V} = \begin{pmatrix} \partial p + q \mathbf{e}_\xi \cdot \partial \mathbf{u} \\ q \mathbf{e}_\eta \cdot \partial \mathbf{u} \\ \partial p \\ q \mathbf{e}_\zeta \cdot \partial \mathbf{u} \end{pmatrix}.$$

As for the two dimensional case, the transport equation for total pressure decouples. In three dimensions, however, the 3×3 diagonal block has a mixed hyperbolic-elliptic character at steady state showing that the decomposition is sub-optimal in three dimensions, though it correctly recovers Eq. 4 in two space dimensions.

4. Numerical results

The low speed flow past the NACA-0012 airfoil ($\alpha_\infty = 2^\circ$) and the ONERA-M6 wing ($\alpha_\infty = 0^\circ$) have been selected to demonstrate the improved accuracy that is obtained when solving the preconditioned formulation of the incompressible Euler equations. For both calculations, the total pressure transport equation has been discretized using a non-linear, positivity preserving, upwind scalar scheme (PSI scheme) while a matrix distribution scheme of the Lax-Wendroff type has been used for the remaining equations. Fig. 1 shows the velocity distribution along both surfaces of the

NACA-0012 profile in the leading edge region. It is evident that the solution of the preconditioned incompressible and compressible ($M_\infty = 0.01$) equations yields results in better agreement with the reference solution obtained from a panel method. Fig. 2 shows the velocity isolines at a span-wise section of the ONERA-M6 wing: very close agreement can be observed between the solutions of the preconditioned incompressible and compressible ($M_\infty = 0.01$) equations.

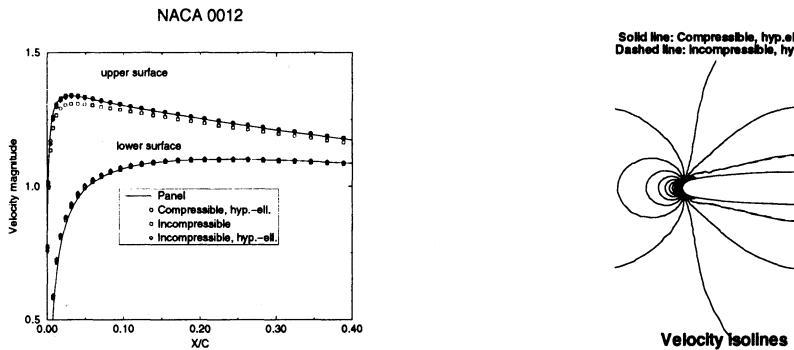


Figure 1. NACA-0012 airfoil.

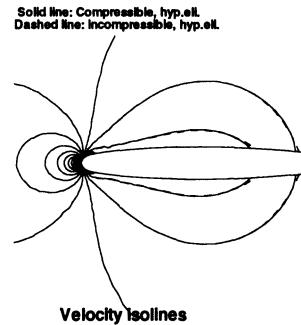


Figure 2. ONERA-M6 wing.

References

- Bonfiglioli A (1998). Multidimensional Residual Distribution Schemes for the Pseudo-Compressible Euler and Navier-Stokes Equations on Unstructured Meshes. *Lecture Notes in Physics*, **515**, pp 254-259.
- Chorin A (1967). A Numerical Method for Solving Viscous Flow Problems. *Journal of Computational Physics*, **2**, pp 12-26.
- Mesaros L and Roe P L (1995). Multidimensional Fluctuation Splitting Schemes Based on Decomposition Methods. 12th AIAA CFD Conference, San Diego, Paper 95-1699.
- Michelsen J A (1996). Multidimensional Upwind Schemes for the Pseudo-Compressible Euler Equations. *Notes on Numerical Fluid Mechanics*, **57**.
- Paillère H and Deconinck H (1995). Conservative upwind residual-distribution schemes based on the steady characteristics of the Euler equations. 12th AIAA CFD Conference, San Diego, Paper 95-1700.
- Sidilkover D (1999). Factorizable schemes for the Equations of Fluid Flow. ICASE Report 99-22.
- Ta'asan S (1993). Canonical Forms of Multidimensional Inviscid Flows. ICASE Report 93-94.
- Ta'asan S (1994). Canonical-Variables Multigrid Method for Steady-State Euler Equations. ICASE Report 94-14.
- Turkel E (1993). Review of Preconditioning Methods for Fluid Dynamics. *Applied Numerical Mathematics*, **12**, pp 257-284.
- Turkel E and Roe P L (1996). Preconditioning Methods for Multidimensional Aerodynamics. VKI LS 1996-06.
- van der Weide E and Deconinck H and Issman E and Degrez G (1999) A parallel, implicit, multi-dimensional upwind, residual distribution method for the Navier-Stokes equations on unstructured grids. *Computational Mechanics*, **23**, pp 199-208.

GODUNOV SOLUTION OF SHALLOW WATER EQUATIONS ON CURVILINEAR AND QUADTREE GRIDS

A.G.L. BORTHWICK

*Reader, Dept of Engineering Science,
Oxford University, Oxford OX1 3PJ, U.K.
Email: alistair.borthwick@eng.ox.ac.uk*

M. FUJIHARA

*Professor, Ehime University, 3-5-7 Tarumi,
Matsuyama, 790-8566, Japan.
Email: fujihara@agr.ehime-u.ac.jp*

AND

B.D. ROGERS

*Research Student, Dept of Engng Science, Oxford University
Email: ben.rogers@eng.ox.ac.uk*

Abstract. This paper describes a second-order accurate Godunov-type scheme for solving the 2D shallow water equations on non-orthogonal curvilinear boundary-fitted and Cartesian quadtree grids. The shallow water equations are written in a new matrix-hyperbolic form suitable for spatially non-uniform bed profiles. The equations are discretised spatially using finite volumes and temporally using the fourth-order Runge-Kutta method. Convection terms are modelled using Roe's flux function with a nonlinear minmod limiter. Validation tests include wind-induced circulation in a dish-shaped basin, an inviscid dam break simulation, and jet-forced flow in a flat-bottomed circular reservoir.

1. Introduction

In 1959, Godunov proposed a first-order accurate scheme for discretising the hyperbolic Euler equations, which permitted flow discontinuities to be modelled properly, and became widely utilised in aerodynamics shock-capturing (e.g. Roe 1981). Godunov-type schemes have also become popular for solv-

ing the matrix-hyperbolic shallow flow equations (e.g. Alcrudo and GarciaNavarro 1993, Ambrosi 1995, Zhao *et al.* 1996, Anastasiou and Chan 1997, and Mingham and Causon 1998). In producing the hyperbolic formulation, the surface gradient terms have been split in a way which does not conserve continuity or momentum when applying Roe's approximate Riemann solver to cases involving non-uniform bathymetry. Although Ambrosi and Zhao *et al.* noticed that problems arose, no corrective actions were suggested. This paper presents a new formulation that properly accounts for non-uniform bed profiles and preserves the accuracy of the scheme.

The boundaries of natural shallow-flow domains are usually geometrically complicated, and directly influence the interior flow behaviour. It is therefore important to be able to model correctly the boundary configuration. Moreover, refined solutions are needed in interior regions where high flow gradients occur. To achieve these goals, adaptive boundary-fitted and locally refined grid generation techniques have been the subject of much recent research. Typical approaches include boundary-fitted curvilinear systems (e.g. Borthwick and Akponasa 1997), unstructured advancing front and Voronoi methods (e.g. Anastasiou and Chan 1997), and hierarchical grid techniques (e.g. Greaves and Borthwick 1999). Herein, the Godunov-type shallow flow solver is based on boundary-fitted curvilinear and hierarchical quadtree grids, and applied to standard test cases.

2. Curvilinear and quadtree grid generation

Non-orthogonal curvilinear grids are generated by the numerical solution of Poisson-type equations, and fitted to the flow domain of interest. The mapping from the physical to the computational domain is given by $\xi_{xx} + \xi_{yy} = p(\xi, \eta)$ and $\eta_{xx} + \eta_{yy} = q(\xi, \eta)$ where p and q are weighting functions. Interchanging the computational co-ordinates (ξ, η) with the physical co-ordinates (x, y) gives $\alpha x_{\xi\xi} - 2\beta x_{\xi\eta} + \gamma x_{\eta\eta} + J^2(px_\xi + qx_\eta) = 0$ and $\alpha y_{\xi\xi} - 2\beta y_{\xi\eta} + \gamma y_{\eta\eta} + J^2(py_\xi + qy_\eta) = 0$, where $\alpha = x_\eta^2 + y_\eta^2$, $\beta = x_\xi x_\eta + y_\xi y_\eta$, $\gamma = x_\xi^2 + x_\eta^2$, $J = x_\xi y_\eta - x_\eta y_\xi$. These are solved in central difference form using successive over-relaxation.

Quadtree grids are created by means of recursive decomposition of a root square panel into which a digitised representation of the flow geometry is inserted after normalisation. Panel decomposition is undertaken as follows: (i) divide the root square into four quadrant panels; (ii) search each panel for seeding points, and subdivide if more than a prescribed number of seeding points; (iii) repeat (ii) until a given level of resolution is reached; and, (iv) implement grid regularisation to minimise hanging nodes. Greaves and Borthwick (1999) give details of the quadtree cell numbering and neigh-

bour finding system.

3. Discretised shallow water equations

In a generalised curvilinear co-ordinate system, the 2-D Cartesian shallow water equations may be expressed as

$$\frac{\partial \mathbf{q}}{\partial t} + \frac{\partial \mathbf{f}}{\partial \xi} + \frac{\partial \mathbf{g}}{\partial \eta} = \mathbf{h}, \quad (1)$$

where

$$\mathbf{q} = J \begin{pmatrix} \zeta \\ uh \\ vh \end{pmatrix}, \quad (2)$$

$$\mathbf{f} = y_\eta \begin{pmatrix} uh \\ u^2 h + g(\zeta^2 + 2\zeta h_s)/2 \\ uvh \end{pmatrix} - x_\eta \begin{pmatrix} vh \\ uvh \\ v^2 h + g(\zeta^2 + 2\zeta h_s)/2 \end{pmatrix}, \quad (3)$$

$$\mathbf{g} = -y_\xi \begin{pmatrix} uh \\ u^2 h + g(\zeta^2 + 2\zeta h_s)/2 \\ uvh \end{pmatrix} + x_\xi \begin{pmatrix} vh \\ uvh \\ v^2 h + g(\zeta^2 + 2\zeta h_s)/2 \end{pmatrix} \quad (4)$$

and

$$\mathbf{h} = J \begin{pmatrix} 0 \\ \frac{1}{J} \frac{\partial}{\partial \xi} \left[\frac{vh}{J} \left\{ (y_\eta^2 + x_\eta^2)u_\xi - (y_\xi y_\eta + x_\xi x_\eta)u_\eta \right\} \right] + \frac{\tau_{sx} - \tau_{bx}}{\rho} + hfv \\ \frac{1}{J} \frac{\partial}{\partial \eta} \left[\frac{vh}{J} \left\{ -(y_\xi y_\eta + x_\xi x_\eta)u_\xi + (y_\xi^2 + x_\xi^2)u_\eta \right\} \right] - g\zeta S_{ox} \\ \frac{1}{J} \frac{\partial}{\partial \xi} \left[\frac{vh}{J} \left\{ (y_\eta^2 + x_\eta^2)v_\xi - (y_\xi y_\eta + x_\xi x_\eta)v_\eta \right\} \right] + \frac{\tau_{sy} - \tau_{by}}{\rho} - hfu \\ \frac{1}{J} \frac{\partial}{\partial \eta} \left[\frac{vh}{J} \left\{ -(y_\xi y_\eta + x_\xi x_\eta)v_\xi + (y_\xi^2 + x_\xi^2)v_\eta \right\} \right] - g\zeta S_{oy} \end{pmatrix} \quad (5)$$

in which t is time, ζ is free surface elevation above still water level, u and v are the depth-averaged velocity components in the x - and y -directions, h is the total water depth, h_s is the still water depth, g is gravitational acceleration, ρ is water density, (τ_{sx}, τ_{sy}) are the surface stresses, (τ_{bx}, τ_{by}) are bed stresses, (S_{ox}, S_{oy}) are bed slopes, f is the Coriolis parameter, and v is the kinematic eddy viscosity coefficient. Note that the surface gradient terms have been split in a novel way for application to non-uniformly varying bathymetries. The bottom slope terms (S_{ox}, S_{oy}) are related to the curvilinear grid by $S_{ox} = (y_\eta S_{o\xi} - y_\xi S_{o\eta})/J$ and $S_{oy} = (x_\xi S_{o\eta} - y_\eta S_{o\xi})/J$ in which $(S_{o\xi}, S_{o\eta})$ are the bottom slope terms in the ξ - and η - directions respectively.

After integration over a control volume, and discretization with Roe's approximate Riemann solver used to evaluate the convection terms \mathbf{f} and \mathbf{g} , equation (1) becomes

$$\frac{\partial}{\partial t}(\mathbf{q}_{i,j}) = \mathbf{f}_{i,j} - \mathbf{f}_{i+1,j} + \mathbf{g}_{i,j} - \mathbf{g}_{i,j+1} + \mathbf{h}_{i,j}. \quad (6)$$

where, $\mathbf{f}_{i,j} = \frac{1}{2} \left[\mathbf{f}(\mathbf{q}_{i,j}^+) + \mathbf{f}(\mathbf{q}_{i,j}^-) - |\mathbf{A}_\xi| (\mathbf{q}_{i,j}^+ - \mathbf{q}_{i,j}^-) \right]$, $\mathbf{g}_{i,j} = \frac{1}{2} \left[\mathbf{g}(\mathbf{q}_{i,j}^+) + \mathbf{g}(\mathbf{q}_{i,j}^-) - |\mathbf{A}_\eta| (\mathbf{q}_{i,j}^+ - \mathbf{q}_{i,j}^-) \right]$ and $\Delta\xi = \Delta\eta = 1$. In the above

$$|\mathbf{A}_\xi| = \mathbf{R}_\xi |\Lambda_\xi| \mathbf{L}_\xi, \quad |\mathbf{A}_\eta| = \mathbf{R}_\eta |\Lambda_\eta| \mathbf{L}_\eta. \quad (7)$$

The flux Jacobians \mathbf{A}_ξ and \mathbf{A}_η are

$$\mathbf{A}_\xi = \frac{\partial \mathbf{f}}{\partial \mathbf{q}} = \begin{bmatrix} 0 & y_\eta & -x_\eta \\ (c^2 - u^2)y_\eta + uvx_\eta & 2uy_\eta - vx_\eta & -ux_\eta \\ -uvy_\eta - (c^2 - v^2)x_\eta & vy_\eta & uy_\eta - 2vx_\eta \end{bmatrix}, \quad (8)$$

and

$$\mathbf{A}_\eta = \frac{\partial \mathbf{g}}{\partial \mathbf{q}} = \begin{bmatrix} 0 & -y_\xi & x_\xi \\ -(c^2 - u^2)y_\xi - uvx_\xi & -2uy_\xi + vx_\xi & ux_\xi \\ uvy_\xi + (c^2 - v^2)x_\xi & -vy_\xi & -uy_\xi + 2vx_\xi \end{bmatrix}. \quad (9)$$

The diagonal eigenvalue matrices $|\Lambda_\xi|$ and $|\Lambda_\eta|$ are

$$|\Lambda_\xi| = \begin{bmatrix} |y_\eta u - x_\eta v| & 0 & 0 \\ 0 & |y_\eta u - x_\eta v - cn_\eta| & 0 \\ 0 & 0 & |y_\eta u - x_\eta v + cn_\eta| \end{bmatrix}, \quad (10)$$

$$\text{and } |\Lambda_\eta| = \begin{bmatrix} |x_\xi v - y_\xi u| & 0 & 0 \\ 0 & |x_\xi v - y_\xi u - cn_\xi| & 0 \\ 0 & 0 & |x_\xi v - y_\xi u + cn_\xi| \end{bmatrix} \quad (11)$$

with $n_\eta = \sqrt{x_\eta^2 + y_\eta^2}$ and $n_\xi = \sqrt{x_\xi^2 + y_\xi^2}$. The right eigenvector matrices are

$$\mathbf{R}_\xi = \begin{bmatrix} 0 & 1 & 1 \\ -x_\eta & u - \frac{y_\eta c}{n_\eta} & u + \frac{y_\eta c}{n_\eta} \\ -y_\eta & v + \frac{x_\eta c}{n_\eta} & v - \frac{x_\eta c}{n_\eta} \end{bmatrix} \quad \text{and} \quad \mathbf{R}_\eta = \begin{bmatrix} 0 & 1 & 1 \\ x_\xi & u + \frac{y_\xi c}{n_\xi} & u - \frac{y_\xi c}{n_\xi} \\ y_\xi & v - \frac{x_\xi c}{n_\xi} & v + \frac{x_\xi c}{n_\xi} \end{bmatrix} \quad (12)$$

and the left eigenvector matrices are

$$\mathbf{L}_\xi = \begin{bmatrix} \frac{x_\eta u + y_\eta v}{n_\eta^2} & \frac{-x_\eta}{n_\eta^2} & \frac{-y_\eta}{n_\eta^2} \\ \frac{y_\eta u - x_\eta v}{2cn_\eta} + \frac{1}{2} & \frac{-y_\eta}{2cn_\eta} & \frac{x_\eta}{2cn_\eta} \\ \frac{x_\eta v - y_\eta u}{2cn_\eta} + \frac{1}{2} & \frac{y_\eta}{2cn_\eta} & \frac{-x_\eta}{2cn_\eta} \end{bmatrix} \text{ and } \mathbf{L}_\xi = \begin{bmatrix} \frac{-x_\xi u - y_\xi v}{n_\xi^2} & \frac{x_\xi}{n_\xi^2} & \frac{y_\xi}{n_\xi^2} \\ \frac{x_\xi v - y_\xi u}{2cn_\xi} + \frac{1}{2} & \frac{y_\xi}{2cn_\xi} & \frac{-x_\xi}{2cn_\xi} \\ \frac{y_\xi u - x_\xi v}{2cn_\xi} + \frac{1}{2} & \frac{-y_\xi}{2cn_\xi} & \frac{x_\xi}{2cn_\xi} \end{bmatrix} \quad (13)$$

where c denotes the wave celerity ($= \sqrt{gh}$). The velocity components (u, v) and the wave celerity (c) in Eqs. (8)-(13) are given by Roe's average state. Interpolation is necessary to evaluate the the Riemann states either side of the cell interfaces. Herein, a minmod slope limiter is used. At closed boundaries, slip or no-slip conditions are applied as appropriate. Riemann invariants are used for open boundary conditions, according to the local Froude number. A fourth-order Runge-Kutta scheme is used for time integration.

4. Results

4.1. SURFACE STRESS-INDUCED CIRCULATION IN A DISH-SHAPED BASIN

The basin has still water depths given by

$h_s = (0.5 + \sqrt{0.5(1 - r/R_o)}) / 1.3$ where r is radial distance from the centre of the basin and R_o is its radius ($R_o = 192$ m). Fig. 1a shows the 8-level quadtree grid, with regularised central region at minimum level 6. A uniform surface stress of 0.02 N/m^2 directed eastward is built up over 1000 s, with a time step of 0.05 s. Bed friction, Coriolis and eddy viscosity are all zero. Steady-state is achieved by 50,000 s. Fig. 1b depicts the depth-averaged velocity vector field. The flow is directed against the wind along

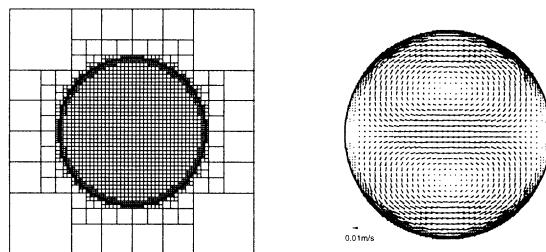


Figure 1. Dish-shaped basin:(a) quadtree grid and (b) velocity vectors.

the east-west axis of the dish, and there are two counter-rotating gyres with

their centres located along the north-south axis of the basin, in agreement with the analytical model of Kranenburg (1992).

4.2. CIRCULAR DAM BREAK SIMULATION

An infinitesimally thin-walled cylinder of radius 11 m contains water of depth 10 m, and is positioned at the centre of a 100 m square tank with a flat base. At $t = 0$, the cylindrical wall is instantaneously removed. Fig. 2a illustrates the initial level 10 quadtree grid, which was dynamically adapted according to criteria based on the magnitude of the free surface gradient. The water level contours at $t = 0.69$ s shown in Fig. 2b are in excellent agreement with numerical predictions by Mingham and Causon (1998). Figs. 3a and 3b present the curvilinear boundary-fitted grid, and corresponding water level contours at $t = 0.69$ s. Again, the results are very similar.

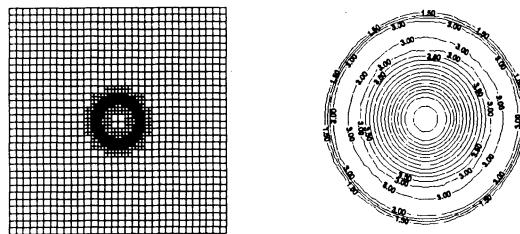


Figure 2. Circular Dam break:(a) initial quadtree grid and (b) depth contours at $t=0.69$ s.

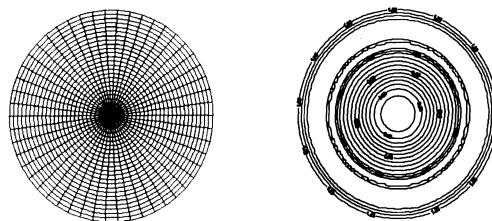


Figure 3. Circular dam break:(a) curvilinear grid and (b) depth contours at $t=0.69$ s.

4.3. JET-FORCED FLOW IN A CIRCULAR RESERVOIR

The models were also used to predict the steady-state velocity pattern for jet-forced circulation in a flat-bottomed circular basin with straight inlet and outlet stems. Fig. 4 shows the curvilinear grid and steady state depth-averaged stream function contours for a flow with inlet Reynolds number, $Re_I = U_I \varepsilon R_o / \nu = 10$. Here, the inlet velocity $U_I = 0.1$ m/s, the radius of the circular basin $R_o = 0.75$ m, the inlet and outlet stem openings subtend $2\varepsilon = \pi/15$ radians, and the kinematic viscosity $\nu = 7.84 \times 10^{-4}$ m²/s. Fig. 5 presents the quadtree grid and corresponding depth-averaged stream function contours.

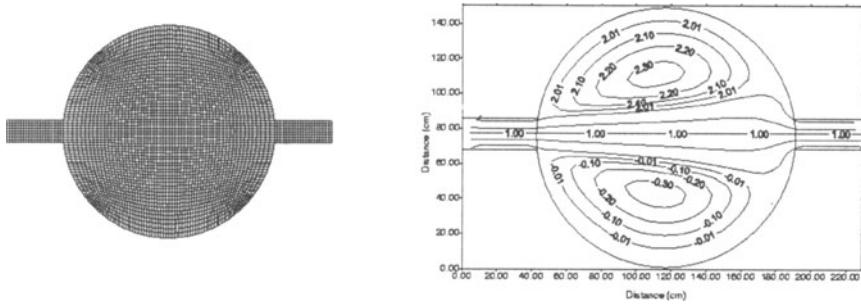


Figure 4. Jet-forced flow in a circular basin:(a) curvilinear grid and (b) streamlines for $Re_I = 10$.

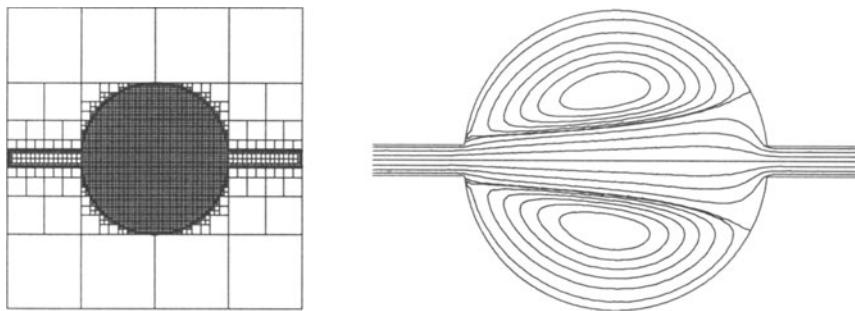


Figure 5. Jet-forced flow in a circular basin:(a) quadtree grid and (b) streamlines for $Re_I = 10$.

In both cases, the streamline pattern consists of an initially slightly diverging throughflow stream after the flow enters the circular basin, followed by a radial outflow at the opening to the outlet stem. Two counter-rotating

vortices are divided by the throughflow. The steady state flow field is almost identical to a semi analytical solution of the 2D Navier Stokes equations obtained by Dennis (1974).

5. Conclusions

This paper has described a Godunov-type solver of the shallow water equations, with the matrix-hyperbolic formulation accounting correctly for non-uniform spatial variations in bathymetry when applied with Roe's approximate Riemann scheme. The model gives accurate results on both curvilinear and quadtree grids.

Acknowledgements

This work has been supported by the Ministry of Education, Science, Sports and Culture of Japan and through U.K. EPSRC Grant GR/L92877, co-investigated by Dr K Anastasiou of Imperial College, London, and Dr P H Taylor of Oxford University.

References

- Alcrudo, F., and Garcia-Navarro, P. (1993). A high-resolution Godunov-type scheme in finite volumes for the 2D shallow-water equations, *Int. J. Num. Meth. Fluids*, **16**, pp 489-505.
- Ambrosi, D. (1995). Approximation of shallow water equations by Roe's Riemann solver, *Int. J. Num. Meth. Fluids*, **20**, pp 157-168.
- Anastasiou, K., and Chan, C. T. (1997). Solution of the 2D shallow water equations using the finite volume method on unstructured triangular meshes, *Int. J. Num. Meth. Fluids*, **24**, pp 1225-1245.
- Borthwick, A. G. L., and Akponasa, G. A. (1997). Reservoir flow prediction by contravariant shallow water equations, *J. of Hydr. Engng, ASCE*, **123**(5), pp 432-439.
- Dennis, S.C.R. (1974). Application of the series truncation method to two-dimensional flows, *Proc. 4th Int. Conf. on Num. Meth. in Fluid Dynamics, New York*, pp 146-151.
- Godunov, S. K. (1959). A difference method for the numerical computation of discontinuous solutions of hydrodynamic equations, *Math. Sbornik*, **3**, pp 271-306 (in Russian).
- Greaves, D.M., and Borthwick, A.G.L. (1999). Numerical Methods for Conservation Laws, *Int. J. Num. Meth. Fluids*, **45**, pp 447-471.
- Kranenburg, C. (1992). Wind-driven chaotic advection in a shallow model lake, *J. Hydraulic Res.*, **30**(1), pp 29-46.
- Mingham, C. G., and Causon, D.M. (1998). High-resolution finite-volume method for shallow water flows, *J. of Hydr. Engng, ASCE*, **124**(6), pp 605-614.
- Roe P. L. (1981). Approximate Riemann solvers, parameter vectors, and difference schemes, *J. Comput. Phys.*, **43**, pp 357-372.
- Zhao, D. H., Shen, H. W., Lai, J. S., and Tabios III, G. Q. (1996), *J. of Hydr. Engng, ASCE*, **122**(12), pp 692-702.

A HIGHER-ORDER-ACCURATE RECONSTRUCTION FOR THE COMPUTATION OF COMPRESSIBLE FLOWS ON CELL-VERTEX TRIANGULAR GRIDS

L. A. CATALANO

*Istituto di Macchine ed Energetica,
Politecnico di Bari,
Via Re David 200, 70125 Bari, ITALY
Email: catalano@poliba.it*

Abstract. A finite-volume method for the solution of two-dimensional compressible flows on cell-vertex unstructured grids is presented. A higher-order-accurate upwind discretization of the inviscid terms is obtained by using a new bi-linear reconstruction and a standard flux-difference-splitting scheme, whereas a standard central discretization is used for the viscous terms. The method is validated by computing the transonic inviscid flow in a two-dimensional cascade and the laminar flow past the NACA-0012 airfoil with incidence.

1. Introduction

In the last decade, Godunov schemes have reached a remarkable level of accuracy and robustness, which make them suitable for the numerical simulation of complex flows, see, *e.g.*, (LeVeque, 1992), (Roe, 1998) and (Toro, 1999). However, engineering applications often require the analysis of complex geometries, thus suggesting the use of unstructured meshes, where a non-trivial reconstruction scheme is needed to get higher-order spatial accuracy. Most upwind finite volume methods for unstructured grids proposed to date employ a cell-vertex discretization, since it allows a natural definition of the flow gradients: using a dual mesh, a gradient-based reconstruction is applied on the two sides of each interface, where an approximate Riemann solver is finally applied to select the proper upwind contributions. However, to knowledge of the author, all of the methods for cell-vertex unstructured grids developed so far reconstruct the flow variables according

to an average between the finite difference on the side itself and the flow variable gradients in the two nodes, which are computed from the gradients in the surrounding cells, see, *e.g.*, (Barth, 1990), (Barth, 1993), (Hallop, 1990). This approach is in contrast with the corresponding one-dimensional scheme, as it will be shown in the next Section. For such a reason, an alternative, much simpler, reconstruction scheme has been recently proposed by the author (Catalano, 1999), and validated by computing the subsonic and the transonic inviscid flows in a two-dimensional turbine cascade.

This paper first concludes the work proposed in Ref. (Catalano, 1999) by including a limiter in the reconstruction scheme; a comparison with the results obtained by using the same numerical discretization on structured grids is then proposed to demonstrate the good shock-capturing properties of the method. Finally, the extension to the discretization of the laminar flow equations is considered and tested.

2. Discretization of the compressible Navier-Stokes equations

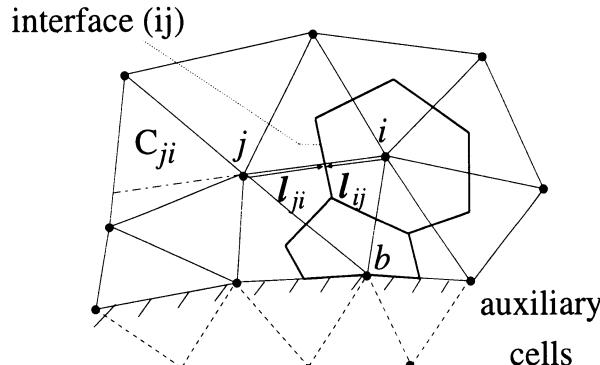


Figure 1. Construction of finite volumes for internal and boundary nodes. Determination of the cell C_{ji} . Construction of the auxiliary cells.

The domain is discretized by means of an unstructured triangular grid with unknowns located at each cell-vertex. The finite volume associated to each node is constructed by connecting the barycentres of two neighbouring cells, see Fig. 1. A higher-order-accurate upwind discretization of the inviscid terms of the Navier-Stokes equations is then obtained as follows: a left state and a right state are reconstructed linearly on the two sides of the interface (ij) associated to each side connecting the node i with each surrounding node j :

$$Q_{(ij)}^L = Q_j + (\nabla Q)_{ji} \cdot l_{ji}, \quad (1)$$

$$Q_{(ij)}^R = Q_i + (\nabla Q)_{ij} \cdot l_{ij}, \quad (2)$$

\mathbf{l}_{ji} and \mathbf{l}_{ij} being the two opposite vectors pointing from the two nodes to the intersection of the interface with the side, see Fig. 1. Different definitions of the gradients characterize the numerical methods cited in the Introduction, but all of them appear in contrast with the corresponding one-dimensional scheme. First consider a uniform one-dimensional grid: on the interface $(k + 1/2)$, $Q_{k+1/2}^L$ is linearly reconstructed as:

$$Q_{k+1/2}^L = \frac{3}{2}Q_k - \frac{1}{2}Q_{k-1} = Q_k + \frac{1}{2}(Q_k - Q_{k-1}). \quad (3)$$

It must be remarked that the reconstruction of the left state (similar arguments hold for the right state) is based on the *gradient* of Q in the left-neighbouring cell, rather than on the *gradient* defined in the node k . Similarly, in two dimensions, a unique left-neighbouring cell can be defined as the cell C_{ji} which contains the prolongation of the side (ji) , plotted as a dot-dashed line in Fig. 1. Clearly, the choice of a cell-vertex triangular grid allows to define a bi-linear variation of Q in each cell, thus defining the cell gradient $(\nabla Q)_{ji} \equiv (\nabla Q)_{C_{ji}}$ uniquely. Standard one-dimensional limiters can also be applied straightforwardly. The flux-difference-splitting of Roe (Roe, 1986) is then used to solve the Riemann problem defined at each interface. In Euler computations, a row of auxiliary cells is added beyond solid walls to retain the higher-order of accuracy of the reconstruction scheme. Isentropic simple radial equilibrium is imposed to compute the unknown values in the auxiliary nodes, see (Catalano, 1999) for details.

Finally, the viscous fluxes through the finite-volume contour are computed on a cell-to-cell basis by using a piecewise constant gradient. Clearly, this discretization coincides with the well-known Galerkin finite-element scheme.

3. Results

The method is firstly applied to the computation of the inviscid transonic flow in a 2D turbine cascade with a complex shock structure; the results are compared with those obtained by using the same numerical discretization (MUSCL extrapolation with the same Van Albada limiter, same Riemann solver) on a structured cell-centered grid, in order to recognize if the good shock-capturing capability of the Roe scheme is preserved with the proposed reconstruction. Since real blades have a rounded trailing edge, an artificial wedge with no load is usually added for Euler computations, to simulate the presence of the recirculation zone; however, the application of two different numerical schemes could lead to a different wedge shape and/or orientation; in order to avoid such strong uncertainty in the blade profile at the trailing edge, where a shock is created, a blade with a fixed sharp trailing edge is

constructed as follows: the blade camberline is defined as

$$y_c = 0.7x^3 - 2.3x^2 + x, \quad x \in [0, 1], \quad (4)$$

with pitch equal to 0.7. Then, a curvilinear coordinate s_{cl} is defined for each point of the camberline, $s_{cl} \in [0, 1]$. The thickness distribution $\delta_b(s_{cl})$ is then prescribed by shifting the thickness distribution of a NACA-0012 airfoil, $\delta_N(x)$ ($x \in [0, 1]$), as follows:

$$\delta_b(s_{cl}) = 2.2\delta_N(\epsilon s_{cl}), \quad s_{cl} \in [0, 1], \quad \epsilon = 1 - 1.2s_{cl} + 1.2s_{cl}^2. \quad (5)$$

Since the number of Riemann problems solved by the present approach is

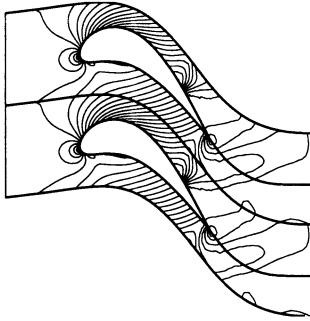


Figure 2. Turbine cascade: Mach number contours ($\Delta M = 0.05$) for the cell-centered structured grid with 129×17 nodes.

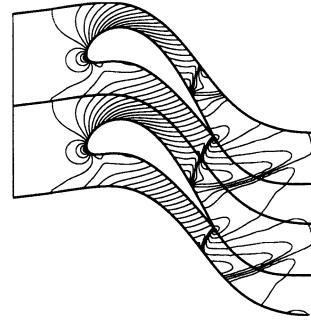


Figure 3. Turbine cascade: Mach number contours ($\Delta M = 0.05$) for the cell-centered structured grid with 257×33 nodes.

50% more than those occurring on a structured grid with the same number of nodes, two nested structured grids with 129×17 nodes and 257×33 nodes, respectively, are considered, whereas the present unstructured method is applied on the coarser mesh only, each quadrilateral cell being subdivided by the shorter diagonal.

The Mach number contours for outlet $M_{is} = 1$, computed on the two nested cell-centered structured grids are provided in Fig. 2 and in Fig. 3, respectively, whereas the results obtained by applying the present method on the coarser unstructured mesh are provided in Fig. 4. As visible, a shock departs from the trailing edge on the pressure side and reflects on the rear part of the suction side, where the flow still accelerates supersonically and seems to create a very weak normal shock, which is visible in Fig. 3 and in Fig. 4 only, as well as in the corresponding wall Mach number distributions

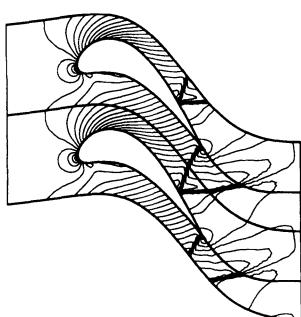


Figure 4. Turbine cascade: Mach number contours ($\Delta M = 0.05$) for the present method (cell-vertex unstructured grid with 129×17 nodes).

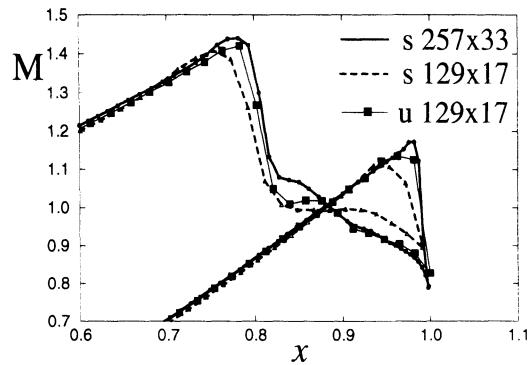
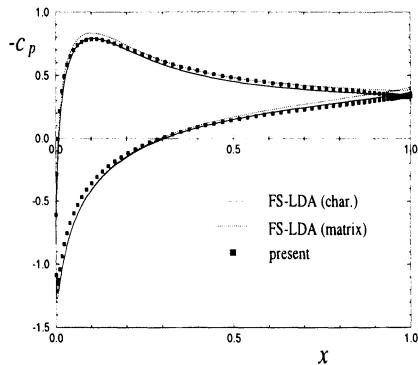
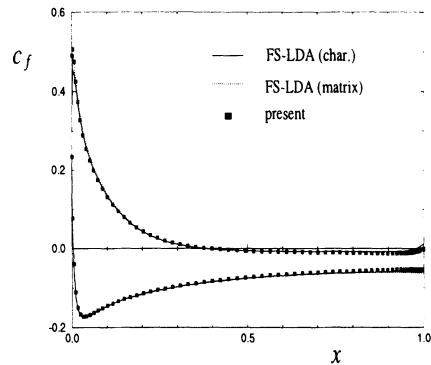


Figure 5. Wall Mach-number distributions near the trailing edge.

near the trailing edge, provided in Fig. 5. It is noteworthy that the shocks captured by the proposed approach on the coarser mesh are even sharper with respect to those computed on the finer structured grid: the shock reflection is captured in about three cells, and both the small expansion which occurs after a shock impinging on a curved wall and the following weak normal shock are visible.

The flow past the NACA 0012 airfoil with $M_\infty = 0.8$, 10° angle of attack and Reynolds number $Re = 500$ (wall temperature equal to the freestream total temperature) is then considered to test the extension of the method to viscous laminar flows. In particular, the present approach is applied on the same grid used in Ref. (Catalano, 1997), obtained from a C-mesh with 257×65 nodes; the results are then compared with those obtained in Ref. (Catalano, 1997) by means of the very accurate Fluctuation Splitting LDA scheme in its scalar (with characteristic decomposition of the flow equations) and matricial versions. The distributions of the pressure and the skin friction coefficients, C_p and C_f , provided in Fig. 6 and in Fig. 7, respectively, are in good agreement with those obtained in Ref. (Catalano, 1997). The recirculation region is visible in the velocity contours (not presented here for lack of space), C_f becoming negative at $x/c = 0.40$ ($x/c = 0.39$ in Ref. (Catalano, 1997)). The values of the lift and the drag coefficients obtained with the present method are $C_L = 0.441$, $C_D = 0.255$, whereas $C_L = 0.430-0.422$ and $C_D = 0.261-0.264$ are reported in Ref. (Catalano, 1997).

Figure 6. C_p distribution.Figure 7. C_f distribution.

4. Conclusions

A finite-volume method for the solution of two-dimensional compressible flows on cell-vertex unstructured grids has been presented. In particular, a new higher-order reconstruction of the unknowns has been proposed, which allows to capture shocks very sharply. The method has been then extended to viscous laminar flows and validated versus the well-known flow past the NACA-0012 airfoil with incidence.

References

- Barth T J (1990). Aspects of unstructured grids and finite-volume solvers for the Euler and Navier-Stokes equations. *Lecture Series 1991-06, Von Karman Institute*.
- Barth T J (1993). A 3-D least-squares upwind Euler solver for unstructured meshes. *Lecture Notes in Physics*, **414**, Springer Verlag, pp 90-94.
- Catalano L A (1999). A higher-order-accurate upwind method for 2D compressible flows on cell-vertex unstructured grids. *Second International Symposium on Finite Volumes for Complex Applications*, Duisburg (Germany).
- Catalano L A, De Palma P, Napolitano M and Pascazio G (1997). Genuinely multidimensional upwind methods for accurate and efficient solutions of compressible flows. *Euler and Navier-Stokes solvers using multi-dimensional upwind schemes and multi-grid acceleration*, Notes on Numerical Fluid Mechanics, **57**, pp 221-250.
- Hallo L, Le Ribault C and Buffat M. An implicit mixed finite-volume-finite-element method for solving 3D turbulent compressible flows. *International Journal for Numerical Methods in Fluids*, **25**, pp 1241-1261.
- LeVeque R J (1992). Numerical Methods for Conservation Laws. Birkhäuser Verlag.
- Roe P L (1998). The Harten Memorial Lecture—New Applications of Upwinding. Numerical Methods for Wave Propagation, pp 1-31. Toro E F and Clarke J F (Editors). Kluwer Academic Publishers.
- Roe P L (1986). Characteristic based schemes for the Euler equations. *Ann. Rev. Fluid Mech.*, **18**, pp 337-365.
- Toro E F (1999). Riemann Solvers and Numerical Methods for Fluid Dynamics. Second Edition, Springer-Verlag.

NUMERICAL EXPERIMENTS WITH MULTILEVEL SCHEMES FOR CONSERVATION LAWS

G. CHIAVASSA AND R. DONAT

*Departament de Matematica Aplicada,
Universidad de Valencia,
46100 Burjassot, Valencia, Spain
Emails: guillaume.chiavassa@uv.es
donat@uv.es*

Abstract. Main steps of a point-value multilevel algorithm are presented and numerical results for a two dimensional test case of gas dynamics are discussed in terms of quality and efficiency.

1. Introduction and general framework

We are concerned with solving the initial value problem for the two dimensional compressible Euler equations :

$$\begin{cases} \partial_t \vec{U} + \vec{f}(\vec{U})_x + \vec{g}(\vec{U})_y = \vec{0} \\ \vec{U}(x, y, 0) = \vec{U}_0(x, y) \end{cases} \quad (1)$$

with $\vec{U} = (\rho, m_x, m_y, E)^t$ and \vec{f} and \vec{g} the classical physical flux functions. The discretization procedure of Shu and Osher (Shu and Osher, 1989) is applied to (1) and the unknowns at each time step $t_n = n\delta t$ are the values of \vec{U} on a Cartesian grid $\mathcal{G}^0 : \vec{U}_{ij}^n = \vec{U}(x_i, y_j, t_n)$. In the conservative discretized form of (1) appears the numerical divergence:

$$\mathcal{D}(\vec{U})_{ij} = \frac{\vec{F}_{i+1/2,j} - \vec{F}_{i-1/2,j}}{\delta x} + \frac{\vec{G}_{i,j+1/2} - \vec{G}_{i,j-1/2}}{\delta y} \quad (2)$$

where $\vec{F}_{i+1/2,j}$ and $\vec{G}_{i,j+1/2}$ are the numerical fluxes in the x and y directions defined from the functions $\vec{F}(\vec{U}_{i-k,j}, \dots, \vec{U}_{i+m,j})$ and $\vec{G}(\vec{U}_{i,j-k}, \dots, \vec{U}_{i,j+m})$. High order shock-capturing schemes are needed to evaluate these numerical fluxes in order to represent accurately the solution without unphysical

oscillations. In our experiments, we will use Marquina's scheme (Donat and Marquina, 1996) with a piecewise hyperbolic reconstruction procedure of order 3 (PHM). Such schemes are very efficient, but suffer from a large computational cost due in particular to their strongly nonlinear structure. In order to reduce this cost, Harten proposed (Harten, 1995) to use a multiresolution representation of the flow to analyze its local smoothness and to replace the expensive flux computation in the smooth zones by a cheaper procedure. These so called *multilevel schemes* have been successfully tested for one dimensional Euler equations in (Harten, 1995) and for two dimensional scalar equations in (Bihari and Harten, 1997). In (Sjögreen, 1995), Sjögreen implements a "dimension by dimension" multilevel algorithm for a bi-dimensional compressible Euler computation. This algorithm uses only one-dimensional procedures which leads to an easier implementation, but tends to be less efficient, in our experimentations, than a fully two-dimensional one (Chiavassa and Donat, 1999).

In this paper, we present a two dimensional algorithm developed in the point values context whereas the previous ones have always been derived in the cell-average framework. A complete description of the multilevel strategy can be found in (Chiavassa and Donat, 1999), and only the main steps are presented in section 2. In section 3, this algorithm is applied to a classical CFD problem, and the results are discussed in terms of quality and efficiency.

2. Description of the algorithm

The three principal ingredients of the multilevel algorithm are the multiresolution transform, the thresholding procedure and the multilevel computation of the numerical divergence.

2.1. MULTIREOLUTION TRANSFORM OF DATA

This transform has been initially presented by Harten in (Harten, 1996) and the reader is referred to this paper for details. One first defines a set of embedded grids $\{\mathcal{G}^l, l = 1, \dots, L\}$ by

$$(x_i, y_j) \in \mathcal{G}^l \iff (x_{2^l i}, y_{2^l j}) \in \mathcal{G}^0. \quad (3)$$

The approximation of a function v on \mathcal{G}^l is therefore obtained from its approximation on \mathcal{G}^0 by :

$$v_{ij}^l = v_{2^l i, 2^l j}^0 \quad i = 0, \dots, Nx/2^l \quad j = 0, \dots, Ny/2^l \quad (4)$$

if $N_p = (Nx + 1) \times (Ny + 1)$ is the size of \mathcal{G}^0 .

Now, the *wavelet* or *detail* coefficients at level l , d_{ij}^l , are computed as the

difference between the approximation of v on \mathcal{G}^l and on the finer grid \mathcal{G}^{l-1} . Essentially, these coefficients are evaluated comparing the exact values of v on \mathcal{G}^{l-1} with some approximated values \tilde{v}_{ij}^{l-1} obtained by a polynomial interpolation procedure using the values v_{ij}^l .

The important property is that the detail coefficients give an easy representation of the smoothness of v since their size depends directly on the local regularity of v (Harten, 1996). Consequently it will help to localize the non-smooth structures of the solution, like shocks or contact discontinuities.

2.2. THE THRESHOLDING PROCEDURE

This procedure associates to each detail coefficient d_{ij}^l a coefficient b_{ij}^l taking value 0 or 1 according to the size of $|d_{ij}^l|$.

For a given parameter ϵ , a tolerance parameter at level l is defined as $\epsilon_l = \epsilon 2^{-l}$. Then, if the size of $|d_{ij}^l|$ is larger than ϵ_l , which means that the point (x_{2i}, y_{2j}) is in a non-smooth region, the value of b_{ij}^l and of its neighboring points are set to 1. Moreover, a test is also performed to check the decrease rate of the detail coefficients to take into account a possible formation of shocks (see (Bihari and Harten, 1997) for details).

Performing this procedure for all the scales $l = L, \dots, 1$, results in a mask, associated to the finest grid \mathcal{G}^0 , which will be used to determine the procedure to evaluate the numerical divergence \mathcal{D} .

2.3. MULTILEVEL COMPUTATION OF \mathcal{D}

Taking for example the simplest time evolution scheme,

$$\vec{U}_{ij}^{n+1} = \vec{U}_{ij}^n - \delta t \mathcal{D}_{ij}(\vec{U}^n), \quad (5)$$

the algorithm is justified intuitively by the following observation : if no non-smooth zone has been detected in \vec{U}^n around the point (x_i, y_j) and if no shock-formation has been forecasted by the thresholding procedure, the region around (x_i, y_j) remains smooth for \vec{U}^{n+1} . Indeed, due to relation (5), the numerical divergence is as smooth as the solution at this point and can be computed with a simple polynomial interpolation procedure. The algorithm is as follows :

- $\mathcal{D}_{ij}(\vec{U}^n)$ is computed using the high order scheme for all the points on the coarsest grid \mathcal{G}^L .
- Once the divergence is known on grid \mathcal{G}^l , its values on the finer grid \mathcal{G}^{l-1} are according to the mask coefficients :

if $b_{ij}^l = 1$ $\mathcal{D}_{ij}(\vec{U}^n)$ is computed with the high order scheme.

if $b_{ij}^l = 0$ $\mathcal{D}_{ij}(\vec{U}^n)$ is interpolated from its values on \mathcal{G}^l .

Decreasing l from L to 1 gives us the values of $\mathcal{D}(\vec{U}^n)$ on the finest grid \mathcal{G}^0 .

At each iteration of the algorithm, we then have to apply the three previous steps to compute \vec{U}^{n+1} from \vec{U}^n . To reduce the computational cost, the multiresolution is only performed for the density ρ since this variable displays all the possible discontinuities.

The time evolution is done with a third order Runge-Kutta method (Shu and Osher, 1989) and the multiresolution and thresholding procedures are only applied to the two first intermediate steps. Justifications of these modifications and quantification of their efficiency can be found in (Chiavassa and Donat, 1999).

3. Numerical experiments

The algorithm described in last section is applied to a classical test: the Double Mach reflexion of a strong shock (see e.g. in (Woodward and Colella, 1984)).

On figure 1(top), we display the density field of the reference solution, *i.e.* without multiresolution, computed at time $t = 0.2$ with Marquina's scheme for a grid of 512×128 points.

On figure 1(middle), the corresponding simulation with the multilevel algorithm is displayed and also the adaptive grid used for the computation. This grid is obtained representing only the points of \mathcal{G}^0 where the numerical divergence is computed with the high order scheme. The smoothness analysis done from the detail coefficients is very efficient and is able to localize perfectly the discontinuities. Other zones are detected near the bottom boundary for example, due to the presence of small oscillations in both reference and multilevel solutions. These perturbations are only due to the initial and boundary conditions (see (Woodward and Colella, 1984)).

It may be noticed that the mixing of interpolated and resolved values in the numerical divergence does not affect the quality of the results.

In figure 2 we display the L_1 norm of the error between the reference density and the one obtained after applying the multilevel algorithm, *i.e.*

$$e_1 = \frac{1}{N_p} \sum_{i=0}^{N_x} \sum_{j=0}^{N_y} |\rho_{ij}^n - \rho_{ref,ij}^n| \quad (6)$$

versus the tolerance ϵ . The fact that this error decreases at least as ϵ is crucial and ensures that we keep a tight control on the quality of the multilevel simulations. The efficiency θ , is measured as the ratio between the cpu time of the reference computation and that of the multilevel one, at time $t = 0.2$. In table 1 we have reported this gain for different sizes of the

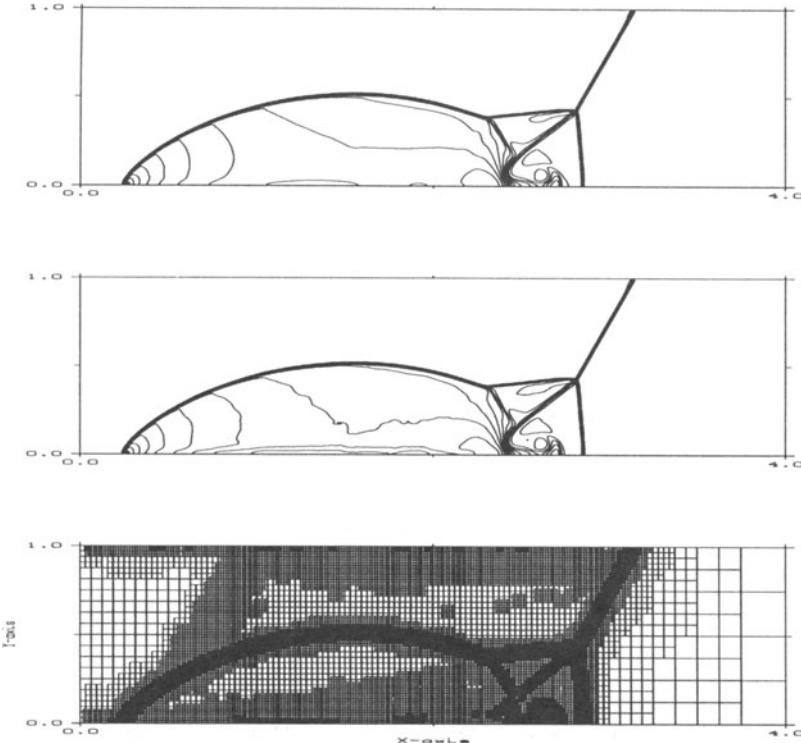


Figure 1. Density field of the reference solution (top), multilevel solution (middle) and corresponding adaptive grid (bottom). Initial grid \mathcal{G}^0 contains 512×128 points, number of levels is $L = 5$, and $\epsilon = 5 \times 10^{-3}$.

initial grid \mathcal{G}^0 . The minimum and maximum percentage of points where the numerical divergence is computed with Marquina's scheme during the simulation is also reported.

It may be seen that the multilevel approach is able to reduce significantly the cpu time. Moreover the gain increases for finer grid computations due to the decrease of the percentage of resolved fluxes. Therefore, this algorithm could be very useful for simulations requiring very fine mesh.

4. Conclusion

In this paper we present the main steps of a multilevel algorithm to be applied to a classical two dimensional test. We observe a significant reduction in execution time without changing the quality of the results. Moreover,

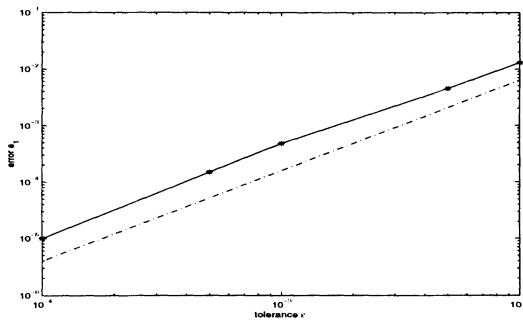


Figure 2. Error e_1 between the multilevel and reference algorithm versus the tolerance ϵ . The dotted line represents the curve of equation: $10 \epsilon^{1.6}$.

grid size \mathcal{G}^0	% f_{\min} – % f_{\max}	cpu gain θ
128×32	17.6 – 52.7	1.7
256×64	8.9 – 33.2	2.45
512×128	4.5 – 23.2	3.8

TABLE 1. percentage of resolved flux and cpu gain θ at time $t = 0.2$.

such approach is relatively simple to implement and does not require complex data structures thus, it could be easily introduced in existing CFD codes.

References

- G. Chiavassa and R. Donat. (1999) Numerical Experiments with Point-Value Multiresolution for 2D Compressible Flows. Preprint University of Valencia, available at <http://www.math.ntnu.no/conservation/2000/013.html>.
- B.L Bihari and A. Harten (1997) Multiresolution schemes for the numerical solutions of 2D conservation laws. *SIAM J. Sci. Comp.* **18**, pp315-354.
- R. Donat and A. Marquina (1996) Capturing shock reflexion: An improved flux formula. *J. Comput. Phys.* **125**, pp 42-58.
- A. Harten (1995) Multiresolution algorithms for the numerical solution of hyperbolic conservation laws. *Comm. Pure Appl. Math.* **48(12)**, pp1305-1342.
- A. Harten (1996) Multiresolution representation of data: a general framework. *SIAM J. Numer. Anal.* **33**, pp1205-1256.
- C. W Shu and S. J. Osher (1989) Efficient implementation of Essentially Non-oscillatory shock-capturing schemes. *J. Comput. Phys.* **83**, pp 32.
- B. Sjögreen (1995) Numerical Experiments with the multiresolution scheme for the compressible Euler equations. *J. Comput. Phys.* **117**, pp 251.
- P. R. Woodward and P. Colella (1984) The numerical simulation of two-dimensional fluid flow with strong shocks. *J. Comput. Phys.* **54**, pp 115.

VOLUME-OF-FLUID METHODS FOR PARTIAL DIFFERENTIAL EQUATIONS

PHILLIP COLELLA

NERSC Division

Lawrence Berkeley National Laboratory

University of California

Berkeley, CA 94720

Email: PColella@lbl.gov

1. Introduction

In this paper, we give an overview of a set of methods being developed for solving classical PDE's in irregular geometries, or in the presence of free boundaries. In this approach, the irregular geometry is represented on a rectangular grid by specifying the intersection of each grid cell with the region on one or the other side of the boundary. This leads to a natural conservative discretization of the solution to the PDE on either side of the boundary. This method has been used for a broad range of free boundary problems, including material interfaces in compressible (Noh and Woodward, 1976; Miller and Puckett, 1996) and incompressible (Hirt and Nichols, 1981; Puckett et al., 1997) flows, shocks (Chern and Colella, 1987; Bell, Colella and Welcome, 1991), and combustion fronts (Bourlioux and Majda, 1995; Hilditch and Colella, 1995; Pilliod and Puckett, 1997). In the case where the surface being represented in this way is an irregular domain boundary, these methods are often referred to as Cartesian grid or embedded boundary methods (Purvis and Burkhalter, 1979; Young et al., 1991). There has been considerable progress in the development of methods for generating the underlying grid description for complex three-dimensional geometries (Aftosmis, Berger and Melton, 1998), which makes this approach particularly attractive for complex engineering problems. These methods have been used in a variety of unsteady fluid dynamics problems, including inviscid compressible flow in three dimensions (Pember et al., 1995a), incompressible flow in two dimensions (Almgren, Bell, Colella and Marthaler, 1997; Calhoun, 1999) and low-Mach number combustion in an axisymmetric burner (Pember et al., 1995b).

The underlying discretization of space is given by rectangular control volumes on a Cartesian grid: $\Upsilon_i = [(\mathbf{i} - \frac{1}{2}\mathbf{u})h, (\mathbf{i} + \frac{1}{2}\mathbf{u})h]$, $\mathbf{i} \in \mathbb{Z}^d$, where d is the dimensionality of the problem, h is the mesh spacing, and \mathbf{u}

is the vector whose entries are all ones. In the case of a fixed, irregular domain Ω , the geometry is represented by the intersection of Ω with Cartesian grid (figure 1). We obtain control volumes $V_i = \Upsilon_i \cap \Omega$, and faces $A_{i \pm \frac{1}{2} e_s}$ that are the intersection of ∂V_i with the coordinate planes $\{\mathbf{x} : x_s = (i_s \pm \frac{1}{2})h\}$. We also define A_i^B to be the intersection of the boundary of the irregular domain with the Cartesian control volume: $A_i^B = \partial \Omega \cap \Upsilon_i$. We will assume here that there is a one-to-one correspondence between the control volumes and faces and the corresponding geometric entities on the underlying Cartesian grid. The description can be generalized to allow for boundaries whose width is less than the mesh spacing, or sharp trailing edges.

In order to construct finite difference methods, we will need only a small number of real-valued quantities that are derived from these geometric objects.

- The areas / volumes, expressed in dimensionless terms: volume fractions $\kappa_i = |V_i| h^{-d}$, face apertures $\alpha_{i \pm \frac{1}{2} e_s} = |A_{i \pm \frac{1}{2} e_s}| h^{-(d-1)}$ and boundary apertures $\alpha_i^B = |A_i^B| h^{-(d-1)}$. We assume that we can compute estimates of the dimensionless quantities that are accurate to $O(h^2)$.
- The locations of centroids, and the average outward normal to the boundary.

$$\begin{aligned}\mathbf{x}_i &= \frac{1}{|V_i|} \int_{V_i} \mathbf{x} dV \\ \mathbf{x}_{i \pm \frac{1}{2} e_s} &= \frac{1}{|A_{i \pm \frac{1}{2} e_s}|} \int_{A_{i \pm \frac{1}{2} e_s}} \mathbf{x} dA \\ \mathbf{x}_i^B &= \frac{1}{|A_i^B|} \int_{A_i^B} \mathbf{x} dA \\ \mathbf{n}_i^B &= \frac{1}{|A_i^B|} \int_{A_i^B} \mathbf{n}^B dA\end{aligned}$$

where \mathbf{n}^B is the outward normal to $\partial \Omega$, defined for each point on $\partial \Omega$. Again, we assume that we can compute estimates of these quantities that are accurate to $O(h^2)$.

Using just these quantities, we can define conservative discretizations for the divergence operator. Let $\vec{F} = (F^1 \dots F^d)$ be a function of \mathbf{x} . Then

$$\begin{aligned}\nabla \cdot \vec{F} &\approx \frac{1}{|V_i|} \int_{V_i} \nabla \cdot \vec{F} dV = \frac{1}{|V_i|} \int_{\partial V_i} \vec{F} \cdot \mathbf{n} dA \\ &\approx \frac{1}{\kappa_i h} \left(\sum_{\pm=+,-} \sum_{s=1}^d \pm \alpha_{i \pm \frac{1}{2} e_s} F^s(\mathbf{x}_{i \pm \frac{1}{2} e_s}) + \alpha_i^B \mathbf{n}_i^B \cdot \vec{F}(\mathbf{x}_i^B) \right)\end{aligned}\tag{1}$$

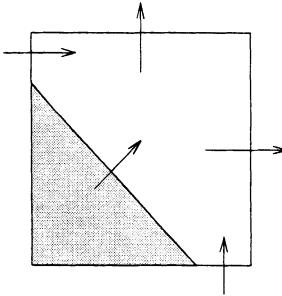


Figure 1. Cartesian grid cell containing volume-of-fluid representation of an irregular boundary. The shaded region indicates the part of the cell not contained in the irregular domain. The arrows indicate the centering of fluxes in a finite volume approximation to the divergence.

where (1) is obtained by replacing the integrals of the normal components of the vector field \vec{F} with the values at the centroids.

This representation of finite differences in irregular geometries is very appealing for a variety of reasons. As we said above, the problem of grid generation has already been solved. More generally, the regular geometric structure of rectangular grids simplifies a variety of issues, including control of the truncation error, the development of efficient iterative methods based on multigrid, the introduction of adaptive meshes, and the coupling to other physical submodels, such as particles or radiation. In order to understand these methods, we want to place them in a more systematic numerical analysis setting. Among the tools we will use are truncation error analysis, with modified equation models having singular right-hand sides, and the development of specialized discretizations that maintain uniform stability and / or conditioning in the presence of arbitrarily small volume fractions.

2. Poisson's Equation on an Irregular Domain

We consider first the Neumann problem for Poisson's equation on an irregular domain Ω . Our treatment follows that in (Johansen and Colella, 1998; Johansen, 1997).

$$\Delta\psi = \rho \text{ on } \Omega \quad (2)$$

$$\frac{\partial\psi}{\partial n} = g \text{ on } \partial\Omega$$

The Laplacian can be written as the divergence of a flux: $\Delta\psi = \nabla \cdot \vec{F}$, $\vec{F} = \nabla\psi$. The discretized solution values approximate the solution to

the PDE at the rectangular cell centers: $\phi_i \approx \psi(ih)$ (figure 2). At first glance, this might be a cause for concern, since some of the centers of Cartesian cells Υ_i might not be contained in Ω . However, it is well-known that, for any domain with smooth boundary, a smooth function can be extended to all of \mathbb{R}^d with a bound on the relative increase in the $C^{k,\alpha}$ norms that depends only on the domain and (k, α) (Gilbarg and Trudinger, 1977). We assume that the values of ψ on the covered cell centers are given by such an extension.

Using the discretization of the divergence defined in (1), we can define a discretization of Poisson's equation as follows.

$$(\Delta^h \phi)_i = \rho_i \quad (3)$$

$$(\Delta^h \phi)_i = \frac{1}{\kappa_i h} \left(\sum_{\pm=+,-} \sum_{s=1}^d \pm \alpha_{i \pm \frac{1}{2} e_s} F_{i \pm \frac{1}{2} e_s}^s + \alpha_i^B g(x_i^B) \right) \quad (4)$$

where $\rho_i = \rho(\mathbf{x}_i)$. The fluxes on the cell faces are computed by linearly interpolating between centered difference approximations. For example, in two dimensions,

$$F_{i+\frac{1}{2},j}^s = \eta \frac{(\phi_{i+1,j} - \phi_{i,j})}{h} + (1 - \eta) \frac{(\phi_{i+1,j \pm 1} - \phi_{i,j \pm 1})}{h} \quad (5)$$

$$\eta = \frac{|y_{i+\frac{1}{2},j} - jh|}{h} \quad (6)$$

where $\pm = + (-)$ if $y_{i+\frac{1}{2},j} > jh (< jh)$.

There has been no rigorous analysis of the conditioning of the linear system arising from solving the equations (3) - (6). The apparent singularity in the operator with respect to κ can be removed by diagonal scaling. If we define L^h by $(L^h \phi)_i = \kappa_i (\Delta^h \phi)_i$, we expect that the system has the same asymptotic conditioning behavior as that of the operator in the absence of irregular boundaries, i.e. that, for homogeneous boundary conditions, L^h is a linear operator with $\text{cond}(L^h) = O(h^{-2})$, uniformly with respect to the range of values taken on by κ .

We define the truncation error in the usual fashion: $\tau = \rho - \Delta^h \phi^{\text{exact}}$, where $\phi_i^{\text{exact}} = \psi(ih)$. We then have the following asymptotic error estimates for the truncation error:

$$\tau_i = O(h^2) \text{ if } i \text{ is an interior cell} \quad (7)$$

$$= O\left(\frac{h}{\kappa_i}\right) \text{ if } i \text{ is an irregular cell} \quad (8)$$

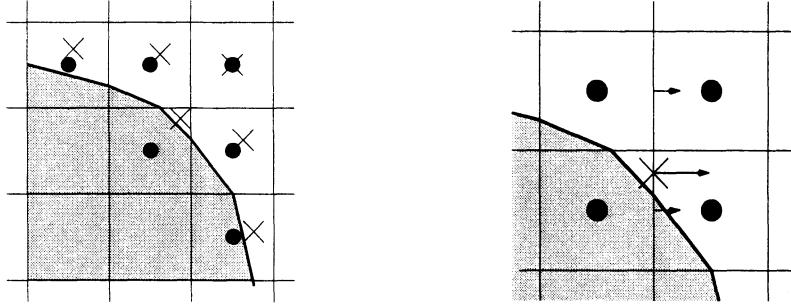


Figure 2. Left: Centering of dependent variables, operators. The primary dependent variables ϕ are centered at the centers of the Cartesian cells (solid circles), while $\nabla \cdot F$ is centered at the centroids of the control volumes (\times). Right: computation of an $O(h^2)$ value of the flux at a cell edge. We compute centered differences at the centers of the two Cartesian grid edge edges indicated in the figure, and interpolate linearly between the them.

We refer to methods that satisfy error estimates of the form (8) on the irregular control volumes as being *formally consistent*.

There are two apparent problems with this truncation error estimate: it is only first order accurate at the boundary, and it is singular in its dependence on κ . Nonetheless, we observe robust second-order convergence of the solution in max norm. These two facts can be reconciled using a modified equation analysis. The solution error $\xi = \phi - \phi^{exact}$ satisfies the error equation $\Delta^h \xi = \tau$ with homogeneous boundary conditions. We expect that the error ξ behaves like a solution to Poisson's equation (2) with homogeneous boundary conditions: $\tilde{\xi}(ih) = \xi_i + O(h^3)$, with $\Delta \tilde{\xi} = \tilde{\tau}$. Here, $\tilde{\tau}$ is a piecewise constant function on each control volume: $\tilde{\tau}|_{V_i} = \tau_i$. In that case, the total charge on each irregular control volume is $O(\frac{h}{\kappa_i}) \times \kappa_i h^2 = O(h^3)$. This leads to a contribution from each irregular control volume to the solution error of a smooth function of magnitude $O(h^3)$. Since there are only $O(h^{-1})$ such cells, their contribution to the error is of magnitude $O(h^2)$. A similar argument applies to the interior cells. There are $O(h^{-2})$ of them, each with a total charge of $O(h^4)$. This leads to the conclusion that $\xi = O(h^2)$.

For Dirichlet boundary conditions (e.g. $\phi = g$ on $\partial\Omega$), the equations are modified to account for the additional flux on the boundary.

$$\vec{F}^B \cdot \mathbf{n}^B = \frac{\partial \phi}{\partial n}$$

where the right-hand side is computed using interpolation from the grid values and the value of the boundary conditions. In order to avoid conditioning problems, we use a stencil for which the boundary centroid and the other interpolation points are separated from one another by

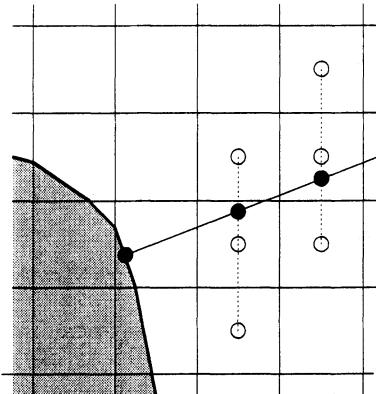


Figure 3. Flux calculation for Dirichlet boundary conditions in 2D. We compute the intersection of a ray originating at x_i^B in the direction normal to the body, and compute its intersection along two parallel coordinate lines whose distance along the ray from x_i^B is at least $\frac{1}{2}h$. We then use quadratic interpolation from the open circles to compute the values at the intersection points, and use those two values and the boundary condition to compute an $O(h^2)$ approximation to the flux.

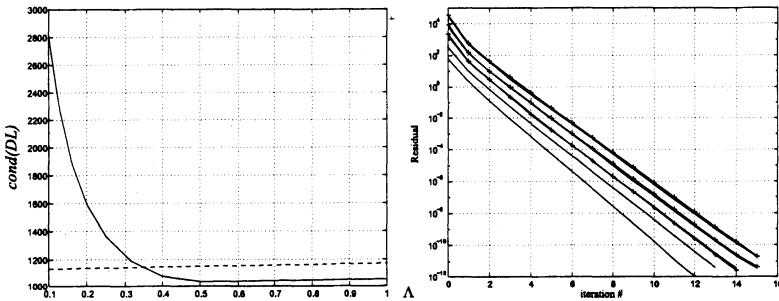


Figure 4. Left: condition number of L^h for a one-dimensional problem containing a single irregular control volume as a function of the volume fraction of that cell. The dashed line is the algorithm described here, while the solid line is the condition number of a standard piecewise linear Galerkin method. Right: residual as a function of number of multigrid V-cycles for a series of two-dimensional problems. The grids range from 40×40 to 640×640 , with the number of iterations required increasing slowly as with the number of grid points. The lines with symbols correspond to calculations done with local mesh refinement.

at least $\frac{1}{2}h$ (figure 3). This leads to a condition number for L^h that is bounded independent of κ , and comparable to that of the uniform grid algorithm (figure 4). An indirect indicator of good conditioning properties is uniform multigrid performance over an order of magnitude in mesh spacing.

If the truncation error is computed as above, the same result as in (8) is obtained. However, the influence on the solution error is somewhat different. The modified equation analysis says that, for each irregular control volume, there is a field induced by a total charge of magnitude

$O(h^3)$. However, that field satisfies a homogeneous Dirichlet boundary condition for a boundary located a distance $O(h)$ from the centroid of the control volume. By the method of images, for example, we see that the field induced by that charge is a dipole of strength $O(h^4)$. For that reason, the total effect on the solution error of the truncation error at the irregular boundary is $O(h^3)$.

3. Hyperbolic Conservation Laws.

We want to apply the ideas in the previous section to understand the various volume-of-fluid methods that have been developed for systems of hyperbolic conservation laws.

$$\frac{\partial W}{\partial t} + \nabla \cdot \vec{F} = 0$$

$$W = W(\mathbf{x}, t), \quad \mathbf{x} \in \Omega; \quad \vec{F} = \vec{F}(W); \quad W, F^s \in \mathbb{R}^m. \quad (9)$$

We discretize the solution to (9) in space and time, using the same spatial discretization as before: $U_i^n \approx W(ih, n\Delta t)$. We can also use the quadrature rule (1) to construct the following conservative discretization of $\nabla \cdot F$.

$$(\nabla \cdot \vec{F})^C = \frac{1}{\kappa_i h} \left(\sum_{\pm=+, -} \sum_{s=1}^d \pm \alpha_{i \pm \frac{1}{2} e_s} F_{i \pm \frac{1}{2} e_s}^s + \alpha_i^B \vec{F}_i^B \cdot \mathbf{n}_i^B \right) \quad (10)$$

Ideally we would like to use an explicit finite difference approximation to compute $F_{i \pm \frac{1}{2} e_s}^s \approx F^s(\mathbf{x}_{i \pm \frac{1}{2} e_s})$, $\vec{F}_i^B \approx \vec{F}(\mathbf{x}_i^B)$, and use (10) to compute the discrete evolution of U .

$$U_i^{n+1} = U_i^n - \Delta t (\nabla \cdot \vec{F})_i^C \quad (11)$$

If the fluxes are computed with sufficient accuracy, the resulting method is formally consistent. In addition, the method satisfies the following discrete conservation identity.

$$\sum_{i \in D} \kappa_i U_i^{n+1} = \sum_{i \in D} \kappa_i U_i^n - \frac{\Delta t}{h} \sum_{f \in \partial D} \alpha_f \vec{F}_f \cdot \mathbf{n}_f \quad (12)$$

where D is any collection of control volumes, and ∂D is the set of cell faces and boundary faces forming the boundary of D . The difficulty with this approach is that the CFL stability constraint on the time step is at best $\Delta t = O(h v_i^{max} (\kappa_i)^{\frac{1}{d}})$, where v_i^{max} is the magnitude

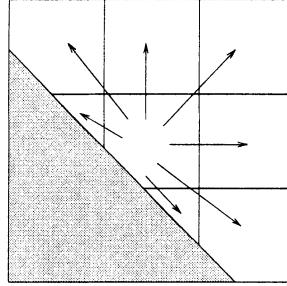


Figure 5. The arrows indicate the control volumes to which δM would be redistributed from the central control volume.

of the maximum wave speed for the i -th control volume. This is the well-known small-cell problem for embedded boundary methods. There have been a number of proposals to deal with this problem, including merging the small control volumes with nearby larger ones, and the development of specialized stencils that guarantee the required cancellations in (10) (Berger and Leveque, 1990). The approach we have taken to this problem has been to expand the range of influence of the small control volumes algebraically to obtain a stable method (Chern and Colella, 1987; Bell, Colella and Welcome, 1991). The starting point for this approach is to compute a stable, but nonconservative approximation to $\nabla \cdot \vec{F}$. One computes the flux difference on the full Cartesian cell, assuming the embedded boundary wasn't there.

$$(\nabla \cdot \vec{F})_i^{NC} = \frac{1}{h} \sum_{\pm=+,-} \sum_{s=1}^d \pm F_{i \pm \frac{1}{2} e_s}^s$$

where the fluxes in this expression are centered at $(i \pm \frac{1}{2} e_s)h$. The initial update uses a linear hybridization of the two estimates of $\nabla \cdot \vec{F}$.

$$U_i^{n+1} = U_i^n - \Delta t (\eta_i (\nabla \cdot \vec{F})_i^C + (1 - \eta_i) (\nabla \cdot \vec{F})_i^{NC}) \quad (13)$$

If we choose, for example, $\eta_i = \kappa_i$, then the small denominator in $(\nabla \cdot \vec{F})^C$ is cancelled, and we obtain a stable method. However, the method fails to conserve, in that it does not satisfy an identity of the form (12). One measure of that lack of conservation is given by the difference in the mass contained in V_i that one would have obtained from using (11) and (13):

$$\kappa_i ((11)_i - (13)_i) = \delta M_i = -\Delta t \kappa_i (1 - \eta_i) ((\nabla \cdot \vec{F})_i^C - (\nabla \cdot \vec{F})_i^{NC})$$

To maintain overall conservation, we redistribute δM into nearby cells (figure 5).

$$U_{i'}^{n+1} := U_{i'}^{n+1} + w_{i,i'} \delta M_i, \quad i' \in N(i). \quad (14)$$

$$w_{\mathbf{i}, \mathbf{i}'} \geq 0, \quad \sum_{\mathbf{i}' \in N(\mathbf{i})} w_{\mathbf{i}, \mathbf{i}'} \kappa_{\mathbf{i}'} = 1 \quad (15)$$

Where $N(\mathbf{i})$ is some set of indices in the neighborhood of \mathbf{i} . The sum condition (15) is what makes the redistribution step conservative. In that case, a relationship of the form (12) is satisfied, with some additional boundary terms corresponding to redistribution into or out of the domain D . In addition, $w_{\mathbf{i}, \mathbf{i}'}$ must be bounded independent of $(\kappa_{\mathbf{i}'})^{-1}$. One example of a redistribution strategy that meets our requirements is $w_{\mathbf{i}, \mathbf{i}'} = (\sum_{\mathbf{i}' \in N(\mathbf{i})} \kappa_{\mathbf{i}'})^{-1}$, where $N(\mathbf{i})$ is a set of indices whose components differ from those of \mathbf{i} by no more than one. For problems in gas dynamics involving strong shocks, mass-weighted redistribution has been observed to be more robust. (Pember et al., 1995a).

We can analyze the accuracy of these methods using the machinery we have developed. If we define $U_i^{n,exact} = W(\mathbf{i}h, n\Delta t)$, then the truncation error is defined to be

$$\tau_i^{n+\frac{1}{2}} = \frac{(U_i^{n+1,exact} - U_i^{n,exact})}{\Delta t} - L(U^{n,exact})_i$$

where $\Delta t L(U)$ denotes the increment of the discrete solution by one time step outlined above, given data U at the beginning of the time step. In (Modiano and Colella, 2000) fluxes are constructed for the calculation of $(\nabla \cdot \vec{F})^C$ that are $O(h^2)$ estimates of the fluxes evaluated at the face centroids and at time $(n + \frac{1}{2})\Delta t$. This is done by using the unsplit second-order Godunov method in (Colella, 1990) to construct fluxes at the centers of the Cartesian grid faces, followed by linear interpolation to obtain $O(h^2)$ fluxes. For example, in two dimensions,

$$F_{i+\frac{1}{2},j}^s = \eta F_{i+\frac{1}{2},j}^{n+\frac{1}{2},G} + (1 - \eta) F_{i+1,j \pm 1}^{n+\frac{1}{2},G}$$

where $F_{i+\frac{1}{2},j}^{n+\frac{1}{2},G}$ is the flux centered at $((i + \frac{1}{2})h, jh)$ computed using the second-order Godunov method on the regular grid, and η is computed as in (6). \vec{F}_i^B is computed using a second-order extrapolation in space and time from the center of the Cartesian cell, in order to obtain an $O(h^2)$ estimate to $\vec{F}(\mathbf{x}_i^B, (n + \frac{1}{2})\Delta t)$. The fluxes in the calculation $(\nabla \cdot \vec{F})^{NC}$ are given by the Godunov fluxes at the centers of the Cartesian grid faces at time $(n + \frac{1}{2})\Delta t$.

This leads to a truncation error estimate

$$\begin{aligned} \tau_i^{n+\frac{1}{2}} &= O(h^2) \text{ if } \mathbf{i} \notin N(\mathbf{i}') \text{ for all irregular control volumes } \mathbf{i}' \\ &= O(h) \text{ otherwise, uniformly with respect to } \kappa \end{aligned} \quad (16)$$

This truncation error estimate follows from the formal consistency of $(\nabla \cdot \vec{F})^C$, while $(\nabla \cdot \vec{F})^{NC}$ has a truncation error of $O(h)$, uniformly with

respect to κ . From this it follows that the truncation error of the hybrid method (13) satisfies (16). Since $\delta M = O(h\Delta t)$, the redistribution step only expands slightly the region near the boundary where the method is first-order accurate.

The behavior of these methods can be understood from a modified equation analysis. We expect that the solution to the modified equation

$$\frac{\partial W^{mod}}{\partial t} + \nabla \cdot (\vec{F}(W^{mod})) = \tilde{\tau} \quad (17)$$

approximates the numerical solution to one order higher accuracy than does W , the solution to the original conservation laws. As before, $\tilde{\tau}$ is the piecewise constant interpolation of the grid function $\tau^{n+\frac{1}{2}}$ in space and time over each control volume.

In (Pember et al., 1995a), the midpoint-centered fluxes used to compute $(\nabla \cdot \vec{F})^{NC}$ are also used to compute $(\nabla \cdot \vec{F})^C$, without applying the linear interpolation step (5) - (6). Since these fluxes approximate the fluxes evaluated at the centroids only to $O(h)$, the analysis here shows that these methods are inconsistent, with a truncation error of $\tau = O(1)$ near the irregular control volumes. Although these algorithms often give satisfactory results, there are cases for which such algorithms fail to converge in max norm, while the the results for the consistent algorithm are second-order accurate in L^1 , and first-order accurate in max norm (Modiano and Colella, 2000).

Critical to the success of this approach is the calculation of $(\nabla \cdot \vec{F})^{NC}$. In control volumes with $\kappa_i << 1$, $(\nabla \cdot \vec{F})^{NC}$ is almost entirely responsible for the update of U_i . For that reason, $(\nabla \cdot \vec{F})^{NC}$ must be designed carefully, so that, for example, the solution on small control volumes comes into equilibrium with the larger ones around it. For the consistent algorithm, there is the additional requirement that the truncation error for all of the fluxes is $O(h^2)$, a condition which is somewhat delicate to obtain for faces whose apertures vanish.

4. Moving and Free Boundaries

We can generalize the approach for hyperbolic problems described above to the case of boundaries that move. Specifically, the domain Ω is now a function of time, $\Omega = \Omega(t)$, and the various geometric quantities can also be computed in a time-dependent way: $\kappa_i(t)$, $\alpha_{i+\frac{1}{2}e_s}(t)$, $\mathbf{x}_{i+\frac{1}{2}e_s}(t)$, etc. We denote the values of these quantities at discrete times by a superscript, e.g., $\kappa_i^n = \kappa_i(n\Delta t)$. For the face-centered quantities, we also need the values of these quantities averaged over a time step, as

well as aperture-weighted averages of the time.

$$\bar{q} = \frac{1}{\Delta t} \int_{n\Delta t}^{(n+1)\Delta t} q(t') dt' , q = \alpha_{\mathbf{i} + \frac{1}{2}\mathbf{e}_s}, \mathbf{x}_{\mathbf{i} + \frac{1}{2}\mathbf{e}_s}, \alpha_{\mathbf{i}}^B, \mathbf{x}_{\mathbf{i}}^B.$$

$$\begin{aligned} \bar{t}_{\mathbf{i} + \frac{1}{2}\mathbf{e}_s} &= \frac{1}{\Delta t \bar{\alpha}_{\mathbf{i} + \frac{1}{2}\mathbf{e}_s}} \int_{n\Delta t}^{(n+1)\Delta t} t' \alpha_{\mathbf{i} + \frac{1}{2}\mathbf{e}_s}(t') dt' \\ \bar{t}_{\mathbf{i}}^B &= \frac{1}{\Delta t \bar{\alpha}_{\mathbf{i}}^B} \int_{n\Delta t}^{(n+1)\Delta t} t' \alpha_{\mathbf{i}}^B(t') dt' \end{aligned} \quad (18)$$

Given these quantities, we can derive a quadrature formula for (9) analogous to (1) by integrating in space-time.

$$\begin{aligned} \frac{1}{h^d} \int_{\mathcal{V}_{\mathbf{i}}^{n,n+1}} \frac{\partial W}{\partial t} + \nabla \cdot \vec{F} dV dt &\approx \\ \kappa_{\mathbf{i}}^{n+1} W(\mathbf{x}_{\mathbf{i}}^{n+1}, (n+1)\Delta t) - \kappa_{\mathbf{i}}^n W(\mathbf{x}_{\mathbf{i}}^n, n\Delta t) \\ + \frac{\Delta t}{h} \left(\sum_{\pm=+, -} \sum_{s=1}^d \pm \bar{\alpha}_{\mathbf{i} \pm \frac{1}{2}\mathbf{e}_s} F^s(\bar{\mathbf{x}}_{\mathbf{i} \pm \frac{1}{2}\mathbf{e}_s}, \bar{t}_{\mathbf{i} \pm \frac{1}{2}\mathbf{e}_s}) \right. \\ \left. + \bar{\alpha}_{\mathbf{i}}^B (\bar{\mathbf{n}}^B \cdot \vec{F}(\bar{\mathbf{x}}_{\mathbf{i}}^B, \bar{t}_{\mathbf{i}}^B) - \bar{s}_{\mathbf{i}}^B W(\bar{\mathbf{x}}_{\mathbf{i}}^B, \bar{t}_{\mathbf{i}}^B)) \right) \end{aligned} \quad (19)$$

where $\mathcal{V}_{\mathbf{i}}^{n,n+1} = \{(\mathbf{x}, t) : \mathbf{x} \in V_{\mathbf{i}}(t), n\Delta t \leq t \leq (n+1)\Delta t\}$. Here $\bar{s}_{\mathbf{i}}^B$ is the space-time averaged value of the normal component of the boundary velocity, computed in a similar fashion to (18).

This formula is the starting point for deriving a conservative difference approximation to (9), replacing the values in (19) by suitable difference approximations.

$$\begin{aligned} (\nabla \cdot \vec{F})^C &\approx \frac{1}{\kappa_{\mathbf{i}}^{n+1} \Delta t} \{ \kappa_{\mathbf{i}}^{n+1} \tilde{U}_{\mathbf{i}}^{ref} - \kappa_{\mathbf{i}}^n \tilde{U}_{\mathbf{i}}^n \\ &+ \frac{\Delta t}{h} \left(\sum_{\pm=+, -} \sum_{s=1}^d F_{\mathbf{i} \pm \frac{1}{2}\mathbf{e}_s}^s \bar{\alpha}_{\mathbf{i} \pm \frac{1}{2}\mathbf{e}_s} \right. \\ &\left. + \bar{\alpha}_{\mathbf{i}}^B (\vec{F}_{\mathbf{i}}^B \cdot \bar{\mathbf{n}}_{\mathbf{i}}^B - \bar{s}_{\mathbf{i}}^B U_{\mathbf{i}}^B)) \right\} \end{aligned} \quad (20)$$

where $\tilde{U}_{\mathbf{i}}^n \approx W(\mathbf{x}_{\mathbf{i}}^{n+1}, (n+1)\Delta t)$ and $\tilde{U}_{\mathbf{i}}^{ref} \approx W(\mathbf{x}_{\mathbf{i}}^{n+1}, n\Delta t)$. The conservative update corresponding to (10) is given by

$$\tilde{U}_{\mathbf{i}}^{n+1} = \tilde{U}_{\mathbf{i}}^{ref} - \Delta t (\nabla \cdot \vec{F})^C \quad (21)$$

The above method satisfies a discrete conservation identity of the following form.

$$\sum_{\mathbf{i} \in D} \kappa_{\mathbf{i}}^{n+1} \tilde{U}_{\mathbf{i}}^{n+1} = \sum_{\mathbf{i} \in D} \kappa_{\mathbf{i}}^n \tilde{U}_{\mathbf{i}}^n + \frac{\Delta t}{h} \sum_{\mathbf{f} \in \partial D} (\dots)_{\mathbf{f}} \quad (22)$$

We note that this identity is satisfied independent of our choice of $\tilde{U}_{\mathbf{i}}^{ref}$.

This method has the same small-cell stability problem as occurs in the fixed-domain case. Indeed, for control volumes such that $\kappa_{\mathbf{i}}^n > 0$, $\kappa_{\mathbf{i}}^{n+1} = 0$, $(\nabla \cdot \vec{F})^C$ would have to vanish for the method to be stable. One can deal with this problem in exactly the same fashion as in the fixed-boundary case, by hybridizing the conservative update with one that is non-conservative, but stable, and redistributing the difference to nearby control volumes.

$$\tilde{U}_{\mathbf{i}}^{n+1} = \tilde{U}_{\mathbf{i}}^{ref} - \Delta t (\eta_{\mathbf{i}} (\nabla \cdot \vec{F})^C + (1 - \eta_{\mathbf{i}}) (\nabla \cdot \vec{F})^{NC}) \quad (23)$$

$$\tilde{U}_{\mathbf{i}'}^{n+1} := \tilde{U}_{\mathbf{i}'}^{n+1} + w_{\mathbf{i}, \mathbf{i}'} \delta M_{\mathbf{i}}, \quad \mathbf{i}' \in N(\mathbf{i})$$

where $\delta M_{\mathbf{i}}$, $w_{\mathbf{i}, \mathbf{i}'}$, and $N(\mathbf{i})$ are computed, for $\kappa_{\mathbf{i}}^{n+1} > 0$, as in the fixed-boundary case by taking $\kappa_{\mathbf{i}} = \kappa_{\mathbf{i}}^{n+1}$. In the case where $\kappa_{\mathbf{i}}^{n+1} = 0$, the denominator in $(\nabla \cdot \vec{F})^C$ vanishes, and the various formulas defining the update of $\tilde{U}_{\mathbf{i}}$ and $\delta M_{\mathbf{i}}$ are taken to be the limit of the expressions for $\kappa_{\mathbf{i}}^{n+1} > 0, \kappa_{\mathbf{i}}^{n+1} \rightarrow 0$. The non-conservative estimate to $\nabla \cdot \vec{F}$ is also computed as in the fixed-boundary case by computing the flux differences as if the boundary was not present, extending the solution as required into covered cells.

Moving-boundary algorithms based on these ideas have been developed in the context of free-boundary problems, such as shock tracking (Chern and Colella, 1987; Bell, Colella and Welcome, 1991), and the tracking of combustion fronts (Hilditch and Colella, 1995; Pilliod and Puckett, 1997). In these problem, the total domain is divided into two time-varying domains $\Omega^{(1)}(t)$, $\Omega^{(2)}(t)$, with a common moving boundary $\partial\Omega^{(1/2)}(t)$. In summary, algorithms for free-boundary problems consist of the following steps.

- Compute the motion of the free boundary and the geometric information required for the conservative updates on either side of the boundary.
- Compute fluxes on the faces on either side of the boundary, using finite difference approximations, and compute the flux across the free boundary using the jump relations

$$s(W^{(1)} - W^{(2)}) = (\vec{F}^{(1)} - \vec{F}^{(2)}) \cdot \mathbf{n}^B$$

- Compute $(\nabla \cdot \vec{F})^C$, $(\nabla \cdot \vec{F})^{NC}$ on either side of the front, and compute the preliminary updates (23).
- Redistribute the increments δM on both sides of the front.

The most substantial modification of the algorithm to accommodate the free boundary is in the redistribution step. In general, signals can propagate across a free boundary, and we design the redistribution step to appropriately represent that wave propagation. For example, we decompose $\delta M_i^{(1)} = \sum_{k=1}^m \beta_k r_k$, where (λ_k, r_k) are the eigenvalues and right eigenvectors of the linearized coefficient matrix for the system projected in the outward ($1 \rightarrow 2$) normal direction. We use this decomposition to split δM into two pieces: one corresponding to signals crossing the free boundary, and the remainder, corresponding to signals that propagate into the domain $\Omega^{(1)}$.

$$\begin{aligned}\delta M_i^{(1),+} &= \sum_{\lambda_k > s} \beta_k r_k \\ \delta M_i^{(1),-} &= \delta M - \delta M^{(1),+}\end{aligned}$$

Then $\delta M_i^{(1),+}$ is distributed into the control volumes in $\Omega^{(2)}((n+1)\Delta t)$, while $\delta M_i^{(1),-}$ is distributed into the control volumes in $\Omega^{(1)}((n+1)\Delta t)$, in both cases using the appropriate version of (14). There is some indeterminacy in situations in which one of the wave speeds coincides with that of the free boundary. For example, if the free boundary is a genuinely nonlinear k_0 -shock, with characteristics approaching from both sides of the discontinuity being overtaken by the shock, we have some freedom as to how we distribute the the k_0 component of the increment. For the calculations of gas-dynamic shocks in (Chern and Colella, 1987), the increments were systematically redistributed to the high-pressure side. On the other hand, if the k_0 wave is linearly degenerate, the increment corresponding to that wave should not cross the free boundary, and should be assigned to $\delta M^{(1),-}$.

The construction of formally consistent discretizations is significantly more complicated in the moving boundary case than in the fixed boundary case, although the principles are the same. The numerical fluxes should approximate those at the appropriate centroids to second order: $F_{i+\frac{1}{2}e_s}^s = F^s(\bar{x}_{i+\frac{1}{2}e_s}, \bar{t}_{i+\frac{1}{2}e_s}) + O(h^2)$, $\vec{F}_i^B = \vec{F}(\bar{x}_i^B, \bar{t}_i^B) + O(h^2)$. Fluxes appropriately centered in space can be obtained as in the fixed boundary case, with a correction to shift the centering in time as required. The centering of \tilde{U}^n , is more complicated, as these values differ from the values at the cell centers by $O(h)$. This difficulty does not arise in the fixed boundary case, since in that limit, the $O(h)$ terms from the

old and new times cancel to leading order. One approach would be to construct an $O(h^2)$ interpolant from the values U_i^n . The difficulty with this approach is that it can lead to an implicit method for the update of U . An approach that avoids that problem is to compute a provisional update using $(\nabla \cdot \vec{F})^{NC}$, and use the values of U at the new time so obtained to compute the corrections to U^{n+1} needed to obtain \tilde{U}^{n+1} . Similar corrections must be computed to obtain an $O(h^2)$ value for \tilde{U}^{ref} , but this is more straightforward, since \tilde{U}^{ref} is centered at the beginning of the time step.

In fact, none of the moving boundary calculations performed to date have been done using formally consistent discretizations. Instead, all of the values have been approximated using $O(h)$ values, with $F_{i+\frac{1}{2}\epsilon_s}^s$ computed as being centered at $(i + \frac{1}{2}\epsilon_s)h$, and $\tilde{U}^n, \tilde{U}^{ref} = U^n$. Nonetheless, these methods have produced convergent results. The reason for this can be explained using the modified equation analysis discussed above. For these problems, $\partial\Omega^{(1/2)}(t)$ is non-characteristic for the equations of gas dynamics being solved on either side of the front in these problems. In that case, one expects that signals would be subject to the region of high truncation error for only an $O(h\lambda^{-1})$ length of time, where λ is the speed of propagation of the signal. So, even though the truncation error is $O(1)$ near the boundary, the contribution to the maximum amplitude of the solution error of that truncation error is $O(h\lambda^{-1})$.

5. Conclusions

We presented here a theory for constructing finite difference methods using volume-of-fluid representations of irregular boundaries. The key component is the systematic interpretation of the dependent variables as approximating the solution evaluated at the centers of rectangular grid cells. We have mainly focused on issues surrounding the construction of stable and consistent discretizations of elliptic and hyperbolic equations. There are several other issues that we have not discussed in detail.

Parabolic equations. The ideas described here have been used to discretize the heat equation for both fixed and moving boundaries, and to solve a simplified form of the Stefan problem (Johansen, 1997). In the fixed domain case, this is a straightforward application of the elliptic discretization ideas discussed above, combined with a suitable time discretization. One issue that arises in that setting is the need to use an L_0 -stable implicit Runge-Kutta method, such as in (Twizell,

Gumel and Arigu, 1996), rather than Crank-Nicholson, particularly for the case of moving boundaries.

Representing moving geometries. The discretization methods described above are compatible with a variety of surface representations, including methods based on surface triangulations (Chern et al., 1986; Aftosmis, Berger and Melton, 1998), level-set methods (Osher and Sethian, 1988) and volume-of-fluid representations of the surface itself (Noh and Woodward, 1976; Hirt and Nichols, 1981). In the latter approach, the geometry of the front is determined from the volume fractions κ_i , whose evolution is given by solving a transport equation for them based on a local reconstruction to the geometry. This leads to a method that is well-behaved under large distortions and changes in the topology. It also provides a particularly robust representation for the free boundary case, as the coupling with the conservative discretizations degrades gracefully if the geometry is underresolved. There has been considerable recent progress in the development of accurate versions of such methods (Helmsen, Colella and Puckett, 1997). Another approach to moving boundaries is to use the discretization methods described here to represent the solution on a combination of overlapping logically rectangular meshes, where some of the meshes are Lagrangian, others Eulerian (Noh, 1963). In fact, many of the discretization ideas described in this paper first appeared in this setting.

Thin bodies and non-rectangular data structures. The method described above can be generalized to allow for geometries for which the irregular control volumes do not have a one-to-one correspondence with the Cartesian grid cells. This occurs, for example, for domains that are the exterior of thin bodies whose width is less than one cell. To represent such cases, one can generalize the Cartesian grid index space to a more general graph structure. The nodes of the graph are the control volumes, each of which is a connected component of $\Omega \cap \Upsilon_i$. The arcs of the graph are faces connecting pairs of control volumes, at which fluxes are defined. This description can be implemented so that many of the advantages of the volume-of-fluid approach are still maintained: most of the data is stored, and most of the computations are performed, on rectangular arrays (Day, et al., 1998).

Multifluid methods. The early development of the volume-of-fluid ideas for computing fluid dynamics problems was done for the specific case of computing the dynamics of interfaces between different materials (Noh and Woodward, 1976; Hirt and Nichols, 1981). From the point of view of the discussion in this paper, such multifluid methods are hybrid methods, with some of the data being represented as double-valued along a sharp front (thermodynamic quantities such as density

and energy), with other quantities represented as having a single value per cell (velocity, pressure). These methods have been quite successful in representing complex material interface problems, and admit a mathematical description that is somewhat different than that described here (Miller and Puckett, 1996).

References

- M. Aftosmis, M. J. Berger and J. Melton, Robust and efficient Cartesian mesh generation for component-based geometry, *AIAA J.*, 36(6):952-960, June, 1998.
- A. S. Almgren, J. B. Bell, P. Colella and T. Marthaler, A Cartesian mesh method for the incompressible Euler equations in complex geometries, *SIAM J. Sci. Comput.*, 18(5):1289-1309, September, 1997.
- J. B. Bell, P. Colella and M. Welcome, A conservative front-tracking for inviscid compressible flow, Proceedings of the Tenth AIAA Computational Fluid Dynamics Conference, 814-822, June, 1991.
- M. J. Berger and R. J. Leveque, Stable boundary conditions for Cartesian grid calculations, ICASE Report 90-37, May, 1990.
- A. Bourlioux and A. J. Majda, Theoretical and numerical structure of unstable detonations, *Phil. Trans. Roy. Soc. London*, 350:29-68, January, 1995.
- D. Calhoun, A Cartesian grid method for solving the streamfunction-vorticity equations in irregular geometries, Ph. D. thesis, Department of Applied Mathematics, University of Washington, July, 1999.
- I.-L. Chern and P. Colella, A conservative front tracking method for hyperbolic conservation laws, Lawrence Livermore National Laboratory Report UCRL-97200, July, 1987.
- I.-L. Chern, J. Glimm, O. McBryan, B. Plohr, and S. Yaniv, Front tracking for gas dynamics, *J. Comput. Phys.*, 62:83-110, 1986.
- P. Colella, Multidimensional upwind methods for hyperbolic conservation laws, *J. Comput. Phys.*, 87:171-200, March, 1990.
- M. S. Day, P. Colella, M. Lijewski, C. Rendleman and D. L. Marcus Embedded boundary algorithms for solving Poisson's equation on complex domains Lawrence Berkeley National Laboratory report LBNL-41811, April, 1998.
- N. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*. Springer-Verlag, 1977.
- J. Helmsen, P. Colella and E.G. Puckett, Non-convex profile evolution in two dimensions using volume of fluids, Lawrence Berkeley National Laboratory report LBNL-40693, July, 1997.
- J. Hilditch and P. Colella, A front tracking method for compressible flames in one dimension, *SIAM J. Sci. Comput.*, 16:755-772, 1995.
- C. W. Hirt and B. D. Nichols, Volume-of-fluid (VOF) method for the dynamics of free boundaries, *J. Comput. Phys.*, 39:201-225, 1981.
- H. Johansen and P. Colella, A Cartesian grid embedded boundary method for Poisson's equation on irregular domains, *J. Comput. Phys.*, 147:60-85, December, 1998.
- H. S. Johansen, Cartesian grid embedded boundary methods for elliptic and parabolic partial differential equations on irregular domains, Ph. D. the-

- sis, Department of Mechanical Engineering, University of California, Berkeley, December, 1997.
- G. H. Miller and E. G. Puckett, A high-order Godunov method for multiple condensed phases, *J. Comput. Phys.*, 128(1):134-164, October, 1996.
- D. Modiano and P. Colella, A higher-order embedded boundary method for time-dependent simulation of hyperbolic conservation laws, Proceedings of the FEDSM 00 - ASME Fluids Engineering Simulation Meeting, June, 2000.
- W. F. Noh, CEL: A time-dependent, two-space-dimensional, coupled Eulerian-Lagrange code, *Methods in Computational Physics*, 3, 1963.
- W. F. Noh and P. R. Woodward, SLIC (simple line interface calculation), *Springer Lecture Notes in Physics*, 25:330-339, 1976.
- S. Osher and J. A. Sethian, Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations, *J. Comput. Phys.*, 79:12-49, 1988.
- R. B. Pember, A. S. Almgren, W. Y. Crutchfield, L. H. Howell, J. B. Bell, P. Colella and V. E. Beckner, An embedded boundary method for the modeling of unsteady Combustion in an industrial gas-fired furnace, Lawrence Livermore National Laboratory report UCRL-JC-122177, October, 1995.
- R. B. Pember, J. B. Bell, P. Colella, W. Y. Crutchfield and M. L. Welcome, An adaptive Cartesian grid method for unsteady compressible flow in irregular regions, *J. Comput. Phys.*, 120:278-304, 1995.
- J. E. Pilliod and E. G. Puckett, An unsplit, second-order accurate Godunov method for tracking deflagrations and detonations, *Proceedings of the 21st International Symposium on Shock Waves*, 1053-1058, July, 1997.
- E. G. Puckett, A. S. Almgren, J. B. Bell, D. L. Marcus and W. J. Rider, A higher-order projection method for tracking fluid interfaces in variable density incompressible flows, *J. Comput. Phys.*, 130(2):269-282, January, 1997.
- J. W. Purvis and J. E. Burkhalter, Prediction of critical Mach number for store configurations, *AIAA J.*, 17:1170-1177, 1979.
- E. H. Twizell, A. B. Gumel and M. A. Arigu, Second-order, L_0 -stable methods for the heat equation with time-dependent boundary conditions, *Adv. Comput. Math.*, 6(3):333-352, 1996.
- D. P. Young, R. G. Melvin, M. B. Bieterman, F. T. Johnson, S. S. Samani and J. E. Bussoletti, A locally refined rectangular grid finite element method: application to computational fluid dynamics and computational physics, *J. Comput. Phys.*, 92:1-66, January, 1991.

SOME NEW GODUNOV AND RELAXATION METHODS FOR TWO-PHASE FLOW PROBLEMS

F. COQUEL, E. GODEWSKI, B. PERTHAME

Laboratoire d'Analyse Numérique,

Université P. et M. Curie,

4 pl. Jussieu, 75252 Paris cedex 05, France.

Emails: coquel@ann.jussieu.fr, godlewski@ann.jussieu.fr

AND

A. IN, P. RASCLE

EDF/DER/RNE/PhR,

1 avenue du Général de Gaulle, 92141 Clamart Cedex, France.

Abstract. We consider a model for two-phase flows for which we construct an explicit finite volume numerical method based on an operator splitting. The extension to general pressure laws is done via an energy relaxation method. A new large-time step explicit scheme is also introduced.

1. Introduction

We consider an equal pressure two-fluid model for two-phase flows with phase change. In order to solve it numerically, stable and efficient methods are necessary which should ideally deal with a wide range of problems simulating off normal conditions for a Pressurized Water Reactor, and in acceptable computational time. We also refer to the works of (Ghidaglia et al., 1999), (Städtker and Holtbecker, 1994), (Toumi and Kumbaro, 1996) for instance for other approaches. Our approach introduces a splitting method which in a first step decouples the two phases and in a second step restores the equality of the pressures. In the first step we solve two classical Euler-type hyperbolic systems by a finite volume method, involving an approximate Riemann solver, such as a Godunov type scheme. The energy relaxation method proposed by (Coquel and Perthame, 1998), allows to extend to real fluids with general pressure laws and leads to “relaxed” Godunov-type solvers. Besides, in order to overcome the usual CFL time step limitation, we have tried a large time-step explicit numerical scheme,

involving another kind of relaxation procedure. We illustrate on some tests how this relaxed scheme can be used.

2. Two-fluid model and method of resolution

We consider a two-fluid model for two-phase flow, governed by the following system of six equations, which represents the balance equations for mass, momentum and energy for each phase k , $k = v, l$ (v for vapor, l for liquid), coupled by interaction terms

$$\left\{ \begin{array}{l} \partial_t \alpha_k \rho_k + \partial_x \alpha_k \rho_k u_k = \Gamma_k, \\ \partial_t \alpha_k \rho_k u_k + \partial_x \alpha_k (\rho_k u_k^2 + p) - p_i \partial_x \alpha_k = \\ \quad \alpha_k \rho_k g + M_{wk} + \Gamma_k u_i + M_{ik}^D, \\ \partial_t \alpha_k \rho_k (e_k + u_k^2/2) + \partial_x \alpha_k \rho_k (h_k + u_k^2/2) u_k + p \partial_t \alpha_k = \\ \quad (\alpha_k \rho_k g + M_{wk}) u_k + \Gamma_k (h_{ik} + u_i^2/2) + M_{ik}^D u_i + q_{ik} + q_{wk}. \end{array} \right. \quad (1)$$

We have denoted for phase k : α_k the volume fraction, ρ_k the density, u_k the velocity, e_k the specific internal energy, h_k the specific internal enthalpy, h_{ik} the interfacial specific internal enthalpy, Γ_k the interfacial mass transfer, q_{ik} the interfacial heat flux, M_{ik}^D the interfacial drag force per unit volume, M_{wk} the wall friction, q_{wk} the wall heat flux, also p is the common pressure for the two phases, p_i the common interfacial pressure, g the gravity, u_i the interface velocity. To this system we add the continuity condition

$$\alpha_v + \alpha_l = 1, \quad (2)$$

and the following interfacial transfer conditions

$$\sum_{k=v,l} \Gamma_k = 0, \quad \sum_{k=v,l} (\Gamma_k u_i + M_{ik}^D) = 0, \quad \sum_{k=v,l} (\Gamma_k (h_{ik} + \frac{u_i^2}{2}) + M_{ik}^D u_i + q_{ik}) = 0. \quad (3)$$

In order to close the problem, the equations of state of the two phases are given either assuming that the gases are ideal polytropic with a γ -law

$$p = (\gamma_k - 1) \rho_k e_k, \quad k = v, l, \quad (4)$$

or given in tabulated form. Another closure law must be given for p_i , which is a key-point for studying the hyperbolicity of the system (1). Several possibilities have been considered (for details see (In, 1999b)), in particular the convenient choice $p_i = p$ leads to a loss of hyperbolicity in the cases of interest, but we will bypass this problem numerically.

Let us now present the splitting method introduced by (Coquel et al., 1997) to solve approximately the systems (1), coupled by (2)(3). It is a two-step in time method, which relies on an operator decomposition. The idea is that during the first step two uncoupled hydrodynamical systems, one for each phase, are solved. Then during the second step coupling terms are taken into account, by restoring the equality of pressure between the two phases. The full discretization moreover involves a finite volume scheme and leads to a first order (in time and space) method.

Given a uniform grid of the x, t plane, with time step Δt and mesh size h , we note $x_j = jh$, $x_{j-1/2} = (j - 1/2)h$, $I_j = (x_{j-1/2}, x_{j+1/2})$, $t^n = n\Delta t$,

$$\mathbf{U}_j^n = \left((\alpha_v \rho_v)_j^n, (\alpha_v \rho_v u_v)_j^n, E_{v,j}^n, (\alpha_l \rho_l)_j^n, (\alpha_l \rho_l u_l)_j^n, E_{l,j}^n \right)^T, \quad (5)$$

where $E_k = \alpha_k \rho_k (e_k + u_k^2/2)$ and we associate to \mathbf{U}_j^n a piecewise constant fonction $\mathbf{U}^h(x, t)$ which is intended to be an approximation of the solution of (1), and (2) by

$$\mathbf{U}^h(x, t) = \mathbf{U}_j^n \text{ for } x \in I_j, t \in (t^n, t^{n+1}). \quad (6)$$

\mathbf{U}_j^n is computed from a given initial condition \mathbf{U}_j^0 (and boundary conditions which we do not describe) via the following steps.

Step 1. We solve the two following systems, one for each phase k

$$\begin{cases} \partial_t m_k + \partial_x m_k u_k = \Gamma_k, \\ \partial_t m_k u_k + \partial_x (m_k u_k^2 + \tilde{p}_k) = p_i \partial_x \alpha_k + m_k g + M_{wk} + \Gamma_k u_i + M_{ik}^D, \\ \partial_t E_k + \partial_x (E_k + \tilde{p}_k) u_k = \\ \quad (m_k g + M_{wk}) u_k + \Gamma_k (h_{ik} + u_i^2/2) + M_{ik}^D u_i + q_{ik} + q_{wk}, \end{cases} \quad (7)$$

with $m_k = \alpha_k \rho_k$, and $\tilde{p}_k = \alpha_k p$. Let us denote

$$\begin{aligned} \mathbf{U}_k &= (m_k, m_k u_k, E_k)^T, \\ \mathbf{F}_k(\mathbf{U}_k) &= (m_k u_k, m_k u_k^2 + \tilde{p}_k, E_k)^T, \\ \mathbf{h}_k &= \begin{pmatrix} \Gamma_k \\ p_i \partial_x \alpha_k + m_k g + M_{wk} + \Gamma_k u_i + M_{ik}^D \\ (m_k g + M_{wk}) u_k + \Gamma_k (h_{ik} + u_i^2/2) + M_{ik}^D u_i + q_{ik} + q_{wk} \end{pmatrix}. \end{aligned}$$

With these notations, (7) writes

$$\partial_t \mathbf{U}_k + \partial_x \mathbf{F}_k(\mathbf{U}_k) = \mathbf{h}_k(\mathbf{U}_v, \mathbf{U}_l), \quad k = v, l. \quad (8)$$

Let us denote $\mathbf{U}_{k,j}^n = \mathbf{U}_k(x, t)$ for $(x, t) \in I_j \times (t^n, t^{n+1})$, and $\mathbf{h}_{k,j}^n = \mathbf{h}_k(\mathbf{U}_{v,j}^n, \mathbf{U}_{l,j}^n)$. We solve approximately these two systems (8) with an explicit first order conservative method under the form

$$\mathbf{U}_{k,j}^{n+1/2} - \mathbf{U}_{k,j}^n + \frac{\Delta t}{h} (\mathbf{F}_{k,j+1/2}^n - \mathbf{F}_{k,j-1/2}^n) = \Delta \mathbf{h}_{k,j}^n, \text{ with} \quad (9)$$

$$\mathbf{F}_{k,j+1/2}^n = \mathbf{F}_k(\mathbf{U}_{k,j}^n, \mathbf{U}_{k,j+1}^n), \text{ for } k = v, l. \quad (10)$$

As we have already pointed out, the resolution of the two systems is uncoupled and reduces to that of two Euler systems for which many solvers have been developed, at least in the case of polytropic ideal gases. In the numerical experiments, we have used for the numerical flux $\mathbf{F}_k(.,.)$ either that of a Godunov type scheme : Godunov (Godunov, 1979), Roe (Roe, 1981), or VFFC (Ghidaglia et al., 1999) or a kinetic numerical flux (Perthame, 1992). The comparison of the Riemann solvers has been done for two numerical tests (separation and water faucet problem) for which we refer to In (Coquel et al., 1998a).

We denote $\alpha = \alpha_v$ thereafter. The discretization of the term $(p_i \partial_x \alpha)_j^n$ must be handled with care since it influences the stability of the scheme in case one phase vanishes. It can be discretized in a centered way or else $p_i \partial_x \alpha$ is rewritten as

$$p_i \partial_x \alpha = p_i \alpha (1 - \alpha) \partial_x \ln\left(\frac{\alpha}{1 - \alpha}\right), \quad (11)$$

then, denoting $\beta = \ln\left(\frac{\alpha}{1 - \alpha}\right)$, it is discretized under the form

$$(p_i \partial_x \alpha)_j^n = (p_i \alpha (1 - \alpha))_j^n \frac{1}{h} \text{minmod}(\beta_{j+1}^n - \beta_j^n, \beta_j^n - \beta_{j-1}^n). \quad (12)$$

Let us remark that (12) also helps to bypass the numerical instabilities due to the non hyperbolicity of the system (1)(since we have assumed $p_i = p$). Other choices are possible, and we refer the reader to (Coquel et al., 1997).

Step 2. The equality of pressure is restored during the second step, in which we solve the two following systems

$$\begin{cases} d_t m_k = 0, \\ d_t m_k u_k = 0, \\ d_t E_k = -p d_t \alpha_k \text{ for } k = v, l, \end{cases} \quad (13)$$

coupled by the equation (2). Since this is a system of ODEs, this problem can be solved locally in space, so we drop the spatial index j in what follows.

The first two equations in (13) lead for $k = v, l$, to

$$m_k^{n+1} = m_k^{n+1/2}, \quad m_k^{n+1} u_k^{n+1} = m_k^{n+1/2} u_k^{n+1/2}. \quad (14)$$

Then we discretize the last equations of (13) under the form

$$E_k^{n+1} - E_k^{n+1/2} = -p^{n+1} (\alpha_k^{n+1} - \alpha_k^n). \quad (15)$$

Considering the two equations of state (4), the continuity equation (2), and the equality of pressures, an easy computation leads to the expressions for

α^{n+1} and p^{n+1} . These values enable us to update the total energies and to compute all the non conservative variables.

3. Relaxed Godunov-type two-phase flow solver

When one solves problems involving “real fluids” the pressure laws are no longer known as explicit functions of the thermodynamic variables but given by tables and the frequent calls to the tables required by the classical methods are highly time-consuming (In, 1999b).

We thus associate to the first step of our splitting method the energy relaxation method (Coquel and Perthame, 1998) that allows to extend to general pressure laws the classical schemes developed for polytropic ideal gas dynamics (i.e. it is applied to each of the decoupled Euler-type systems (7) solved in the first step; details can be found in (In, 1999a)). Such a method uses the exact pressure law only once per time step and mesh cell.

Let us just give some hints of the relaxation procedure for a typical Euler system (for a full study, we refer to (Coquel and Perthame, 1998))

$$\begin{cases} \partial_t \rho + \partial_x(\rho u) = 0, \\ \partial_t \rho u + \partial_x(\rho u^2 + p) = 0, \\ \partial_t E + \partial_x(E + p)u = 0, \end{cases} \quad (16)$$

where $E = \frac{1}{2}\rho u^2 + \rho e$. We assume that this system is supplemented by an incomplete equation of state (EOS) $p = p(\rho, e)$. The main idea is to introduce a decomposition of the internal energy under the form $e = e_1 + e_2$ where e_1 governs a simple polytropic pressure law $p_1(\rho, e_1) = (\gamma_1 - 1)\rho e_1$ with γ_1 a given constant greater than 1, while e_2 includes the non-linearities and will be simply advected by the flow. Let us just say that it is possible to exhibit an entropy inequality, compatible with the relaxation procedure if γ_1 is chosen so that the subcharacteristic conditions holds which reads

$$\gamma_1 > \sup C^2 \rho / p \quad (17)$$

where C^2 denotes the equilibrium sound speed.

Now, we apply this energy relaxation theory to build finite volume schemes for general pressure laws. We define a piecewise constant approximation $\mathbf{U}^h(x, t)$ of the solution of (16), under the form

$$\mathbf{U}^h(x, t) = \mathbf{U}_j^n = (\rho_j^n, \rho_j^n u_j^n, E_j^n), \quad x \in I_j, \quad t \in (t^n, t^{n+1}). \quad (18)$$

We use again an operator splitting technique in order to advance in time from the approximation $\mathbf{U}^h(x, t^n)$ to the approximation $\mathbf{U}^h(x, t^{n+1})$.

First step (relaxation): The states $\mathbf{U}_{1,j}^n = (\rho_j^n, \rho_j^n u_j^n, \frac{1}{2} \rho u_j^n {}^2 + \rho_j^n e_{1,j}^n)^T$ and the internal energy $\rho_j^n e_{2,j}^n$ are obtained from the relaxation procedure by the consistency relations

$$\begin{aligned} (\gamma_1^n - 1) \rho_j^n e_{1,j}^n &= p(\rho_j^n, e_j^n), \\ \rho_j^n e_{2,j}^n &= \rho_j^n e_j^n - \rho_j^n e_{1,j}^n, \end{aligned} \quad (19)$$

where γ_1^n is defined so that (17) holds.

Second step (evolution): We solve approximately the Cauchy problem

$$\partial_t \mathbf{U}_1 + \partial_x \mathbf{F}_1(\mathbf{U}_1) = \mathbf{0}, \quad (20)$$

$$\partial_t \rho e_2 + \partial_x \rho e_2 u = 0, \quad (21)$$

for $t^n \leq t < t^{n+1}$, with the initial data $(\mathbf{U}_{1,j}^n, \rho_j^n e_{2,j}^n)^T$ for $x \in I_j$. Note that in this step, the conservation laws (20) governing \mathbf{U}_1 are completely decoupled from (21) that governs e_2 and we can use for (20) any first order conservative finite volume scheme. In practice, we have tested the Godunov and the Roe scheme (also a kinetic scheme) to get $\mathbf{U}_{1,j}^{n+1-}$ for all $j \in Z$. Denoting $\bar{\mathbf{F}}_1$ the numerical flux function for the system (20) (respectively $\bar{\mathbf{F}}_2$ for the equation (21)), we get at time t^{n+1-}

$$\mathbf{U}_{1,j}^{n+1-} = \mathbf{U}_{1,j}^n - \frac{\Delta t}{h} (\bar{\mathbf{F}}_1(\mathbf{U}_{1,j}^n, \mathbf{U}_{1,j+1}^n) - \bar{\mathbf{F}}_1(\mathbf{U}_{1,j-1}^n, \mathbf{U}_{1,j}^n)), \quad (22)$$

$$(\rho e_2)_j^{n+1-} = (\rho e_2)_j^n - \frac{\Delta t}{h} (\bar{\mathbf{F}}_2(\mathbf{U}_j^n, \mathbf{U}_{j+1}^n) - \bar{\mathbf{F}}_2(\mathbf{U}_{j-1}^n, \mathbf{U}_j^n)) \quad (23)$$

(where we have defined $e_{2,j}^{n+1-}$ in a way which ensures positivity).

At last, we compute \mathbf{U}_j^{n+1} , thanks to the following relations which are implied by the relaxation procedure

$$\begin{cases} \rho_j^{n+1} = \rho_j^{n+1-}, \\ (\rho u)_j^{n+1} = (\rho u)_j^{n+1-}, \\ E_j^{n+1} = E_{1,j}^{n+1-} + (\rho e_2)_j^{n+1-}. \end{cases} \quad (24)$$

In this way, we obtain a conservative finite volume scheme under the usual form

$$\mathbf{U}_j^{n+1} = \mathbf{U}_j^n - \frac{\Delta t}{h} (\bar{\mathbf{F}}(\mathbf{U}_j^n, \mathbf{U}_{j+1}^n) - \bar{\mathbf{F}}(\mathbf{U}_{j-1}^n, \mathbf{U}_j^n)), \quad (25)$$

provided we define the numerical flux

$$\bar{\mathbf{F}}(\mathbf{U}_j^n, \mathbf{U}_{j+1}^n) = \left(\begin{array}{c} \bar{F}_{1,1}(\mathbf{U}_{1,j}^n, \mathbf{U}_{1,j+1}^n) \\ \bar{F}_{1,2}(\mathbf{U}_{1,j}^n, \mathbf{U}_{1,j+1}^n) \\ \bar{F}_{1,3}(\mathbf{U}_{1,j}^n, \mathbf{U}_{1,j+1}^n) + \bar{F}_2(\mathbf{U}_j^n, \mathbf{U}_{j+1}^n) \end{array} \right) \quad (26)$$

where we denote $\bar{\mathbf{F}}_1 = (\bar{F}_{1,1}, \bar{F}_{1,2}, \bar{F}_{1,3})^T$.

The extensions by the above procedure of the usual solvers used for the resolution of (20) are straightforward and preserve all the desirable properties of stability and accuracy of the original method, for a correct choice of \bar{F}_2 (for instance, see Theorem 4.7 in (Coquel and Perthame, 1998) for a precise statement concerning the relaxed Godunov method).

As already said, we can inject any of these “relaxed Euler solvers” in the first step of our splitting method for multiphase flow and we get “relaxed two-phase flow solvers”. Numerical experiments for two-phase flows can be found in (Coquel et al., 1998b).

4. Large time-step method

The stability of the above split scheme depends on the greatest sound velocity of the two phases and not on the mixture sound velocity. This can be very restrictive on the time step and we now present a large time-step method which relies on a different relaxation procedure performed on the system in Lagrangian coordinates. The Lagrangian part is then followed by a projection back on the Eulerian grid (we refer to (Godlewski and Raviart, 1996)). Again this block will be incorporated in the first step of our split method and so we present it on a classical Euler system written in Lagrangian coordinates.

Consider the Euler system (16) written in Lagrangian coordinates

$$\begin{cases} \partial_t \tau - \partial_m u = 0, \\ \partial_t u + \partial_m p = 0, \\ \partial_t e + \partial_m p u = 0, \end{cases} \quad (27)$$

where $\tau = \rho^{-1}$, with the entropy inequality

$$\partial_t s \leq 0 \quad (28)$$

(where we have improperly noted in the same way the functions in Lagrangian coordinates m, t and in the laboratory frame x, t). It is well known that smooth solutions satisfy also $\partial_t s = 0$ and the following equation

$$\partial_t p + c^2 \partial_m u = 0, \quad (29)$$

where c defined by $\partial_\tau p(\tau, s) = -c^2$ is the Lagrangian sound speed (see (Godlewski and Raviart, 1996) for instance). The idea is then to solve in the first step a linear isentropic system with $p = p(\tau, s_0)$

$$\begin{cases} \partial_t \tau - \partial_m u = 0, \\ \partial_t u + \partial_m \pi = 0, \\ \partial_t \pi + a^2 \partial_m u = 0, \end{cases} \quad (30)$$

with a greater than the exact sound speed c (this condition happens to be necessary in the theoretical analysis of the relaxation procedure) and in the second step, the pressure is “relaxed” (see (In, 1999b)). Note that the eigenvalues of the linear system (30) are $-a, 0, a$ with the characteristic variables $W_1 = \pi - au, W_2 = \pi + a^2\tau, W_3 = \pi + au$ so that system (30) is equivalent to the decoupled system

$$\begin{cases} \partial_t W_1 - a\partial_m W_1 = 0, \\ \partial_t W_2 = 0, \\ \partial_t W_3 + a\partial_m W_3 = 0. \end{cases} \quad (31)$$

Note that

$$u = (W_3 - W_1)/2a, \pi = (W_1 + W_3)/2, \pi u = (W_3^2 - W_1^2)/4a. \quad (32)$$

This will be used below to compute the numerical fluxes for system (30).

Indeed, we first define the Lagrangian grid by $m_{1/2} = 0$, and at time t^n $m_{j+1/2} = \sum_{k=1}^j \rho_k^n \Delta x$, then we set $\Delta m_j = \rho_j^n \Delta x$, $M_j = (m_{j-1/2}, m_{j+1/2})$. Given an initial condition $\mathbf{V}^n(m)$ at time t^n , we want to approximate the conservative variable $\mathbf{V} = (\tau, u, e)^T$ at t^{n+1} , by a piecewise constant defined for $m \in M_j$ by $\mathbf{V}^h(m, t^{n+1}) = \mathbf{V}_j^{n+1} = (\tau_j^{n+1}, u_j^{n+1}, e_j^{n+1})^T$.

When it comes down to it, the conservative numerical scheme will be written in the usual way, but the derivation needs some steps. The solution τ, u of (30) with the initial condition at t^n for $m \in M_j$

$$(\tau, u, s, \pi)(m, t^n) = (\tau_j^n, u_j^n, s_j^n, p_j^n), \quad (33)$$

is given by

$$\begin{aligned} \tau_j^{n+1} &= \tau_j^n - \frac{\Delta t}{\Delta m_j} (F_{\tau, j+1/2} - F_{\tau, j-1/2}), \\ u_j^{n+1} &= u_j^n - \frac{\Delta t}{\Delta m_j} (F_{u, j+1/2} - F_{u, j-1/2}), \end{aligned} \quad (34)$$

and the numerical fluxes are computed as now described.

One chooses a so that

$$a > \max_j \left(\sqrt{-\frac{\partial p}{\partial \tau}} \right)_j^n, \quad (35)$$

and we set $\pi(m, t^n) = p_j^n$ for $m \in M_j$, and

$$W_{1,j}^n = W_1(m, t^n) = p_j^n - au_j^n, \quad W_{3,j}^n = W_3(m, t^n) = p_j^n + au_j^n. \quad (36)$$

Then from (32)

$$\begin{aligned} F_{\tau,j+1/2} &= \frac{1}{2a\Delta t} \int_{t^n}^{t^{n+1}} (W_1 - W_3)(m_{j+1/2}, t) dt = \frac{1}{2a} (\bar{W}_1 - \bar{W}_3)_{j+1/2}, \\ F_{u,j+1/2} &= \frac{1}{2\Delta t} \int_{t^n}^{t^{n+1}} (W_1 + W_3)(m_{j+1/2}, t) dt = \frac{1}{2} (\bar{W}_1 + \bar{W}_3)_{j+1/2}. \end{aligned} \quad (37)$$

The exact computation of the fluxes can be carried out for a CFL much greater than one by adding the contributions of the cells concerned. Let us recall that the relaxation system is linear. We skip the details which can be found in (In, 1999b). We then denote

$$F_{e,j+1/2} = \frac{1}{4a\Delta t} \int_{t^n}^{t^{n+1}} (W_3^2 - W_1^2)(m_{j+1/2}, t) dt \quad (38)$$

and set

$$e_j^{n+1} = e_j^n - \frac{\Delta t}{\Delta m_j} (F_{e,j+1/2} - F_{e,j-1/2}), \quad (39)$$

then

$$\pi_j^{n+1} = p(\tau_j^{n+1}, e_j^{n+1}). \quad (40)$$

Formulas (34) and (39) give a scheme for the Euler system in Lagrangian coordinates. Again, this relaxation procedure can be justified when analyzing the entropy inequalities and it can be shown provided (35) holds that this scheme satisfies a discrete entropy inequality. This Lagrangian step is then followed by a projection back onto the Eulerian grid. The “Lagrange + projection” block can then be incorporated in the first step our splitting method.

We illustrate the saving of CPU time on a simple example (Sod shock tube problem, see Figure 1). The computations use the CFL 5, 10 et 50 for 1000 meshes, for Godunov with a CFL 0.9, $t_{final} = 0.05$ s. One gets the following cpu for $dx = 10^{-3}$ m: Godunov CFL= 0,9 cpu=57, and for the large time-step scheme for CFL= 5, cpu=9, for CFL= 10, cpu=5, CFL= 50, cpu=2.

References

- Coquel F and Perthame B (1998). Relaxation of energy and approximate Riemann solvers for general pressure laws in fluid dynamics. *SIAM Journal on Numerical Analysis* **35**, pp 2223–2249.
- Coquel F El Amine K Godlewski E Perthame B and Rascle P (1997). A Numerical Method Using Upwind schemes for the Resolution of Two-Phase Flows. *J. Comp. Phys.* **136**, pp. 272-288.
- Coquel F Godlewski E In A Perthame B and Rascle P (1998a). Influence of the Riemann Solver over the Splitting Method for the Resolution of Two-phase Flows. Proc. Sixteenth international conference on Numerical methods in fluid dynamics Lecture Notes in Physics **515**, Springer-Verlag.

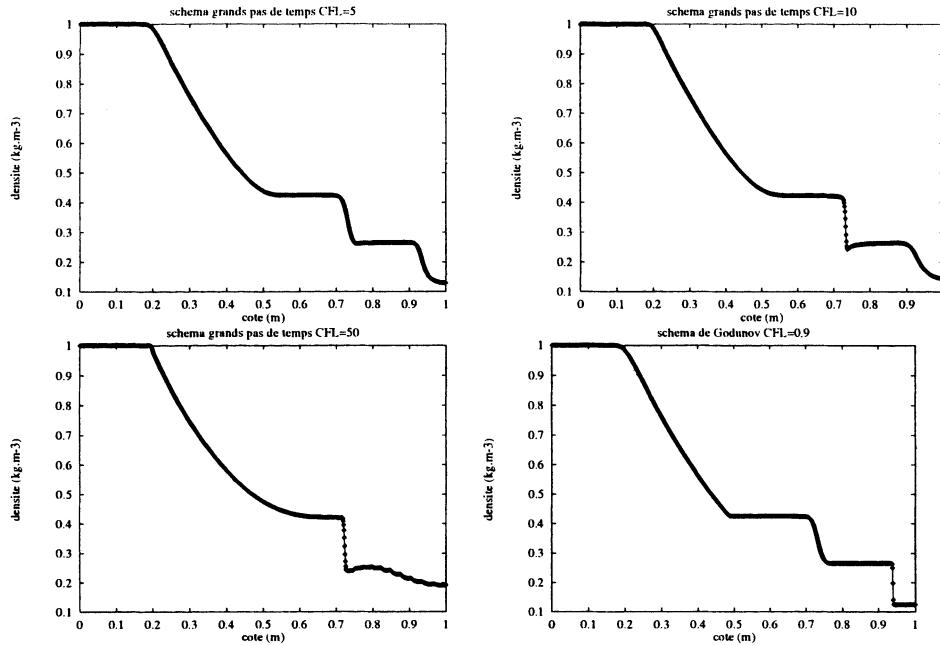


Figure 1. Sod tube, 1000 meshes : density.

- Coquel F Godlewski A Perthame E In B and Rascle P (1998b). An Energy Relaxation method for Inviscid Real Fluids and Application to Two-phase Flows. Proc. ICFD Conference on Numerical Methods for Fluid Dynamics, Numerical Methods for Fluid Dynamics VI, M.J. Baines Ed. ICFD, Oxford.
- Ghidaglia J-M Lecoq G and Toumi I (1999). Two Flux Schemes for computing Two-Phase Flows through Multidimensional Finite Element Method. Preprint CMLA, Ens Cachan, France.
- Godlewski E and Raviart P A (1996). Numerical Approximation of Hyperbolic Systems of Conservation Laws. Applied Mathematical Sciences 118, Springer, New York.
- Godunov S et coll. (1979). Résolution Numérique des Problèmes Multidimensionnels de la Dynamique des Gaz. Translated from Russian, Editions Mir Moscou.
- In A (1999a) Numerical Evaluation of an Energy Relaxation Method for Inviscid Real Fluids. *SIAM J. Sc. Comp.* **21**, pp. 340-365.
- In A (1999b) Méthodes Numériques pour les équations de la Dynamique des Gaz complexes et Écoulements Diphasiques. Doctoral dissertation, Université Paris 6, Paris.
- Perthame B (1992) Second-order Boltzmann schemes for compressible Euler equations in one and two space dimensions. *SIAM J. Numer. Anal.* **29**, pp. 1-19.
- Roe P L (1981) Approximate Riemann Solvers, Parameter Vectors and Difference Schemes. *J. Comput. Phys.* **43**, pp. 357-372.
- Städtk H and Holtbecker R (1994) Numerical Simulation of Multi-Dimensional Two-Phase Flow Based on Hyperbolic Flow Equations. submitted to the 30th Meeting of the European Two-phase Flow Group, Piacenza, 6-8 June 1994, Italie.
- Toumi I and Kumbaro A (1996) An approximate linearized solver for a two-fluid model. *J. Comput. Phys.* **124**, pp 286-300.

DEVELOPMENT OF GENUINELY MULTIDIMENSIONAL UPWIND RESIDUAL DISTRIBUTION SCHEMES FOR THE SYSTEM OF EIGHT WAVE IDEAL MAGNETOHYDRO-DYNAMIC EQUATIONS ON UNSTRUCTURED GRIDS

ÁRPÁD CSÍK

*Centre for Plasma-Astrophysics, K.U.Leuven,
Celestijnenlaan 200B, 3001 Leuven, Belgium.
and*

*Aeronautics and Aerospace Department,
Von Karman Institute for Fluid Dynamics,
72, Chaussée de Waterloo 1640, Rhode-Saint-Genèse, Belgium.
Email: arpi@vki.ac.be*

HERMAN DECONINCK

*Aeronautics and Aerospace Department,
Von Karman Institute for Fluid Dynamics,
72, Chaussée de Waterloo 1640, Rhode-Saint-Genèse, Belgium.
Email: deconinck@vki.ac.be*

AND

STEFAAN POEDTS

*Centre for Plasma-Astrophysics, K.U.Leuven,
Celestijnenlaan 200B, 3001 Leuven, Belgium.
Research Associate of the Flemish Fund for Scientific Research
(FWO-Vlaanderen)
Email: Stefaan.Poedts@wis.kuleuven.ac.be*

Abstract. Multidimensional upwind residual distribution schemes are applied to the eight-wave equations of ideal magnetohydrodynamics. Both first and second order linear, and second order nonlinear distribution schemes are presented. The solenoidal condition of the magnetic field is enforced by Powell's source term approach. Time integration is done by using both explicit and implicit strategies. The spatial accuracy and the shock capturing properties of the schemes in the steady state are investigated numerically. Computational results are shown for a *super-magnetosonic* channel flow.

1. Introduction

The magnetohydrodynamics (MHD) equations describe the behaviour of conducting plasmas embedded in magnetic fields. They combine the equations of fluid dynamics with Maxwell's equations of electrodynamics. Many important and interesting macroscopic phenomena in both laboratory and astrophysical plasmas can be successfully described by the MHD model.

A higher order Godunov method for the solution of the ideal MHD equations has been developed by (Zachary A L, Colella P, 1992). The main difficulties in the development of higher order Godunov schemes are the non-strict hyperbolicity of the MHD equations and the fact that the genuinely nonlinear waves can be locally linearly degenerate. Unlike the Euler equations, the ideal MHD equations are nonconvex (Brio M and Wu C C, 1988), consequently the wave structure becomes more complicated. Compound waves may appear for example, consisting of a shock and an attached rarefaction wave.

In the past fifteen years there have been many developments in the field of the solution of the ideal MHD equations (Brio M and Wu C C, 1988; Zachary A L, Colella P, 1992; Dai W and Woodward P R, 1994; Powell K G, Roe P L, Myong R S, Gombosi T and De Zeeuw D L, 1995) yielding a wide variety of robust and relatively accurate numerical schemes, which have been successfully applied in space physics applications. However, an important drawback of these methods is that the multidimensional problem is usually split into 1D Riemann-problems, thereby loosing important information coming from the multidimensional nature of the equations. This increases the numerical diffusivity and decreases the accuracy of the scheme.

In the past few years an attractive genuinely multidimensional upwind method has emerged in computational fluid dynamics, as an alternative to these classical finite volume schemes. The idea of the residual distribution (fluctuation splitting) method as a multidimensional upwind spatial discretisation technique for the Euler equations of gas dynamics has originated by Roe (Roe P L, 1982) and was further developed by Deconinck, Roe, Struijs, and Paillère (Deconinck H, Struijs R and Roe P L, 1990; Paillère H, Deconinck H and Roe P L, 1995). The multidimensional upwind schemes for scalar advection (Paillère H, Deconinck H and Roe P L, 1995) were extended to non-commuting hyperbolic systems by van der Weide (van der Weide E, 1998; van der Weide E, Deconinck H, Issmann E and Degrez G, 1999), who introduced positive matrix distribution schemes.

In this paper we give an overview of the development of monotone upwind multidimensional numerical schemes for the ideal MHD equations, based on these latest results. A detailed analysis of the problem can be

found in a recent work of the authors of this paper (Csík Á, Deconinck H and Poedts S, 1999). The fluctuation splitting method requires regular Jacobian matrices per cell. Since the original form of the MHD equations leads to singular Jacobian matrices, we solve the regularized eight-wave Galilean invariant ideal MHD equations as given in (Godunov S K, 1972). In this approach, which was first proposed by Powell (Powell K G, Roe P L, Myong R S, Gombosi T and De Zeeuw D L, 1995), the singularity is removed by adding a source term proportional to $\nabla \cdot \mathbf{B}$ to the conservative form, hence introducing the so-called divergence wave. Consequently, the resulting equations are no longer in conservation form. This weak enforcement of the so-called divergence free condition stabilizes the numerical scheme against the instabilities related to the numerically non vanishing divergence of the magnetic field.

2. The governing equations of ideal MHD

The hyperbolic eight wave system of ideal MHD equations is given by:

$$\frac{\partial U}{\partial t} + \nabla \cdot \mathbf{F} = S, \quad (1)$$

where the state vector $U = (\rho, \rho u, \rho v, \rho w, B_x, B_y, B_z, E)^T$ contains the conservative variables and $\mathbf{F} = (F_x, F_y, F_z)$ is the conservative flux vector:

$$\mathbf{F} = \begin{bmatrix} \rho \mathbf{v} \\ \rho \mathbf{v} \mathbf{v} + \hat{I}(p + \frac{\mathbf{B} \cdot \mathbf{B}}{2}) - \mathbf{B} \mathbf{B} \\ \mathbf{v} \mathbf{B} - \mathbf{B} \mathbf{v} \\ (E + p + \frac{\mathbf{B} \cdot \mathbf{B}}{2}) \mathbf{v} - \mathbf{B}(\mathbf{v} \cdot \mathbf{B}) \end{bmatrix}.$$

The source term S is proportional to the divergence of the magnetic field:

$$S = -s \nabla \cdot \mathbf{B},$$

with $s = [0, B_x, B_y, B_z, u, v, w, \mathbf{v} \cdot \mathbf{B}]^T$. In the above equations we use the standard notations: \hat{I} represents the 3×3 identity matrix, ρ is the density, u, v , and w are the x, y , and z components of the velocity \mathbf{v} , respectively, \mathbf{B} is the magnetic field, p is the thermal pressure, and E is the total energy density defined by

$$E = \frac{p}{\gamma - 1} + \frac{1}{2} \rho \mathbf{v} \cdot \mathbf{v} + \frac{1}{2} \mathbf{B} \cdot \mathbf{B},$$

where γ is the ratio of specific heats. Equation (1) has to be supplemented by the divergence free condition of the magnetic field:

$$\nabla \cdot \mathbf{B} = 0, \quad (2)$$

expressing the observational fact that magnetic monopoles do not exist. In an analytical treatment, equation (2) is an initial constraint and the evolution equation of the magnetic field guarantees its validity at all times if it is initially satisfied. Consequently, equation (1) becomes a set of conservation laws describing the conservation of mass, momentum, magnetic flux and energy. However, in a numerical approach, due to the discretisation and the round off errors, the divergence of the magnetic field may deviate from zero, which has a destabilizing effect on the numerical algorithm. As was shown by Powell (Powell K G, Roe P L, Myong R S, Gombosi T and De Zeeuw D L, 1995), the source term in equation (1) stabilizes the MHD equations against this numerical instability by avoiding the accumulation of these numerical errors.

The source term S can be incorporated into the singular Jacobian matrices $A'_U = \frac{\partial F_x}{\partial U}$, $B'_U = \frac{\partial F_y}{\partial U}$ and $C'_U = \frac{\partial F_z}{\partial U}$ resulting in regular coefficient matrices A_U , B_U and C_U in the quasi-linear form of the governing equations (1):

$$\frac{\partial U}{\partial t} + A_U \frac{\partial U}{\partial x} + B_U \frac{\partial U}{\partial y} + C_U \frac{\partial U}{\partial z} = 0. \quad (3)$$

3. Fluctuation Splitting Spatial Discretisation

We solve the eight wave system of ideal MHD equations (1) on an arbitrary triangulation of the 2D spatial domain Ω , under the assumption of invariance on the coordinate z ($\frac{\partial}{\partial z} \equiv 0$), based on the residual distribution method (van der Weide E, Deconinck H, Issmann E and Degrez G, 1999). Just as in linear finite element methods, the solution is approximated by a continuous function, varying linearly over each triangle,

$$U(x, y, t) = \sum_i U_i(t) N_i(x, y),$$

where $U_i(t)$ is the time dependent value $U(x_i, y_i, t)$ at node i and $N_i(x, y)$ is the piecewise linear shape function equal to unity at node i and vanishing outside the support of all triangles meeting at node i .

Assume that equation (1) has been linearized over a triangular cell T , such that it is equivalent to equation (3) at the discrete level. Integration of equation (3) then leads to the definition of the fluctuation Φ^T in triangle T :

$$\Phi^T = \int_T \left(A_U \frac{\partial U}{\partial x} + B_U \frac{\partial U}{\partial y} \right) d\Omega = \sum_{i=1}^3 K_i U_i,$$

where $K_i = (A_U n_{x,i} + B_U n_{y,i})/2$ is the linear combination of matrices A_U and B_U , which are constant for each triangle. $n_{x,i}$, and $n_{y,i}$ are the components of the inner normal vector scaled with the length of the edge. Diagonalization of matrix K_i yields:

$$K_i = \frac{1}{2} R_i \Lambda_i L_i,$$

where the columns of R_i contain the scaled right eigenvectors, the rows of L_i contain the scaled left eigenvectors and Λ_i is the diagonal matrix of the eigenvalues of matrix K_i . The matrices $K_i^+ = \frac{1}{2} R_i \Lambda_i^+ L_i$ and $K_i^- = \frac{1}{2} R_i \Lambda_i^- L_i$ are the so called generalized upwind parameters, where Λ_i^+ contains the positive and Λ_i^- contains the negative eigenvalues: $\Lambda_i^\pm = (\Lambda_i \pm |\Lambda_i|)/2$.

In the fluctuation splitting approach we compute the cell residual Φ^T and we distribute its fractions to the nodes of triangle T according to the eigenvalues of matrix K_i . We assemble the contributions from all cells to the nodes and we obtain the nodal update. This treatment leads to a very compact stencil, which is useful for efficient coding of parallel and implicit algorithms. The semi-discretised form of equation (3) at point i with the lumped Galerkin mass matrix is

$$\frac{dU_i}{dt} = -\frac{1}{S_i} \sum_T \beta_i^T \Phi^T = -\frac{1}{S_i} \sum_T \Phi_i^T.$$

Here, S_i is the area of the median dual cell around node i equal to one third of the area of triangles sharing node i , β_i^T is the distribution matrix to node i in triangle T , and $\Phi_i^T = \beta_i^T \Phi^T$ is the distribution function in triangle T to node i . In order to keep the consistency of the scheme, we require that $\beta_1^T + \beta_2^T + \beta_3^T = \hat{I}$.

The properties of the different schemes in the fluctuation splitting context are determined by the way the β_i^T matrix (or the distribution function $\beta_i^T \Phi^T$) is defined (see Table 1). A monotonic scheme can be obtained by demanding positivity. The solution of a positive scheme is free of spurious oscillations close to discontinuities. A scheme is called linearity preserving, if it gives the exact solution of a problem in a steady state, when this is a linear function of the space variables. In smooth flows linearity preserving schemes in practice yield second order accuracy at the steady state. A more

detailed description of these properties is given by van der Weide (van der Weide E, 1998; van der Weide E, Deconinck H, Issmann E and Degrez G, 1999).

4. Numerical Results and Discussion

In order to integrate equation (1) in time, we use either multi-stage Runge Kutta explicit or Backward Euler implicit time marching procedures. The implicit solution strategies are the extended version of the ones developed by Issman (Issman E, Degrez G and Deconinck H, 1996) for solving compressible flow equations. For fully *super-magnetosonic* test problems the implicit schemes required 2-3 times less CPU time than the explicit schemes.

scheme	type	U	P	LP	Φ_i^T
N	linear	yes	yes	no	$K_i^+(U_i - U_{in})$
LDA	linear	yes	no	yes	$K_i^+(\sum_{j=1}^3 K_j^+)^{-1} \Phi^T$
B	nonlinear	yes	yes	yes	$\theta \Phi_i^N + (\hat{I} - \theta) \Phi_i^{LDA}$

Table 1: Properties of three fluctuation splitting schemes. From left to right: name of the scheme, type, upwind (U), positivity (P), linearity preservation (LP) and the definition of the distribution function Φ_i^T . In the last row of the table, θ is a nonlinear diagonal blending matrix. The letters N, LDA, and B stand for the Narrow, Linearity Preserving A, and Blending schemes, respectively.

One definite advantage of our approach compared to standard finite volume schemes is the compactness of the stencil. The stencil only includes the immediate surrounding nodes of a certain node, even for second order schemes. This enables a very efficient coding, in particular for parallel implicit schemes. Another advantage compared to dimensionally split schemes is that the presented fluctuation splitting schemes use true multidimensional information. This results in a much sharper shock capturing property and higher accuracy.

In order to demonstrate this property, we perform the magnetic nozzle test case described as follows. The *super-magnetosonic* flow enters the nozzle from the left (see figures) and leaves it at the right with a *super-magnetosonic* speed. At the bottom and at the top of the nozzle we impose symmetric and ideal perfectly conducting wall boundary conditions, respectively. In the initially imposed uniform flow field the density $\rho = 1$, the velocity vector $\mathbf{v} = (3, 0, 0)$, the magnetic vector $\mathbf{B} = (2, 0, 0)$, and the thermal pressure of the plasma $p = 0.6$.

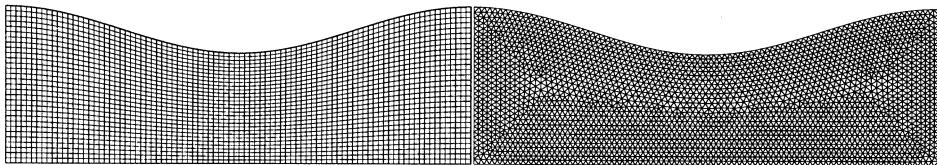


Figure 1. Structured quadrilaterals for the Roe type finite volume code with 2700 finite volumes (left) and unstructured triangulation for the fluctuation splitting code with 2662 nodes (right)

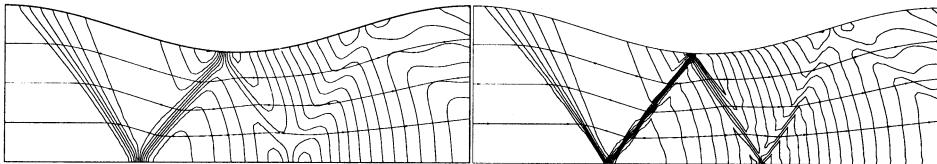


Figure 2. Density isolines superimposed by magnetic field lines in the magnetic nozzle test case for the second order Roe type finite volume scheme (left) and for the fluctuation splitting second order B scheme (right)

The results for the second order nonlinear residual distribution B scheme and for a standard, second order accurate Roe type dimensionally split finite volume scheme on structured quadrilaterals show that the fluctuation splitting scheme is much less diffusive, therefore more accurate than the classical dimensionally split finite volume scheme (see figure 2.).

We also investigated the spatial accuracy of the fluctuation splitting schemes by means of grid convergence studies on smooth test cases and we found that the numerical results were close to the theoretical expectations (first order N scheme and second order B, LDA scheme). The study of the shock capturing properties of the nonconservative Galilean invariant governing equations with the source term is the subject of a recent work of the authors of this paper (Csik Á, Deconinck H and Poedts S, 2000). It is shown that the proper discretization of the source term yields a consistent scheme, *i.e.* discontinuities are captured more accurately on finer meshes.

Acknowledgement

This research was supported by the F.W.O.-Vlaanderen.

References

- Brio M and Wu C C (1988). An upwind Differencing Scheme for the Equations of Ideal Magnetohydrodynamics. *Journal of Computational Physics* **75**, pp 400-422.
 Csik Á, Deconinck H and Poedts S (2000). On the Shock Capturing Properties of the Non Conservative Symmetrizable Form of the Ideal Magnetohydrodynamic Equations. *submitted to the Journal of Computational Physics*.

- Csik Á, Deconinck H and Poedts S (1999). Monotone Residual Distribution Schemes for the Ideal 2D Magnetohydrodynamic Equations on Unstructured Grids. *AIAA-CP 99-3325*, pp 644-656.
- Dai W and Woodward P R (1994). Extension of the piecewise parabolic method (PPM) to multidimensional ideal magnetohydrodynamics. *Journal of Computational Physics* **115**, pp 485-514.
- Deconinck H, Struijs R and Roe P L (1990). Fluctuation Splitting for multidimensional convection problems: an alternative to finite volume and finite element methods. *VKI LS 1990-03, Computational Fluid Dynamics*.
- Godunov S K (1972). Symmetric form of the equations of magnetohydrodynamics. *Numerical Methods for Mechanics of Continuum Medium* **1**, pp 26-31.
- Issman E, Degrez G and Deconinck H (1996). Implicit upwind residual distribution Euler/Navier-Stokes solver on unstructured meshes. *AIAA Journal* **34/10**, pp 2021-2028.
- Pailiere H, Deconinck H and Roe P L (1995). Conservative upwind residual distribution schemes based on the steady characteristics of the Euler equations. *AIAA-CP 95-1700*, pp 592-605.
- Powell K G, Roe P L, Myong R S, Gombosi T and De Zeeuw D L (1995). An upwind scheme for ideal magnetohydrodynamics. *AIAA-CP 95-1704*, pp 661-675.
- Roe P L (1982). *Fluctuations and Signals - A Framework for Numerical Evolution Problems*. *Numerical Methods for Fluid Dynamics*. Academic Press.
- van der Weide E, Deconinck H, Issmann E and Degrez G (1999). Fluctuation Splitting Schemes for multidimensional convection problems: an alternative to finite volume and finite element methods. *Computational Mechanics* **23/2**, pp 199-208.
- van der Weide E (1998). Compressible Flow Simulation on Unstructured Grids using Multi-dimensional Upwind Schemes. *PhD thesis*, Technische Universiteit Delft, The Netherlands.
- Zachary A L, Colella P (1992). A Higher-Order Godunov Method for the Equations of Ideal Magnetohydrodynamics. *Journal of Computational Physics* **99**, pp 341-347.

APPLICATION OF TVD HIGH RESOLUTION SCHEMES TO THE VISCOUS SHOCK TUBE PROBLEM

VIRGINIE DARU

*Lab. SINUMEF, ENSAM,
151 Bd de l'Hôpital,
75013 PARIS, FRANCE.
Email: Virginie.Daru@paris.ensam.fr*

AND

CHRISTIAN TENAUD

*LIMSI-UPR CNRS 3251,
BP 133, ORSAY Cedex, FRANCE.
Email: tenaud@limsi.fr*

Abstract. In order to evaluate the accuracy of several high resolution TVD schemes in solving complex unsteady viscous shocked flows, we study the flow produced by the interaction of a reflected shock wave with the incident boundary layer in a shock tube, resulting in a lambda-shaped like bifurcated pattern. Two types of discretization, namely a combined time and space discretization and an independent time and space discretization are considered. Both methods are associated with several limiters, among which a more accurate recent family of limiters depending on the local wave velocity. Our results highlight the difficulty to obtain a converged solution for a Reynolds number value of 1000. In the meantime the above family of limiters, associated with a MacCormack type scheme, is shown to be the more accurate for solving this kind of problem.

1. Introduction

Our aim here is to evaluate the accuracy of a number of TVD high resolution schemes to solve a complex unsteady viscous flow, namely the viscous interaction between the boundary layer generated behind a shock wave

travelling in a shock tube after reflection of the latter at the end wall. Within a Mach number range of the incident shock, the interaction results in a lambda-shape like bifurcated shock wave pattern (Mark, 1958). This is due to the fact that the stagnation pressure of the boundary-layer becomes lower than the pressure behind the reflected shock. Therefore, the fluid cannot pass under the reflected wave and a separated flow region appears, generating a "bubble" which is dragged upstream with the reflected shock (Fig. 1). We compare the numerical solutions obtained for this problem

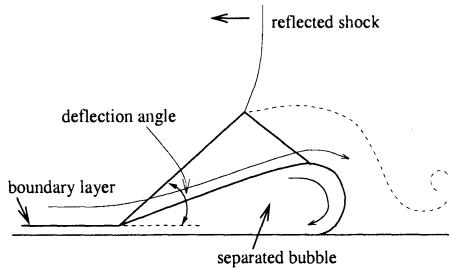


Figure 1. Bifurcated shock pattern.

using several numerical schemes and limiter functions. Beside the classical limiters, which retain only a simplified condition to satisfy the TVD constraints, we also use a family of limiters depending on the local Courant number which have more recently been used with some success for inviscid flows by several authors.

2. Numerical schemes

The study is limited to the two-dimensional laminar case. We solve the Navier-Stokes equations with constant viscosity coefficients written in cartesian coordinates, completed with a perfect gas law.

Two types of discretization, namely a combined time and space discretization (CTS) and an independent time and space discretization (ITS) are considered. Both methods are associated with several limiters.

In the CTS approach, we use the MacCormack scheme, to which is added a limiting correction term in order to render the scheme TVD. Beside classical limiters, we have used recent specific limiters which hybridize a third order scheme and a first order scheme in the scalar linear 1D case. By taking the upper bound of the TVD constraints, one obtains the limiter used by (Arora and Roe, 1997) and (Jeng and Payne , 1995), which we call here O3Sup :

$$\Phi^{O3S_{up}}(r, \nu) = \max \left[0, \min \left(\frac{2r}{\nu}, 1 - \frac{1+\nu}{3}(1-r), \frac{2}{1-\nu} \right) \right]$$

where ν is the local wave velocity and r is the gradient ratio. In the system case, the underlying high order scheme is no more third order but we expect the correction term to reduce the truncation error in smooth regions.

In the ITS approach, the time integration is performed by means of a third order Runge-Kutta method (Shu et al., 1991). The space discretization of the convective terms is based on a Roe's approximate Riemann solver. Two shock capturing schemes have been implemented: a second-order upwind TVD scheme , developed in (Harten, 1983) and (Yee and Harten, 1987) and a MUSCL-TVD scheme , using the local characteristic approach. We only used classical limiters in conjunction with the ITS approach, as the family of limiters described above is not adequate in this framework. More details about the schemes can be found in (Daru and Tenaud, 2001).

Before reporting our results about the viscous shock tube problem, let us summarize here some of the results we obtained using the different methods on selected 1D and 2D inviscid test cases. First, when using the CTS approach, the O3Sup limiter improves the solution whatever the CFL number is, compared to classical limiters which are more diffusive. Second, when using the same limiter, the CTS approach gives better results than the ITS method when the CFL number increases. This can be explained by the fact that around extrema the leading first order term of the truncation error is independent of the CFL number for the ITS method, which is not the case when a CTS discretization is used. These errors are equivalent only if one uses a very small CFL number. Third, when using the ITS approach, the Harten-Yee and MUSCL schemes give nearly the same results.

Following these results, we have retained only the TVD Harten-Yee scheme with the Van-Leer Harmonic limiter (HYVL) and the Mack-Cormack scheme using the O3Sup limiter function (MCO3Sup) to compute the viscous shock tube.

3. Description of the flow in the viscous shock tube

We consider a unit side length square shock tube with insulated walls. The diaphragm is initially located in the middle of the tube ($x = 0.5$). The initial state, in terms of dimensionless quantities, is on the left of the diaphragm: $\rho_L = 120$, $p_L = \rho_L/\gamma$, $u_L = v_L = 0$, and on the right: $\rho_R = 1.2$, $p_R = \rho_R/\gamma$, $u_R = v_R = 0$. At the initial time, the diaphragm is broken. A shock wave, followed by a contact discontinuity, moves to the right (the shock Mach number is equal to 2.37), creating a boundary layer along the horizontal wall. The incident shock wave reflects at the right end wall approximately at time $t = 0.2$. The interaction with the boundary layer results in a lambda-shape like shock pattern, under which a separated boundary layer "bubble" takes place. The bubble is delimited by a very unstable su-

personnic shear layer. The triple point emerging from the lambda-shape like shock pattern generates a slip line which rolls up around the boundary layer bubble. The contact discontinuity stays stationary, close to the right end wall, after interaction with the reflected shock.

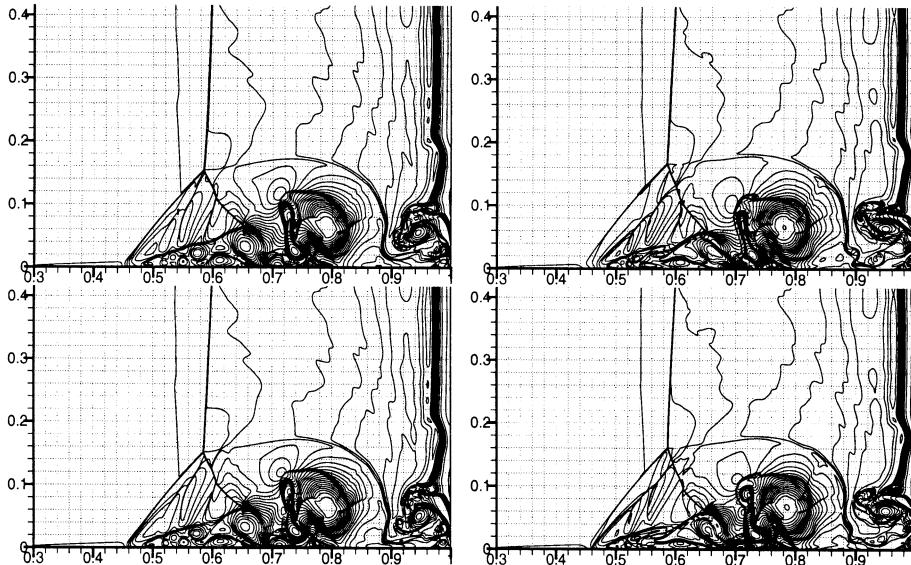


Figure 2. $Re=1000$, $t=1$, density contours. Left: HYVL scheme, right: MCO3Sup scheme. Mesh size $\delta x = 6.7 \times 10^{-4}$ (top), $\delta x = 5 \times 10^{-4}$ (bottom).

4. Results for $Re = 1000$

We compare in Fig. 2 the density fields obtained at a dimensionless time $t = 1$ using the HYVL and MCO3Sup schemes, in two cartesian uniform meshes defined by $\delta x = 6.7 \times 10^{-4}$ and $\delta x = 5 \times 10^{-4}$. We can see that small structures are generated just downstream of the boundary layer separation, producing distortions of the slip line delimiting the boundary-layer bubble and inducing oblique shock waves moving along it. These oblique shock waves perturb the lambda-shape like shock pattern. Shocklets are generated inside the large vortices. All these flow characteristics being very sensitive to small perturbations, it is extremely difficult to ascertain the quality of the schemes (there is probably not a well defined convergence process). However, as a general comment, we can observe that the HYVL results are smoother (see for instance the vortical structures inside the boundary layer bubble), and the variation of the density field when the

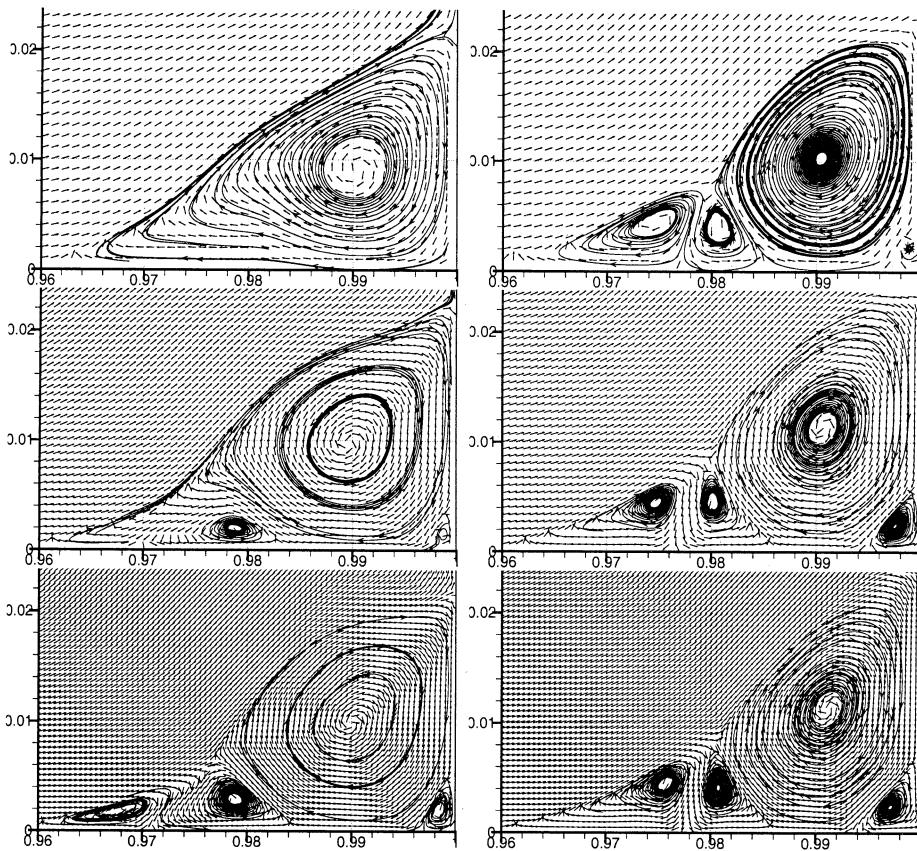


Figure 3. $Re=1000$, $t=1$, velocity vector field and instantaneous streamlines. Left: HYVL scheme, right: MCO3Sup scheme. Mesh size $\delta x = 10^{-3}$ (top), $\delta x = 6.7 \times 10^{-4}$ (middle), $\delta x = 5 \times 10^{-4}$ (bottom).

mesh is refined seems weak. In fact we have noticed that the MCO3Sup results are very sensitive to the mesh size for coarse meshes.

Now let us have a closer insight in the right bottom corner : we have represented in Fig. 3 the velocity vector field obtained using the two schemes in three successively refined meshes. Here we clearly see that the MCO3Sup scheme converges faster than the HYVL scheme : the results on the two finer meshes are almost the same, indicating that the solution is almost converged in this area on the intermediate mesh. Moreover, the four vortices are already captured in the coarsest mesh, though the small one in the corner is barely outlined. In contrast, we can see that three vortices are hardly captured using the HYVL scheme on the intermediate mesh. The four vortices appear only on the finer mesh. The coarsest mesh solution

reproduces only one vortex.

Although we cannot assure that we have reached grid convergence in all regions of the flow, the close examination of all these results leads us to conclude that the MCO3Sup scheme has better convergence properties, and that the solution given by this scheme in the finest grid is nearly converged. A comparison with the solutions obtained using the MacCormack scheme associated to a Van Leer limiter shows that the O3Sup limiter reduces the overall dissipation of the scheme, but the most important effect for the convergence properties is due to the choice of the CTS approach rather than the ITS approach.

5. Conclusions

The CTS method associated with the O3Sup limiter is clearly the best scheme in terms of accuracy and convergence properties. A sensible improvement is notably due to the limiter itself, compared to good quality classical limiters.

This viscous shock tube problem could constitute an interesting test case to compare high resolution numerical schemes in the case of compressible viscous unsteady flows. We have considered in this work only TVD schemes. There exists other approaches, particularly the ENO and WENO schemes which could be interesting to test on this flow case.

Acknowledgements

Part of the computations have been carried out on the Cray C98 of the IDRIS/CNRS. The authors greatly acknowledge the support of these institutions.

References

- Arora M. and Roe P.L. (1997). A well-behaved TVD limiter for high-resolution calculations of unsteady flows. *J. of Comp. Phys.* **132**, pp 3-11.
- Daru V. and Tenaud C. (2001). Evaluation of TVD high resolution schemes for unsteady viscous shocked flows. *Computers and Fluids* **30**, pp 89-113.
- Harten A. (1983). High Resolution Schemes for Hyperbolic Conservation Laws. *J. of Comp. Phys.* **49**, pp 357-393.
- Jeng Y.N. and Payne U.J. (1995). An adaptive TVD limiter. *J. of Comp. Phys.* **118**, pp 229-241.
- Mark H. (1958). The interaction of a reflected shock wave with the boundary layer in a shock tube. *NACA TM 1418*.
- Shu C.W., Erlebacher G., Zang T.A., Whitaker D. and Osher S. (1991). High-Order ENO Schemes Applied to Two- and Three-Dimensional Compressible Flow. *ICASE Report 91-38*.
- Yee H.C. and Harten A. (1987). Implicit TVD Schemes for Hyperbolic Conservation Laws in Curvilinear Coordinates. *AIAA Journal* **25** (2), pp 266-274.

COMPARISON OF NUMERICAL SOLVERS WITH GODUNOV SCHEME FOR MULTICOMPONENT TURBULENT FLOWS

E. DECLERCQ

*C.E.M.I.F. Evry University,
40, rue du Pelvoux,
Courcouronnes Val d'Essonne
91020 Evry Cedex, France
Email:declercq@worldonline.fr*

Abstract. This contribution's topic is the resolution by different numerical solvers of a multicomponent, compressive, turbulent flow. The unique associated Riemann Problem's solution is identified thanks to an entropy characterization. An exact Riemann solver is implemented and called by Godunov scheme. Some simulations are introduced to exhibit a comparison between Godunov scheme, Vfroe-nc and Rusanov scheme.

1. A turbulence model for multicomponent flows

The model is written for a polytropic isentropic gas. In this work, we are interested by the one order closure model and particularly the coupling between turbulence and pressure. Reynolds' tensor is described through the turbulent kinetic energy K of the mixture. The system is closed thanks to the K evolution equation.

The Favre's average variables describing the flow are : $(\rho, \rho\alpha, \rho U, K)$
The density of the mixture is noted by ρ . $\rho\alpha$ stands for the density of one of the fluids, with α the mass fraction of one component in the mixture. U stands for the velocity and K for the turbulence of the mixture.

$$C = \begin{pmatrix} \rho \\ \rho\alpha \\ \rho U \end{pmatrix} \quad F(C) = \begin{pmatrix} \rho U \\ \rho\alpha U \\ \rho U \otimes U + \left(\frac{2}{3}K + P\right)I \end{pmatrix} \quad (1)$$

Setting $W = (C, K)$, this system is conservative in C but not in K variable

$$(\mathcal{S}) \begin{cases} \partial_t C + \nabla F(C) = 0 \\ \partial_t K + \nabla(KU) + \frac{2}{3}K\nabla U = 0 \end{cases} \quad (2)$$

1.1. FROM A 3D PROBLEM TO THE 1D RIEMANN PROBLEM

It is well known that finite volume upwinding schemes are efficient methods to solve such no linear system. The most natural finite volume method is the Godunov's method (Godunov S K , 1957). It requires the exact solution on each interface of the Riemann Problem associated to (\mathcal{S}) . The Riemann solution of the multidimensional (\mathcal{S}) system is unknown. So we have to come down to the solution of the associated one-dimensional problem (\mathcal{S}_n) written in the normal direction of the interface. We set by (PR) the associated Riemann Problem, and by $W^*(\frac{x}{t}, W_l, W_r)$ its solution :

$$(PR) \begin{cases} \partial_t W_n + N \partial_n W_n = 0 & (\mathcal{S}_n) \\ W_n(X, t=0) = W_l & \text{if } X.n < 0 \\ W_n(X, t=0) = W_r & \text{if } X.n > 0 \end{cases} \quad \text{I. C.} \quad (3)$$

The (\mathcal{S}_n) system is hyperbolic. The exact solution of (PR) is composed of four constant states separated by shock waves or rarefactions and a contact discontinuity (Godlewski and Raviart, 1996). Shock curves must verify the Rankine-Hugoniot jump relations. To the non-conservative equation on K corresponds an approximate jump relation, which depends of the choice of the integration's way of the non-conservative product. The construction of the rarefactions waves is based on the fact that the Riemann invariants remain constant on rarefactions curves (see (Declercq E, 1999)).

Come back to the Godunov scheme, we introduce the notations :

$V(i)$ the neighboring cells of cell "i" (not including cell i)

l_{ij} the interface measure : $l_{ij} = |\partial\Omega_i \cap \partial\Omega_j|$

n_{ij} the unit normal vector between cell i and cell j.

We set by $\mathcal{F}_{jk}(C^*, n_{jk})$ the normal numerical flux between W_j and W_k :

$$\mathcal{F}_{jk}(C^*, n_{jk}) = \vec{F}(C^*(0, W_j, W_k)) \cdot \vec{n}_{jk} = \begin{pmatrix} \rho^* u_n^* \\ \rho^* \alpha^* u_n^* \\ \rho^* u_n^* \vec{u}^* + (\frac{2}{3}K^* + P^*) \cdot \vec{n} \end{pmatrix}$$

$$C_j^{n+1} \equiv C_j^n - \frac{\Delta t}{|\Omega_j|} \sum_{k \in V(j)} \mathcal{F}_{jk}(C^*, n_{jk}) l_{jk} \quad (4)$$

With \bar{K}_j defined like in (Herard JM, Forestier A, Louis X , 1994) as :

$$\bar{K}_j = \left(\sum_{j \in V(i)} K_{ij}^* \right) / \left(\sum_{j \in V(i)} 1 \right) \quad (5)$$

$$K_j^{n+1} = K_j^n - \frac{\Delta t}{|\Omega_j|} \sum_{k \in V(j)} (K^* u_n^*) l_{jk} - \frac{2}{3} \frac{\Delta t}{|\Omega_j|} \bar{K}_j \sum_{k \in V(j)} u_n^* l_{jk} \quad (6)$$

1.2. ENTROPY CHARACTERIZATION AND UNIQUENESS OF THE SOLUTION

Our system (\mathcal{S}) admits two supplementary conservative variables :

$$\mathcal{F} = \frac{K}{\rho^{\frac{2}{3}}} \text{ and } \mathcal{E} = \frac{\rho U^2}{2} + K + \rho \int \frac{P(\rho\alpha)}{\rho^2} d\rho \quad (7)$$

In keeping with the second thermodynamic principle, a φ convex entropy is growing on a physical shock : $\sigma[\varphi(\mathcal{U})] \geq [\psi(\mathcal{U})]$.

The equivalence between Lax inequalities and compressive shock is shown. But the equivalence between entropy shock and compressive shock is demonstrated for only the physical entropy \mathcal{E} . \mathcal{F} has no physical sense, because its growing on 1-shock curve implies incompressive shock. The computation of the Riemann solution on the entropy shock curves ensure the uniqueness of the result.

2. Other numerical methods and preferences

The advantages of the exact solver are well known : positivity respect, entropy solution. But we have to balance these advantages by the fact that the method uses more CPU than linearized solver which doesn't require analytical computations. Then we introduce different schemes to analyze where a method is more or less efficient than an other.

2.1. A LINEARIZED SOLVER : VFROE-NC

Vfroe scheme was introduced by Faille, Gallouët, Masella in 1996. It is based on a local resolution of a linearized Riemann problem. The numeric flux is defined, like Godunov scheme, by the physical flux computed at the interface solution of the linearized problem. An extension of this scheme was introduced by Buffard, Gallouët and Hérard (Buffard T, Gallouët T, Hérard JM, 1998). Vfroe-nc uses the nonconservative variables to preserve Riemann invariants through contact discontinuities.

Thus we prefer the variables $(P(\rho\alpha), U)$ to $(\rho\alpha, \rho U)$.

With $Y = (\alpha, u_n, u_t, K, P)$, $\tau = \frac{1}{\rho}$ and the linearized variable

$\hat{Y} = \frac{Y_l + Y_r}{2}$, we have to solve the linear system :

$$\frac{\partial Y}{\partial t} + A_n(\hat{Y}) \frac{\partial Y}{\partial n} = 0 \quad (8)$$

With the turbulent celerity $\hat{c}'^2 = (\gamma \hat{P} + \frac{10}{9} \hat{K}) \hat{\tau}$ we obtain the eigenvalues

$$\lambda_1(\hat{Y}) = \hat{u}_n - \hat{c}', \quad \lambda_{2,3,4} = \hat{u}_n, \quad \lambda_5 = \hat{u}_n + \hat{c}' \quad (9)$$

Defining the intermediate states by a combination of the eigenvectors r_i

$$Y_r = Y_l + \sum_{i=1,5} \alpha_i r_i(\hat{Y}) \quad (10)$$

$$\begin{cases} Y_1 = Y_l + \frac{1}{2\hat{c}'} \{ [u_n] - \frac{\hat{\tau}}{\hat{c}'} [\frac{2}{3}K + P] \} r_1(\hat{Y}) \\ Y_2 = Y_r - \frac{1}{2\hat{c}'} \{ [u_n] + \frac{\hat{\tau}}{\hat{c}'} [\frac{2}{3}K + P] \} r_5(\hat{Y}) \end{cases} \quad (11)$$

The linearized Riemann solution $Y^*(Y_l, Y_r, 0)$ is given by :

$$\begin{cases} Y^* = Y_l & \text{if } \hat{u}_n - \hat{c}' > 0 \\ Y^* = Y_1 & \text{if } 0 < \hat{u}_n < \hat{c}' \\ Y^* = Y_2 & \text{if } -\hat{c}' < \hat{u}_n < 0 \\ Y^* = Y_r & \text{if } \hat{u}_n + \hat{c}' < 0 \end{cases} \quad (12)$$

The extension of Vfroe-nc to nonconservative systems is given by (13), with \bar{K}_j defined in (5)

$$W_j^{n+1} = W_j^n - \frac{\Delta t}{|\Omega_j|} \sum_{k \in V(j)} (\{\mathcal{F}_{jk}(W^*, n_{jk})\} l_{jk} - \frac{2\bar{K}_j}{3} u_k^* n_k l_{jk} \delta_{\{W=K\}}) \quad (13)$$

The linearized solver generates intermediate states that may not respect physical positivity. For example in the mirror states for a double rarefaction wave $U_{ni} < 0$, we have to set :

$$K_1 = \max(0, K_i(1 + \frac{5}{3} \frac{U_{ni}}{\hat{c}'_i})), \quad P_1 = \max(0, P_i(1 + \gamma \frac{U_{ni}}{\hat{c}'_i})) \quad (14)$$

It is also possible that the linearized velocity \hat{u}_n implies a bad choice of α . A big jump of α may generate, in the next time step, such a pressure that the deduced mass fraction would be out of the $(0, 1)$ definition set.

2.2. RUSANOV SCHEME

Rusanov scheme (Rusanov V, 1961) was introduced in 1961, it doesn't need any exact or approximative Riemann problem's resolution.

$$\partial_t W + \sum_{i=1,2} (H_i(W))_{,i} + \sum_{i=1,2} B_i W_{,i} = 0 \quad (15)$$

$$\tilde{A}_n = \sum_{i=1,2} \frac{\partial H_i(W)}{\partial W} \cdot n_i + \sum_{i=1,2} n_i B_i \quad (16)$$

We introduce $\tilde{S}_{i,j} = \max_{i,j}(\max_k |\lambda_k(W_i)|, |\lambda_k(W_j)|)$ with λ_k the eigenvalues of \tilde{A}_n , and $W_{ij} = \frac{W_i + W_j}{2}$. With \mathcal{F}^R defined by (17), we get the following expression of Rusanov scheme :

$$\begin{cases} C_i^{n+1} = C_i^n - \frac{\Delta t}{|\Omega_i|} \sum_{j \in V(i)} \mathcal{F}^R(W_i^n, W_j^n, n_{ij}) l_{ij} \\ K_i^{n+1} = K_i^n - \frac{\Delta t}{|\Omega_i|} \sum_{j \in V(i)} ((K u_n)_{ij} - \frac{1}{2} \tilde{S}_{ij} (K_j - K_i) + \frac{2}{3} K_i u_{ij} n_{ij})^n l_{ij} \\ \mathcal{F}^R(W_i, W_j, n_{ij}) = \frac{1}{2} (F(W_i) n_{ij} + F(W_j) n_{ij} - \tilde{S}_{ij} (C_j - C_i)) \end{cases} \quad (17)$$

Rusanov scheme has been chosen for its respect of $\rho, \rho\alpha$ and K positivity, and because it preserves the mass fraction $0 < \alpha < 1$.

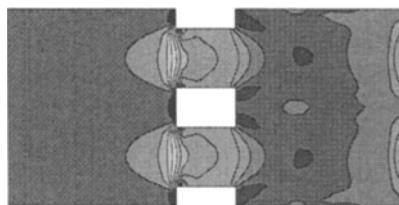
2.3. NUMERICAL RESULTS

We present, in the following page, a 2D test case of a onecomponent flow through a tube with obstacles. The air is coming from the left to the right side of the tube.

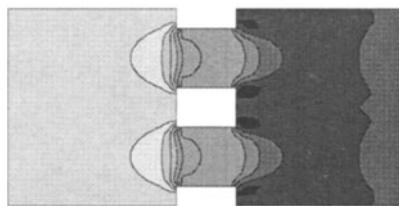
The initial conditions are : $(\rho, u, v, \alpha, K, P) = (1.1, 600, 0, 1, 1000, 100000)$ This case of a tube with obstacles reveals the advantages and deficiencies of the different methods. Vfroe-nc generates some negative turbulence in spite of the boundary positivity. With Rusanov method, positivity problems do not appears, but the K evaluation is very approximate.

3. Conclusion

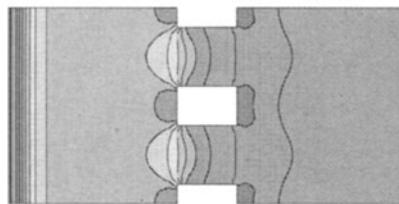
In general we advocate to use Vfroe scheme, but in some extreme tests (strong rarefactions) the cost of an exact solver is justified by the robustness required. A good compromise is to use successively the two schemes, the exact one for initialization and to deal with the boundary conditions, and a linearized one to go on with a reduced cost.



Godunov scheme



Vfroe-nc scheme



Rusanov scheme

Figure 1. Comparison of turbulence at the critical time where $K < 0$ by Vfroe

References

- Buffard T, Gallouët T, Hérard JM (1998). A sequel to a rough Godunov scheme : application to real gases. *Écoles CEA-EDF-INRIA*, pp.363-382.
- Declercq E, (1999) Comparaison de solveurs numériques pour le traitement de la turbulence bifluide. *PhD Thesis of Evry University*.
- Godlewski E and Raviart P A (1996). Numerical Approximation of Hyperbolic Systems of Conservation Laws. Springer.
- Godunov S K (1957) A finite difference method for the numerical computation of discontinuous solutions of the equations of fluid dynamics. *Math. Sb. 47*, pp.357-393
- Herard JM, Forestier A, Louis X (1994) A Non Strictly Hyperbolic System To Describe Compressible Turbulence. *Rapport E.D.F/D.E.R.*, HE-41/94/11A.
- Rusanov V (1961) Calculation of interaction of non-steady shock waves with obstacles. *Journal of Computational Mathematics and Physics USSR*, Vol 1 PP.267-279.

GODUNOV-TYPE SCHEMES FOR THE MHD EQUATIONS

A. DEDNER, D. KRÖNER, C. ROHDE, M. WESENBERG

Institute for Applied Mathematics,

Freiburg University, Germany

dedner|dietmar|chris|wesenber@mathematik.uni-freiburg.de

Abstract. If dissipative effects are neglected, the equations of ideal magnetohydrodynamics (MHD) are a mathematical model for the flow of a compressible, electrically conducting fluid which is influenced by a magnetic field. They are derived from the Euler equations of fluid dynamics and the Maxwell equations and form a hyperbolic system of conservation laws. Since its behaviour is much more complicated than the Euler system's, theoretical results and numerical schemes have not yet reached the same level as in the Euler case. This paper focuses on available approximate MHD Riemann solvers, which can be used in Godunov-type finite volume schemes: we present results of an extensive comparison, which justify the choice of the solver we use in our multidimensional code for astrophysical simulations. Moreover, we summarize some important properties of the MHD system and explain how they may influence the solutions' structure. We conclude with two 2D applications from solar physics.

1. Equations

The MHD equations in 3D constitute the following system of partial differential equations:

$$\begin{aligned} \partial_t \rho + \nabla \cdot (\rho \vec{u}) &= 0, \\ \partial_t (\rho \vec{u}) + \nabla \cdot (\rho \vec{u} \vec{u}^T + \mathcal{P}) &= 0, \\ \partial_t \vec{B} + \nabla \times (\vec{B} \times \vec{u}) &= 0, \\ \partial_t (\rho E) + \nabla \cdot (\rho E \vec{u} + \mathcal{P} \vec{u}) &= 0, \\ \nabla \cdot \vec{B} &= 0. \end{aligned} \tag{1}$$

For the density ρ , the momentum $\rho\vec{u}$, the magnetic field \vec{B} , the total energy ρE , the gas pressure p and the unit matrix \mathcal{I} , the pressure tensor \mathcal{P} is defined as

$$\mathcal{P} := \left(p + \frac{1}{8\pi} |\vec{B}|^2 \right) \mathcal{I} - \frac{1}{4\pi} \vec{B} \vec{B}^T.$$

In order to close the system, we add the relations for an ideal polytropic gas

$$E = \varepsilon + \frac{1}{2} |\vec{u}|^2 + \frac{1}{8\pi\rho} |\vec{B}|^2, \quad p = (\gamma - 1)\rho\varepsilon,$$

where ε is the internal energy and γ stands for the adiabatic exponent. As usual, suitable initial and boundary conditions have to be added. The last equation in (1) has the physical meaning that magnetic monopoles must not arise; this condition is always fulfilled, if it is satisfied by the initial conditions.

If we define $\vec{U} := (\rho, \rho\vec{u}, \vec{B}, \rho E)$ and denote by \vec{F} , \vec{G} and \vec{H} the fluxes in the different coordinate directions, we may rewrite (1) as a system of conservation laws in the conserved variables \vec{U} :

$$\begin{aligned} \partial_t \vec{U} + \partial_x \vec{F}(\vec{U}) + \partial_y \vec{G}(\vec{U}) + \partial_z \vec{H}(\vec{U}) &= 0, \\ \partial_x B_x + \partial_y B_y + \partial_z B_z &= 0. \end{aligned} \tag{2}$$

If all unknowns are independent of z or of y and z , the system (2) can be reduced to two or one spatial dimensions respectively. Note that due to the cross coupling between momentum and magnetic field the number of equations is not reduced in 2D. Only in 1D, where (2) reads

$$\begin{aligned} \partial_t \vec{U} + \partial_x \vec{F}(\vec{U}) &= 0, \\ \partial_x B_x &= 0 \end{aligned} \tag{3}$$

and the first induction equation yields $\partial_t B_x = 0$, one equation can be dropped because B_x is constant in this case. Since the MHD system is invariant with respect to rotations, i.e. for any unit vector $\vec{n} \in \mathbb{R}^3$ there is a corresponding rotation matrix $\mathcal{T}(\vec{n})$ with

$$n_x \vec{F}(\vec{U}) + n_y \vec{G}(\vec{U}) + n_z \vec{H}(\vec{U}) = \mathcal{T}^{-1}(\vec{n}) \vec{F}(\mathcal{T}(\vec{n}) \vec{U}), \tag{4}$$

a multidimensional scheme can be built by using 1D fluxes in the normal directions of the grid cell interfaces.

2. Riemann Problems

The Euler equations in 1D are a strictly hyperbolic system, i.e. the three eigenvalues of the flux Jacobian are always distinct, and their characteristic

fields can either be classified as linearly degenerate (i.e. $\nabla\lambda \cdot \vec{r} := \nabla_{\vec{U}}\lambda(\vec{U}) \cdot \vec{r}(\vec{U}) \equiv 0$ for an eigenvalue λ and the corresponding right eigenvector \vec{r}) or as genuinely nonlinear (i.e. $\nabla\lambda \cdot \vec{r} \neq 0$ for all admissible values of the conservative variables), see (Kröner, 1997). These two properties do not hold for the MHD equations.

The MHD system (3) can be reduced to seven equations in 1D (see Section 1). This system is still hyperbolic, i.e. the flux Jacobian can always be diagonalized, but the relationship between its eigenvalues $\lambda_1, \dots, \lambda_7$ depends on the values of the conservative variables: If $B_x = 0$ we have $\lambda_1 < \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_6 < \lambda_7$, while the system is strictly hyperbolic if $B_x \neq 0$ and $B_y^2 + B_z^2 \neq 0$. In the remaining case ($B_x \neq 0$ and $B_y^2 + B_z^2 = 0$) we can have either two double or two triple eigenvalues. Moreover, the characteristic fields belonging to the Alfvén waves (λ_2 and λ_6) and the entropy wave (λ_4) are linearly degenerate, while $\nabla\lambda \cdot \vec{r}$ can be zero and nonzero for the remaining slow (λ_3 and λ_5) and fast (λ_1 and λ_7) magnetoacoustic waves, see (Brio and Wu, 1988).

In the classical approach “simple waves” (Laxian shocks, rarefaction waves and contact discontinuities) are attached to each other (with constant intermediate states in between) to construct a solution for a given Riemann problem (Smoller, 1983). The behaviour of the fast and slow waves mentioned above means that also “compound waves” (e.g. consisting of a shock and a directly attached rarefaction wave as in the Brio–Wu example below) have to be included in this construction (Liu, 1975), while in the non-strictly hyperbolic case this ansatz cannot be applied directly. Therefore the solution theory for the MHD Riemann problem is not completely settled so far. For the ongoing research in this direction we refer to (Beyerstedt and Freistühler, 1997) and the references therein.

The following two examples of Riemann problems illustrate some of these MHD-specific properties. The first one was suggested in (Brio and Wu, 1988); it is the classical Sod problem for Euler with an additional magnetic field. The second problem was introduced in (Ryu and Jones, 1995):

Brio–Wu Riemann problem ($\gamma = 2.0, B_x = 3/4\sqrt{4\pi}$)						
	ρ	u_x	u_y	u_z	B_y	B_z
$x < 0$	1.0	0.0	0.0	0.0	$\sqrt{4\pi}$	0.0
$x \geq 0$	0.125	0.0	0.0	0.0	$-\sqrt{4\pi}$	0.0

Ryu–Jones Riemann problem ($\gamma = 5/3, B_x = 2.0$)						
	ρ	u_x	u_y	u_z	B_y	B_z
$x < 0$	1.08	1.2	0.01	0.5	3.6	2.0
$x \geq 0$	1.213883	0.196413	0.127421	0.103348	3.815995	1.924447
						1.403094

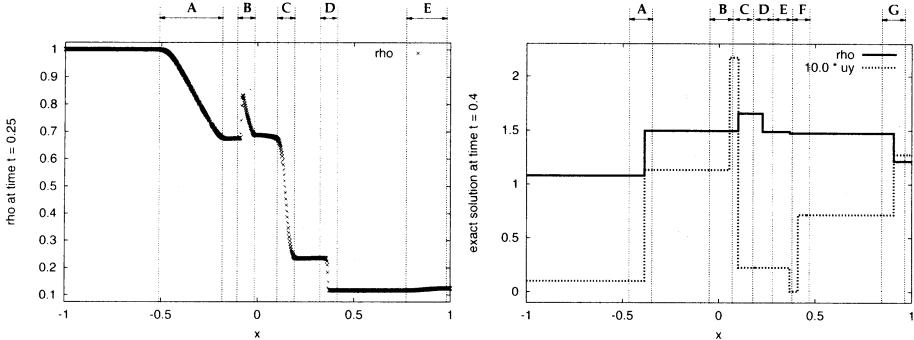


Figure 1. *Left:* Numerical solution of the Brio–Wu problem (ρ , $t = 0.25$); two rarefaction waves (A,E), one contact discontinuity (C), one Laxian shock (D) and one compound wave (shock + rarefaction, B). *Right:* Exact solution of the Ryu–Jones problem (ρ and $10 \cdot u_y$, $t = 0.4$); four Laxian shocks (A,C,E,G) and three contact discontinuities (B,D,F).

For the Brio–Wu problem there is no exact solution available, but numerical studies reveal the types of the arising waves. At the point where the compound wave switches from the shock to the rarefaction wave (see Fig. 1 left), λ_5 and λ_6 coincide and $\nabla \lambda_5 \cdot \vec{r}_5$ — which is non-zero in general — vanishes. In the Ryu–Jones problem, whose exact solution is known, all seven possible waves occur. This demonstrates the general complexity of MHD Riemann solutions. Moreover in this example there is no single primitive variable out of $\{\rho, u_x, u_y, u_z, B_y, B_z, p\}$ which is affected by all waves (see Fig. 1 right). This is another difference from the Euler equations and means that indicators for adaptive grid refinement as in (Dedner et al., 1999) should not be based on a single quantity.

3. Riemann Solvers

In the following we present some of the most important results we obtained from our comparison; the full results can be found in (Wesenberg, 2000). We implemented and tested the BCT (Bell et al., 1989), DW (Dai and Woodward, 1995), HLLE (and two new modifications we call HLLEMG / HLLEML, which extend the ideas in (Einfeldt et al., 1991) to MHD) and the Roe scheme as described in (Brio and Wu, 1988). Since the results of the Roe scheme with respect to CPU time, L^1 -error and experimental order of convergence (EOC) are almost identical with those of the DW scheme, we do not include the results for the Roe scheme. Similarly, the errors and EOCs of the HLLEMG and the DW scheme coincide for most test cases.

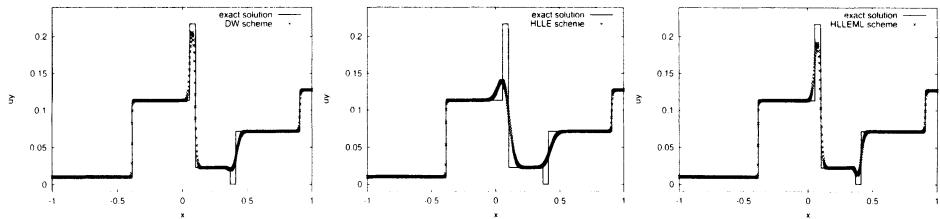
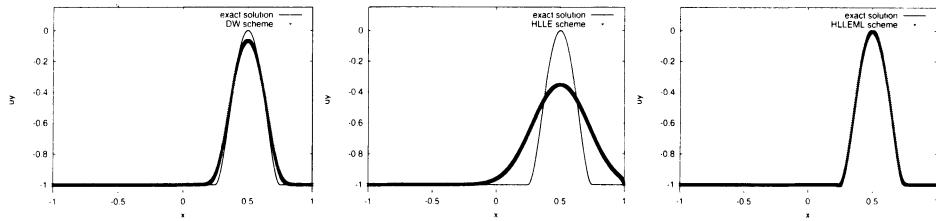


Figure 2. Numerical results for the Ryu-Jones problem; top: u_y of approximated and exact solution for DW (left), HLLE (middle) and HLLEML (right) scheme. Bottom: CPU-time, L^1 -errors and EOC for four solvers.

The numerical results for the Ryu-Jones Riemann problem (see Fig. 2) show that errors and EOCs for BCT, DW and HLLEML are on the same level, while BCT needs much more time to achieve these results than the two other fluxes. The errors and even the convergence rates of the HLLE scheme are much worse, and if we compare the CPU time needed to reach a fixed error bound (e.g. 0.15, see bold figures in the table of Fig. 2), it is even slower than DW and HLLEML.

As a further test problem we choose initial data such that the density and velocity distributions are simply advected with a constant speed. The corresponding results are shown in Fig. 3. Again errors and EOCs for BCT and DW are about the same, while the HLLE results are quite bad and those of HLLEML are by far the best. The CPU times are omitted because their relations are the same as in the Ryu-Jones example above.

In view of these results, we decided to use the DW scheme for our simulations, since BCT is too costly, HLLE performs too bad, HLLEMG is unstable for one test problem and DW turned out to be more reliable and

results for advection of ρ and \vec{u} ($t = 1.0$, $cfl = 0.9000$)

flux	dw		hlle		hlleml		
	h	$\ u - u_h\ _{L^1}$	EOC	$\ u - u_h\ _{L^1}$	EOC	$\ u - u_h\ _{L^1}$	EOC
0.0200		0.5839		1.1994		0.2839	
0.0100		0.3896	0.5838	1.1057	0.1173	0.1558	0.8653
0.0050		0.2396	0.7014	0.9225	0.2614	0.0846	0.8820
0.0025		0.1374	0.8016	0.7087	0.3803	0.0447	0.9183

Figure 3. Numerical results for advection of density and velocity; top: u_y of approximated and exact solution for DW (left), HLLE (middle) and HLLEML (right) scheme. Bottom: L^1 -errors and EOC for these solvers.

robust in multidimensional applications than HLLEML and Roe (Wesenberg, 2000).

4. Applications

Our code is designed for multidimensional simulations in solar physics. In particular we are interested in the rise of concentrated magnetic structures (“magnetic flux tubes”) through the sun’s atmosphere, which is assumed to be responsible for the development of sun spots. For this application we have to add gravitational source terms to system (1) and prescribe a static background “atmosphere”, which is only dependent on height. Starting from a perturbed horizontal layer of lower density which rises under the influence of gravitation, Fig. 4 shows that an increasing tangential magnetic field stabilizes this layer more and more against the arising Rayleigh–Taylor instabilities. A similar effect can be seen in the case of rising magnetic flux tubes: if the tube is “twisted” by an additional tangential magnetic field, it gets less fragmented and can therefore rise higher, see Fig. 5.

ACKNOWLEDGEMENTS: The authors were partially supported by the DFG–Schwerpunktprogramm “Analysis und Numerik von Erhaltungsgleichungen” and the EU–TMR research network for Hyperbolic Conservation Laws.

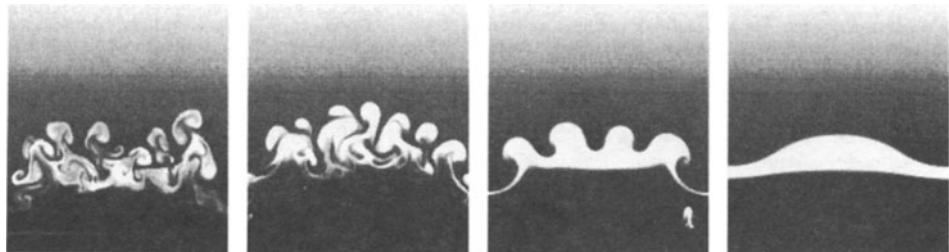


Figure 4. Rayleigh-Taylor instability with a tangential magnetic field increasing from left to right.

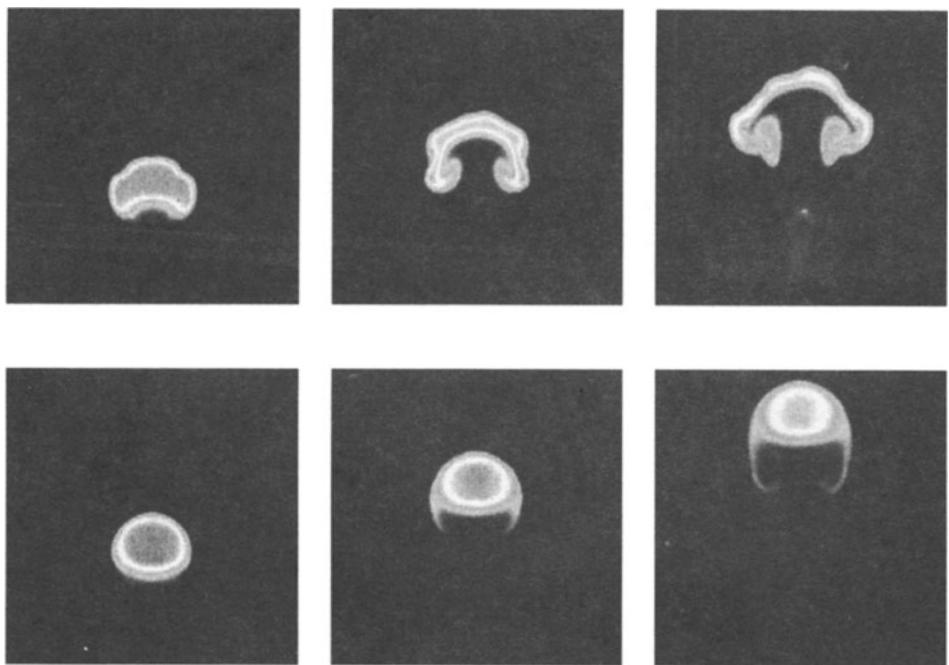


Figure 5. Rising magnetic flux tubes; *top*: untwisted, *bottom*: twisted.

References

- Bell JB, Colella P, and Trangenstein JA (1989): Higher Order Godunov Methods for General Systems of Hyperbolic Conservation Laws. *J. Comput. Phys.*, **82**: 362.
 Beyerstedt R, and Freistühler H (1997): A Class of Algorithms Solving Riemann's Initial Value Problem for Ideal Magnetohydrodynamics. *Preprint 4-97*, Institut für Mathematik, RWTH Aachen.
 Brio M, and Wu CC (1988): An Upwind Differencing Scheme for the Equations of Ideal Magnetohydrodynamics. *J. Comput. Phys.*, **75**: 400.

- Dai W, and Woodward PR (1995): A Simple Riemann Solver and High-Order Godunov Schemes for Hyperbolic Systems of Conservation Laws. *J. Comput. Phys.*, **121**: 51.
- Dedner A, Rohde C, Wesenberg M (1999): A MHD-Simulation in Solar Physics. Finite Volumes for Complex Applications II: 491. Vilsmeier R et al. (Editors). Hermès (Paris).
- Einfeldt B, Munz CD, Roe PL, and Sjögren B (1991): On Godunov-Type Methods near Low Densities. *J. Comput. Phys.*, **92**: 273.
- Kröner D (1997): Numerical Schemes for Conservation Laws. Wiley Teubner.
- Liu TP (1975): The Riemann Problem for General Systems of Conservation Laws. *J. Differ. Equations*, **18**: 218.
- Ryu D, and Jones TW (1995): Numerical Magnetohydrodynamics in Astrophysics: Algorithm and Tests for One-Dimensional Flow. *ApJ*, **442**: 228.
- Smoller J (1983): Shock Waves and Reaction-Diffusion Equations. Springer (Heidelberg, Berlin).
- Wesenberg M (2000) A Note on MHD Riemann Solvers. *In preparation*.

ABSORBING BOUNDARY CONDITIONS FOR ASTROPHYSICAL MHD SIMULATIONS

A. DEDNER, D. KRÖNER, M. WESENBERG

*Institute for Applied Mathematics,
Freiburg University, Germany
dedner|dietmar|wesenber@mathematik.uni-freiburg.de*

AND

I. SOFRONOV

*Keldysh Institute of Applied Mathematics RAS,
Moscow, Russia
sofronov@spp.keldysh.ru*

Abstract. Many problems for systems of conservation laws are formulated either on infinite domains or on domains which are by orders of magnitude larger than the interesting structures. In the first case, it is often impossible to find an exact representation of the problem which is suitable for numerical simulations. But even in the second case it can be difficult to perform the simulation on the whole domain, since much computational effort is wasted in uninteresting regions. Therefore the size of the computational domain has to be reduced, which introduces new boundaries without physical meaning. At these artificial boundaries suitable boundary conditions for the PDEs have to be formulated.

In this paper we will discuss a method to derive artificial non-reflecting boundary conditions for systems of conservation laws. We will concentrate on the equations of ideal magnetohydrodynamics (MHD) in a gravitationally balanced, stratified atmosphere. We will state the main results, discuss implementational aspects and show results in 1D and in 2D.

The equations of ideal magnetohydrodynamics — modeling the flow of an electrically conducting fluid under the influence of a magnetic field — are given by a system of partial differential equations:

$$\partial_t \rho + \nabla \cdot (\rho \mathbf{u}) = 0 \quad (1)$$

$$\partial_t (\rho \mathbf{u}) + \nabla \cdot (\rho \mathbf{u} \mathbf{u}^T + \mathcal{P}) = \rho \mathbf{g} \quad (2)$$

$$\partial_t \mathbf{B} + \nabla \times (\mathbf{B} \times \mathbf{u}) = 0 \quad (3)$$

$$\partial_t (\rho E) + \nabla \cdot (\rho E \mathbf{u} + \mathcal{P} \mathbf{u}) = \rho \mathbf{g} \cdot \mathbf{u} \quad (4)$$

$$\nabla \cdot \mathbf{B} = 0. \quad (5)$$

The total pressure tensor \mathcal{P} is given by

$$\mathcal{P} = \left(p + \frac{1}{8\pi} |\mathbf{B}|^2 \right) \mathcal{I} - \frac{1}{4\pi} \mathbf{B} \mathbf{B}^T$$

and $\mathbf{g} = (0, 0, g)^t$ is the gravitational force. Using the equation of state for an ideal polytropic gas (with the adiabatic exponent $\gamma > 1$)

$$p = (\gamma - 1) \left(\rho E - \rho \frac{1}{2} |\mathbf{u}|^2 - \frac{1}{8\pi} |\mathbf{B}|^2 \right),$$

the system can be rewritten as a (constrained) hyperbolic system in the density ρ , momentum $\rho \mathbf{u}$, magnetic field \mathbf{B} and the total energy ρE . We will denote the vector of conserved quantities with \mathbf{U} .

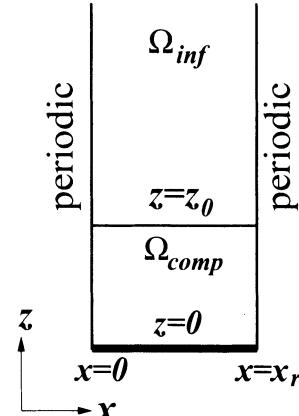
For the numerical solution of the MHD equations we rely on Godunov type schemes as developed in (Bell et al., 1989; Brio and Wu, 1988; Dai and Woodward, 1995; Wesenberg, 2000). For these schemes, boundary conditions are often implemented by introducing an additional layer of cells at the boundaries. These so called ghost cells lead to a simple algorithm for the time evolution, since fluxes over the boundaries do not require a special treatment. For this approach it is necessary to compute the values in the ghost cells before updating the values in the interior cells. At artificial boundaries — introduced for purely numerical reasons — formulas to compute the ghost cell values are often difficult to derive. Some techniques for handling this problem and further references can be found in (Grote and Keller, 2000; Kröner, 1991; Sofronov, 1998).

In this paper we want to demonstrate the importance of using suitable boundary conditions at certain artificial boundaries. The application we are interested in is the interaction of plasma motion and magnetic fields in the solar convection zone, which can be modeled by the compressible MHD equations (1) — (5). For some of these problems in lower regions of the convection zone we can use a stratified background atmosphere with local perturbations as initial conditions. To minimize the computational effort of resolving the evolving small scale structures, the size of the computational domain has to be reduced as far as possible. This requires the introduction of artificial boundaries. Our method for formulating transparent boundary conditions for these artificial boundaries generalizes the technique presented

in (Sofronov, 1998) to MHD. It is based on the derivation of an equation for the pressure perturbation at the boundary, which includes a nonlocal term in time; by using a special approximation, this term can be evaluated in a time-stepping manner, so that the numerical method stays local in time. We arrive at this equation by linearizing the MHD equations around a background atmosphere, thus assuming that the perturbations at the boundary are smooth and small enough. In our numerical examples we will show that, to a certain extent, even larger perturbations are hardly reflected at the boundary.

1. Derivation of the boundary condition

We will derive artificial boundary conditions for the top boundary at $z = z_0 > 0$. The problems are formulated in a domain $\Omega := [0, x_r] \times [0, \infty)$ with periodic boundary conditions on the vertical boundaries. On the lower boundary (at $z = 0$) we will use Dirichlet conditions for all our computations. (A discussion of transparent boundary conditions for the lower boundary can be found in (Dedner et al., 2000).) To derive the boundary conditions we split Ω in the computational domain $\Omega_{comp} := [0, x_r] \times [0, z_0]$ and the far field domain $\Omega_{inf} := [0, x_r] \times [z_0, \infty)$.



We find the boundary conditions at the top boundary of Ω_{comp} by studying the behavior of solutions to the linear model of (1) — (5) in Ω_{inf} . We decompose the unknown quantities $\rho, \mathbf{u} = (u, v, w)^t, \mathbf{B}$ and p using the background atmosphere and the perturbations, i.e. $\rho = \bar{\rho} + \tilde{\rho}$. For the background atmosphere we assume

$$\dot{\bar{\rho}} = \dot{\bar{\rho}}(z), \quad \dot{\mathbf{u}} \equiv 0, \quad \dot{\mathbf{B}} \equiv 0, \quad \mathbf{g}(z) = (0, 0, g(z)) \quad \text{and} \quad \dot{\bar{p}} = \dot{\bar{\rho}}^\gamma.$$

We then derive the following linear equations for the perturbations:

$$\partial_t \tilde{\rho} + \bar{\rho} \partial_x \tilde{u} + \partial_z (\bar{\rho} \tilde{w}) = 0 \quad (6)$$

$$\dot{\bar{\rho}} \partial_t \tilde{u} + \partial_x \tilde{p} = 0 \quad (7)$$

$$\ddot{\bar{\rho}} \partial_t \tilde{w} + \partial_z \tilde{p} = \tilde{\rho} g \quad (8)$$

$$\partial_t \tilde{p} + \gamma \dot{\bar{\rho}} (\partial_x \tilde{u} + \partial_z \tilde{w}) = -\dot{\bar{\rho}} g \tilde{w} \quad (9)$$

$$\partial_t \tilde{v} = 0, \quad \partial_t \tilde{\mathbf{B}} = 0. \quad (10)$$

After some calculations, we arrive at a wave equation for the pressure perturbation with varying coefficients:

$$\partial_{tt} \tilde{p} - \frac{\gamma \ddot{\bar{\rho}}}{\dot{\bar{\rho}}} (\partial_{xx} \tilde{p} + \partial_{zz} \tilde{p}) + g \partial_z \tilde{p} + \left(\partial_z g - \frac{\gamma - 1}{\gamma} g^2 \frac{\dot{\bar{\rho}}}{\dot{\bar{\rho}}} \right) \tilde{p} = 0 \quad (11)$$

The main work is now concentrated on finding a boundary condition for this equation. Using Fourier transformation in the x -direction, we reduce (11) to an equation in (z, t) -space for each Fourier coefficient \tilde{p}_λ . Using Laplace transformation in time, we arrive at an ODE. We now consider a background atmosphere, in which the pressure and the density decrease exponentially with z

$$\dot{\rho}(z) = (\theta + 1)^{-\theta} a^{-\theta} e^{-2\alpha\theta z}, \quad \theta := (\gamma - 1)^{-1},$$

where a, α are some positive constants. For this setting the exact solution to the ODE consists of Bessel functions. We then find a first order hyperbolic equation for the pressure perturbation with a convolution term:

$$e^{\alpha z} \sqrt{a\theta} \partial_t \tilde{p}_\lambda + \partial_z \tilde{p}_\lambda + \left(\alpha\theta + \frac{\alpha}{2} \right) \tilde{p}_\lambda - (\mathcal{L}^{-1} A_\lambda) * \tilde{p}_\lambda = 0. \quad (12)$$

The convolution term

$$((\mathcal{L}^{-1} A_\lambda) * \tilde{p}_\lambda)(t, z) := \int_0^t (\mathcal{L}^{-1} A_\lambda)(t - \tau, z) \tilde{p}_\lambda(\tau, z) d\tau$$

is responsible for the nonlocal character of the boundary condition. Its kernel is given by the inverse Laplace transformation (\mathcal{L}^{-1}) of a complex function A_λ , which contains the Bessel functions from the exact solution of the ODE. The full derivation of the equations can be found in (Dedner et al., 2000).

The equation for the pressure perturbation (12), together with the equations (6), (7), (8) and (10), are now used as top boundary conditions for the full MHD equations in Ω_{comp} .

2. Implementational aspects

The presented test calculations for (1) — (5) were performed on a regular cartesian grid. We use an explicit finite volume scheme with the approximate Riemann solver presented in (Dai and Woodward, 1995). To calculate the values in the ghost cells at the top boundary, we first use the Fourier transformation w.r.t. x of the discrete pressure perturbation. Then we solve (12) for a finite number of Fourier coefficients separately. This leads to discrete values $\tilde{p}_{\lambda,h}$, from which we obtain the new values of \tilde{p} by means of the inverse Fourier transformation. Then we calculate the other quantities $\tilde{\rho}, \tilde{u}, \tilde{v}, \tilde{w}, \tilde{\mathbf{B}}$ at the boundary by approximating (6) — (10) with second order finite differences.

To solve equation (12) we use a second order finite difference scheme for the differential terms; the main problem lies in the discretization of the

convolution term. Two problems have to be solved: In the first place, the kernel is an inverse Laplace transformation of a quite complicated function and therefore has to be approximated. Secondly, the equation (12) is non-local in time. We want to construct a scheme which is local in time; this can be achieved by using a special rational approximation to the function A_λ :

$$A_\lambda(s) \approx \sum_{i=1}^M \frac{a_i}{s + \alpha_i} =: A_{\lambda,M}(s).$$

Note that $A_{\lambda,M}(s)$ will depend on z . We calculate this approximation using the Chebyshev–Padé algorithm. This representation of A_λ enables us to calculate the inverse Laplace transformation analytically:

$$(\mathcal{L}^{-1} A_{\lambda,M})(t) = \left(\mathcal{L}^{-1} \left(\sum_{i=1}^M \frac{a_i}{s + \alpha_i} \right) \right)(t) = \sum_{i=1}^M a_i \exp(-\alpha_i t).$$

The coefficients a_i, α_i can be calculated in advance. Now we can evaluate the convolution operator at a time t^{n+1} through the sum of the approximation to the convolution at time t^n

$$\left((\mathcal{L}^{-1} A_{\lambda,M}) * \tilde{p}_{\lambda,h} \right) (t^n) = \sum_{i=1}^M a_i B_i^n,$$

where

$$B_i^n := \int_0^{t^n} \exp(-\alpha_i(t^n - t)) \tilde{p}_{\lambda,h}(t) dt,$$

and a term involving the values of the pressure perturbation between t^n and t^{n+1} :

$$\begin{aligned} & \left((\mathcal{L}^{-1} A_{\lambda,M}) * \tilde{p}_{\lambda,h} \right) (t^{n+1}) \\ &= \int_0^{t^{n+1}} (\mathcal{L}^{-1} A_{\lambda,M})(t^{n+1} - t) \tilde{p}_{\lambda,h}(t) dt \\ &= \sum_{i=1}^M a_i \left\{ \exp(-\alpha_i(t^{n+1} - t^n)) B_i^n \right. \\ & \quad \left. + \int_{t^n}^{t^{n+1}} \exp(-\alpha_i(t^{n+1} - t)) \tilde{p}_{\lambda,h}(t) dt \right\}. \end{aligned}$$

Thus we approximate the convolution using only data between t^n and t^{n+1} . For further details we refer to (Dedner et al., 2000).

3. Numerical results

To show the strength of our method we will concentrate on one test case (a Rayleigh–Taylor type instability), but will perform our calculations using three different types of boundary conditions. All of these boundary conditions are chosen in a way which will guarantee that the background atmosphere remains static during the simulation if no initial perturbation is prescribed.

- In the first method we use **Dirichlet boundary condition**. To fulfill the stated requirement, the values of the background atmosphere are prescribed:

$$\mathbf{U}(t) = \dot{\mathbf{U}} \quad \text{at the boundary}$$

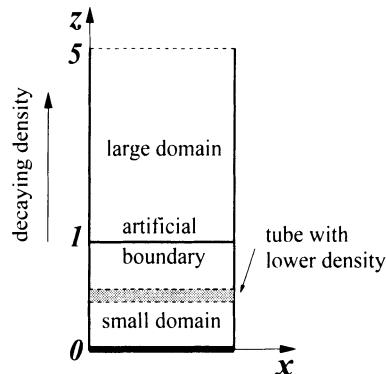
- The second approach (**normal boundary condition**) is a modified Neumann condition:

$$\partial_n(\mathbf{U}(t) - \dot{\mathbf{U}}) = 0 \quad \text{at the boundary}$$

(To maintain a static atmosphere it is not sufficient to prescribe vanishing normal derivatives.)

- The third approach are our **absorbing boundary conditions**.

Since no non-trivial analytical solution is known, we compare the performance of the different boundary conditions by first solving the whole problem on a large domain with **dbc** and a given grid spacing h . We then use this solution as a reference solution for the calculations with the different boundary conditions on a far smaller domain with the same grid spacing h . At a given height the density is lowered in a tube parallel to the x -axis; this region of lower density rises through the atmosphere. An additional sinusoidal perturbation is prescribed in the z -velocity, which is zero at the vertical boundaries and positive everywhere else. This leads to the development of the typical mushroom shape of the Rayleigh–Taylor instability.



We first show results in one space dimension using the values described above for $x = 0$ as initial conditions; in the tube we additionally prescribe a magnetic field in the x -component of B . The evolution of the density calculated on the large domain is shown at two different times in Figure 1 (left). A comparison of the different boundary conditions together with the reference solution is shown in Figure 1 (right). The **nbc** leads to a totally

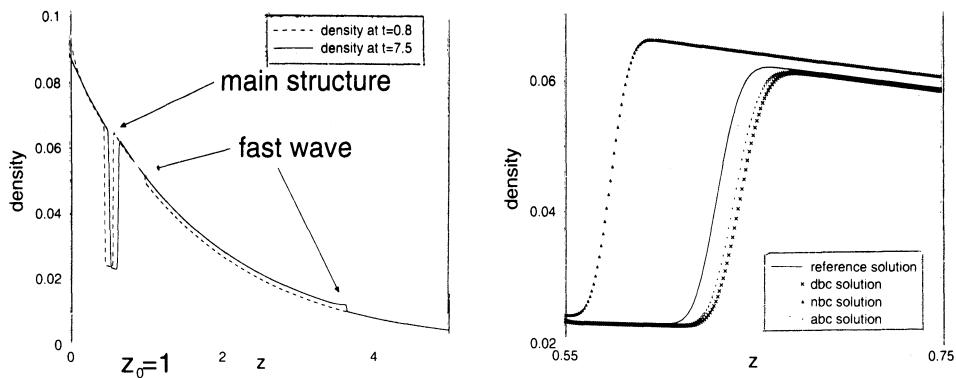


Figure 1. **Left:** Reference solution on large domain at two different times. **Right:** Comparison of three boundary conditions vs. reference solution at time $t = 7.5$.

wrong position of the lower density region. **dbc** and **abc** show more or less the same position of the tube, which is only slightly above the position of the tube in the reference solution.

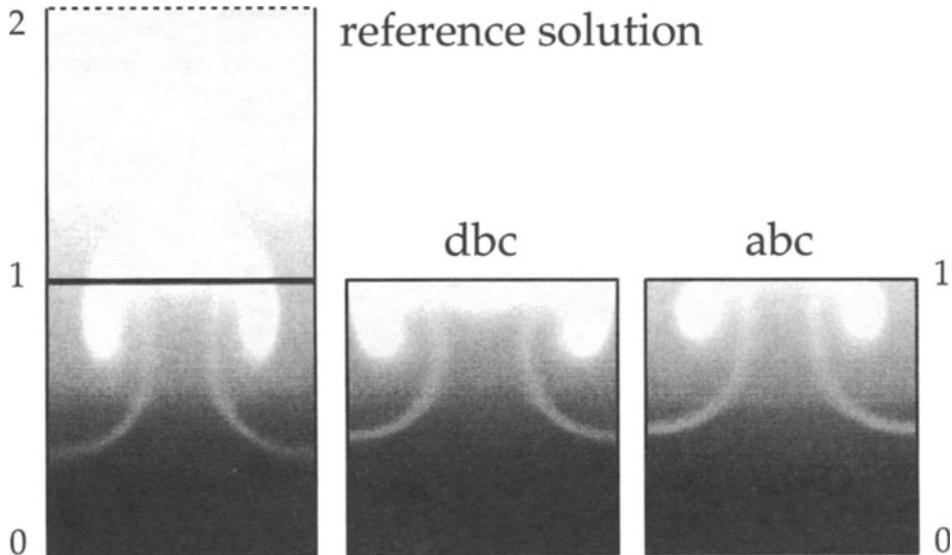


Figure 2. Comparison of the boundary condition in 2D.

In Figure 2 we have plotted the density of a calculation in two space dimensions with the initial conditions stated above, but without magnetic field. Since the results with the **nbc** show again a totally wrong position and shape of the tube, we have not included the outcome of that calculation. The

calculations with **dbc** do not lead to the correct shape of the instability (Figure 2 (middle)). Although major perturbations in the density have moved through the upper boundary, the results with the **abc** are still close to the reference solution. The shape of the instabilities is quite well resolved, only the position is again slightly too high (Figure 2 (right)).

ACKNOWLEDGEMENTS: The authors were partially supported by the DFG–Schwerpunktprogramm “Analysis und Numerik von Erhaltungsgleichungen” and by RFBR Grant No. 98-01-00177 and the University of Freiburg.

References

- Bell JB, Colella P, and Trangenstein JA (1989): Higher Order Godunov Methods for General Systems of Hyperbolic Conservation Laws. *J. Comput. Phys.*, **82**: 362.
- Briø M, and Wu CC (1988): An Upwind Differencing Scheme for the Equations of Ideal Magnetohydrodynamics. *J. Comput. Phys.*, **75**: 400.
- Dai W, Woodward PR (1995): A Simple Riemann Solver and High-Order Godunov Schemes for Hyperbolic Systems of Conservation Laws, *J. Comput. Phys.*, **121**: 51.
- Dedner A, Kröner D, Sofronov IL, Wesenberg M (2000): Transparent Boundary Conditions for MHD Simulations in Stratified Atmospheres, *Preprint*.
- Grote MJ, Keller JB (2000): Nonreflecting Boundary Conditions for Maxwell’s Equations, *J. Comput. Phys.*, **139**: 327.
- Kröner D (1991): Absorbing boundary conditions for the linearized Euler equations in 2-D, *Math. Comput.*, **57**: 153.
- Sofronov IL (1998): Non-reflecting inflow and outflow in a wind tunnel for transonic time-accurate simulation, *J. Math. Anal. Appl.*, **221**: 92.
- Wesenberg M (2000) A Note on MHD Riemann Solvers. *In preparation*.

ABOUT KINETIC SCHEMES BUILT IN AXISYMMETRICAL AND SPHERICAL GEOMETRIES

S. DELLACHERIE

Commissariat à l'Énergie Atomique

Direction du Cycle du Combustible

DPE/SPCP/LMEP

91191 Gif sur Yvette, France

Email: stephane.dellacherie@cea.fr

Abstract. In that paper, we show that the technic of kinetic schemes previously used in cartesian geometry to solve the compressible Euler equations can be naturally applied to axisymmetrical and spherical geometries. Indeed, we show that it is easy to build an explicit monodimensionnal axisymmetrical or spherical kinetic scheme which preserves the positivity of the density and of the internal energy under a CFL criteria.

1. Introduction

The french Atomic Energy Commission (CEA) develops an axisymmetrical code which simulates the vapor created by electron beam metal evaporation by solving the Boltzmann equation with a Monte Carlo method [see (Chatain, Gonella and Roblin, 1997)]. But, although the metal is almost everywhere a rarefied gas - we say that it is in a *kinetic area* -, there exists an area - which is called the *fluid area* - where the metal has to be considered as a fluid because the metal density is very high: then, the Monte Carlo method needs a lot of iterations to converge to a stationnary solution which induces a high CPU time. To diminish this CPU time, a possible way is to simulate the Boltzmann equation only in the kinetic area and the compressible Euler equations only in the fluid area. Up to now, some authors have developed such technics in the field of aerodynamics [see (Qiu, 1993)] and in the field of thermonuclear plasma [see (Dellacherie, 1998)]: they used the *kinetic schemes* to solve the Euler equations [see (Perthame, 1992)] which are a family of finite volume schemes designed for resolution of

the compressible Euler equations. These schemes allow to couple naturally a kinetic area with a fluid area. But they always developed this method in cartesian geometries. In that paper, we extend the kinetic schemes in axisymmetrical and spherical geometries. In the second section, we recall the kinetic formalism used to build a kinetic scheme. The third section is devoted to the monodimensionnal axisymmetrical kinetic scheme: we propose a stability result which proves that the axisymmetrical kinetic scheme preserves the positivity of the density and of the internal energy. In the fourth section, we show the connection between the kinetic axisymmetrical transport equation and the way we solve the monodimensionnal axisymmetrical Euler equations. In the fifth section, we give numerical results (Sod's tube and Noh's tube) and we conclude in the last section.

2. Preliminaries

The monodimensionnal cartesian, axisymmetrical or spherical compressible Euler equations are given by

$$\begin{cases} \partial_t(r^d\rho) + \partial_r(r^d\rho U) = 0, \\ \partial_t(r^d\rho U) + \partial_r(r^d\rho U^2) = -r^d\partial_r P, \\ \partial_t(r^d\rho\xi) + \partial_r[r^d(\rho\xi + P)U] = 0, \\ P \equiv \rho T/m = (\gamma - 1)\rho\varepsilon, \\ \xi = \frac{1}{2}U^2 + \varepsilon, \\ \gamma \in]1, 3]. \end{cases} \quad (1)$$

U , ξ and ε are respectively the radial velocity, the total energy and the internal energy, and we only consider a *perfect gas*. m is the atomic mass and $d \in \{0, 1, 2\}$ defines the geometry. Let us recall that the building of a kinetic scheme comes from the resolution of the transport equation

$$\partial_t f + \vec{v} \cdot \nabla_x f = 0 \quad (2)$$

where $f(t, \vec{x}, \vec{v})$ is a distribution function and where $\vec{v} \in \mathbb{R}^3$ is the microscopic velocity. For example, in monodimensionnal cartesian geometry ($x \equiv r$ and $v \in \mathbb{R}$), we use an upwind scheme to solve $\partial_t f + v\partial_x f = 0$ with an initial condition f_i^n given by (n is the time subscript and i is the space subscript) $f_i^n(v) = \frac{\rho_i^n/m}{\sqrt{T_i^n/m}} \chi\left(\frac{v-U_i^n}{\sqrt{T_i^n/m}}\right)$ where χ is a positive function such that

$$\int_{\mathbb{R}} (1, w^2) \chi(w) dw = (1, 1) \quad \text{and} \quad \chi(-w) = \chi(w). \quad (3)$$

Then, we obtain the numerical scheme

$$\begin{cases} \rho_i^{n+1} &= \rho_i^n - \frac{\Delta t}{\Delta x_i} [\mathfrak{S}_{i+1/2} - \mathfrak{S}_{i-1/2}], \\ U_i^{n+1} &= \frac{\Delta m_i^n}{\Delta m_{i+1}^{n+1}} U_i^n - \frac{\Delta t}{\Delta m_i^{n+1}} [\wp_{i+1/2} - \wp_{i-1/2}], \\ \xi_i^{n+1} &= \frac{\Delta m_i^n}{\Delta m_{i+1}^{n+1}} \xi_i^n - \frac{\Delta t}{\Delta m_i^{n+1}} [\mathfrak{N}_{i+1/2} - \mathfrak{N}_{i-1/2}] \end{cases}$$

($\Delta m_i^n \equiv \rho_i^n \Delta x_i$) where the numerical fluxes are defined by

$$\begin{pmatrix} \mathfrak{S} \\ \wp \\ \mathfrak{N} \end{pmatrix}_{i+1/2}^+ = \begin{pmatrix} \mathfrak{S} \\ \wp \\ \mathfrak{N} \end{pmatrix}_{i+1/2}^- + \begin{pmatrix} \mathfrak{S} \\ \wp \\ \mathfrak{N} \end{pmatrix}_{i+1/2}^- \quad (4)$$

with the half fluxes

$$\begin{pmatrix} \mathfrak{S} \\ \wp \\ \mathfrak{N} \end{pmatrix}_{i+1/2}^+ = \int_{v>0} mv \begin{pmatrix} 1 \\ v \\ \frac{v^2}{2} + \frac{\lambda(\gamma)}{2} \cdot \frac{T_i}{m} \end{pmatrix} f_i^n(v) dv \quad (5)$$

and

$$\begin{pmatrix} \mathfrak{S} \\ \wp \\ \mathfrak{N} \end{pmatrix}_{i+1/2}^- = \int_{v<0} mv \begin{pmatrix} 1 \\ v \\ \frac{v^2}{2} + \frac{\lambda(\gamma)}{2} \cdot \frac{T_{i+1}}{m} \end{pmatrix} f_{i+1}^n(v) dv \quad (6)$$

where $\lambda(\gamma) = \frac{3-\gamma}{\gamma-1}$ is the number of kinetic degrees of freedom minus the one due to the translation in the x direction. When the function χ has a *compact support*, it is easy to prove that the numerical cartesian scheme preserves the positivity of the density and of the internal energy under a classical CFL criteria. Then, we would like to use the strong properties of the monodimensionnal cartesian kinetic scheme to build a monodimensionnal axisymmetrical or spherical kinetic scheme having good stability properties. The difficulty is that, due to the term $r^d \partial_r P$, the system (1) is not conservative when $d \in \{1, 2\}$ and we can not directly apply the results obtained in cartesian geometry.

3. Positivity of the explicit monodimensionnal axisymmetrical kinetic scheme

The first stage is to write the system (1) in the equivalent form (now $d = 1$)

$$\begin{cases} \partial_t(r\rho) + \partial_r(r\rho U) = 0, \\ \partial_t(r\rho U) + \partial_r[r(\rho U^2 + P)] = P, \\ \partial_t(r\rho \xi) + \partial_r[r(\rho \xi + P)U] = 0 \end{cases} \quad (7)$$

to obtain the following numerical scheme:

$$\begin{cases} \rho_i^{n+1} &= \rho_i^n - \frac{\Delta t}{V_i} (r_{i+1/2} \mathfrak{J}_{r,i+1/2} - r_{i-1/2} \mathfrak{J}_{r,i-1/2}), \\ U_i^{n+1} &= \frac{\Delta m_i^n}{\Delta m_i^{n+1}} U_i^n - \frac{\Delta t}{\Delta m_i^{n+1}} (r_{i+1/2} \wp_{r,i+1/2} - r_{i-1/2} \wp_{r,i-1/2}) + \frac{\Delta t \Delta r_i}{\Delta m_i^{n+1}} P_i^n, \\ \xi_i^{n+1} &= \frac{\Delta m_i^n}{\Delta m_i^{n+1}} \xi_i^n - \frac{\Delta t}{\Delta m_i^{n+1}} (r_{i+1/2} \aleph_{r,i+1/2} - r_{i-1/2} \aleph_{r,i-1/2}) \end{cases} \quad (8)$$

with $V_i = r_i \Delta r_i$ and $\Delta m_i^n = \rho_i^n V_i$. The radial fluxes \mathfrak{J}_r , \wp_r and \aleph_r are defined by the relations (4), (5) and (6). We have the following result:

Theorem 1 *Let us suppose that χ has a compact support i.e. $\exists X > 0 / |x| > X \Rightarrow \chi(x) = 0$ and let us take $\gamma \in]1, 2]$. Then, under the CFL criteria $\Delta t < \min(\Delta t_1^n, \Delta t_2^n)$ where*

$$\Delta t_1^n = \frac{\min_i \Delta r_i / 4}{\max_i (|U_i^n| + \frac{X}{\sqrt{\gamma}} C_i^n)} \quad (9)$$

and where

$$\Delta t_2^n = \frac{1}{\gamma - 1} \cdot \frac{\Delta r_1 / 2}{\max_i \left\{ [1 + \text{sgn}(U_i^n)] |U_i^n| + \sqrt{\frac{2}{\gamma(\gamma-1)}} C_i^n \right\}}, \quad (10)$$

the numerical scheme defined with (8) preserves the positivity of the density and of the internal energy.

Let us note that C_i^n is the sound velocity given by $C_i^n \equiv \sqrt{\frac{\gamma P_i^n}{\rho_i^n}}$ and that $\text{sgn}(x) = x/|x|$ if $x \neq 0$ (we define $\text{sgn}(0) = 0$). We recall that $i = 1$ refers to the central mesh.

Moreover, we have $\gamma \in]1, 2]$: the reason is that, due to the axisymmetrical geometry, the number of degrees of freedom is at least equal to 2 and we can not have $\gamma \in]2, 3]$ from a physical point of view (see also the following section). The CFL criteria (10) is due to the source term $(0, P, 0)$ included in the system (7). Although the theorem 1 is written in axisymmetrical geometry, let us remark that it is possible to find a similar result in spherical geometry.

We will see that the proof of the theorem 1 can be easily applied to other Euler equations having a source term. For example, we can improve the stability result proposed in (Coquel et. al., 1997) where it is simulated monodimensionnal cartesian Euler diphasic equations ($x \equiv r$ in that case) with a kinetic scheme. We could verify with the under-mentioned technic that the CFL criteria due to the source term $(0, P \partial_x \alpha, 0)$ appearing in the gas phase equations (α is the volume fraction of the gas phase) would be

$$\Delta t < \frac{1}{\gamma - 1} \cdot \min_i \left\{ \frac{\alpha_i^n}{|\partial_x \alpha_i^n|} \cdot \frac{1}{[1 + \text{sgn}(U_i^n \cdot \partial_x \alpha_i^n)] |U_i^n| + \sqrt{\frac{2}{\gamma(\gamma-1)}} C_i^n} \right\}.$$

Proof of the theorem 1. Let us consider the two numerical schemes

$$\begin{cases} \widehat{\rho}_i &= \rho_i^n - \frac{2\Delta t}{V_i} (r_{i+1/2} \mathfrak{S}_{r,i+1/2} - r_{i-1/2} \mathfrak{S}_{r,i-1/2}), \\ \widehat{U}_i &= \frac{\Delta m_i^n}{\Delta m_i} U_i^n - \frac{2\Delta t}{\Delta m_i} (r_{i+1/2} \varphi_{r,i+1/2} - r_{i-1/2} \varphi_{r,i-1/2}), \\ \widehat{\xi}_i &= \frac{\Delta m_i^n}{\Delta m_i} \xi_i^n - \frac{2\Delta t}{\Delta m_i} (r_{i+1/2} \mathfrak{N}_{r,i+1/2} - r_{i-1/2} \mathfrak{N}_{r,i-1/2}) \end{cases} \quad (11)$$

and

$$\begin{cases} \widetilde{\rho}_i &= \rho_i^n, \\ \widetilde{U}_i &= U_i^n + \frac{2\Delta t \Delta r_i}{\Delta m_i^n} P_i^n, \\ \widetilde{\xi}_i &= \xi_i^n. \end{cases} \quad (12)$$

Then, it is easy to see that it is possible to write the explicit numerical scheme (8) in the following way

$$\begin{cases} \rho_i^{n+1} &= \frac{1}{2} \widehat{\rho}_i + \frac{1}{2} \widetilde{\rho}_i, \\ \Delta m_i^{n+1} U_i^{n+1} &= \frac{\Delta m_i}{2} \widehat{U}_i + \frac{\Delta m_i}{2} \widetilde{U}_i, \\ \Delta m_i^{n+1} \xi_i^{n+1} &= \frac{\Delta m_i}{2} \widehat{\xi}_i + \frac{\Delta m_i}{2} \widetilde{\xi}_i. \end{cases}$$

In other words, $(\rho_i^{n+1}, \rho_i^{n+1} U_i^{n+1}, \rho_i^{n+1} \xi_i^{n+1})$ is a convex combinaison of $(\widehat{\rho}_i, \widehat{\rho}_i \widehat{U}_i, \widehat{\rho}_i \widehat{\xi}_i)$ and of $(\widetilde{\rho}_i, \widetilde{\rho}_i \widetilde{U}_i, \widetilde{\rho}_i \widetilde{\xi}_i)$. Knowing that

$$\left\{ (\rho, \rho U, \rho \xi) \text{ such that } \rho > 0 \text{ and } \xi - \frac{1}{2} U^2 > 0 \right\}$$

is a convex cone [see lemma 2 in (Gressier, Villedieu and Moschetta, 1999)], we deduce that if $\widehat{\rho}_i > 0$, $\widehat{\xi}_i > 0$, $\widetilde{\rho}_i > 0$ and $\widetilde{\xi}_i > 0$ then $\rho_i^{n+1} > 0$ and $\xi_i^{n+1} > 0$. By applying the classical technic used to prove the positivity of standard kinetic scheme, we show that if $\Delta t < \Delta t_1^n$ then $\widehat{\rho}_i > 0$ and $\widehat{\xi}_i > 0$. More over

$$\widetilde{\rho}_i > 0 \text{ and } \widetilde{\xi}_i > 0 \iff 2 \frac{(C_i^n)^4}{(\gamma r_i)^2} \Delta t^2 + 2 \Delta t \frac{U_i^n (C_i^n)^2}{\gamma r_i} - \frac{(C_i^n)^2}{\gamma(\gamma-1)} < 0$$

which is equivalent to write

$$\widetilde{\rho}_i > 0 \text{ and } \widetilde{\xi}_i > 0 \iff \Delta t < \frac{1}{\gamma-1} \cdot \frac{r_i}{\sqrt{(U_i^n)^2 + \frac{2(C_i^n)^2}{\gamma(\gamma-1)}} + U_i^n}$$

which easily gives the CFL criteria (10). \square

4. Kinetic interpretation of the scheme

The transport equation (2) is defined by

$$\partial_t(rf) + \partial_r[v_r(rf)] + \partial_{v_r} \left[\frac{v_\theta^2}{r}(rf) \right] - \partial_{v_\theta} \left[\frac{v_r v_\theta}{r}(rf) \right] = 0 \quad (13)$$

in monodimensionnal axisymmetrical geometry where v_r is the kinetic radial velocity and where v_θ is the kinetic orthoradial velocity. We can also write this equation in the following way:

$$\partial_t(rf) + \partial_r[v_r(rf)] = Q(f) \quad (14)$$

with $Q(f) = -v_\theta^2 \partial_{v_r} f + v_r \partial_{v_\theta} (v_\theta f)$ which can be seen as a collision operator in the phase space.

By integrating (14) with the measure $m(1, v_r, \frac{1}{2}(v_r^2 + v_\theta^2))dv_r dv_\theta$ and by taking

$$f(t, r, v_r, v_\theta) = \frac{\rho(t, r)/m}{2\pi P(t, r)/\rho(t, r)} \exp \left\{ -\frac{v_\theta^2 + (v_r - U(t, r))^2}{2P(t, r)/\rho(t, r)} \right\},$$

we obtain the Euler system (7) and we verify that the source terme $(0, P, 0)$ is only due to $Q(f)$. Then, the axisymmetrical kinetic scheme (8) can be seen as a splitting between the pure transport part of (13)

$$\partial_t(rf) + \partial_r[v_r(rf)] = 0 \quad (15)$$

and the local collision equation

$$\partial_t(rf) = Q(f). \quad (16)$$

Let us note that the proof of the theorem 1 is also based on this splitting: see the decomposition (11) and (12).

Of course, we obtain a fully conservative scheme for the Euler equation in the sense that the first and the third equations in (8) are conservative. Moreover, it is easy to verify that the uniform state $(\rho, U = 0, P)$ is preserved. From a kinetic point of view, it is difficult to verify these properties by preserving the positivity of f at the same time when the equation (13) is solved [see (Mieussens, 1999)]. Here, we easily obtain these properties because, *from a kinetic point of view*, we solve the transport part (15) but we just take into account the macroscopic effect of the collision part (16). This technic is also justify for the kinetic-fluid coupling because the boundary limit at the kinetic-fluid interface is taking into account just in the divergence part of the Euler equations (7) which is only due to the pure transport equation (15): see (Dellacherie, 1999).

5. Numerical results

Now, we propose some classical numerical tests to show the good behaviour of the monodimensionnal axisymmetrical and spherical kinetic schemes. We take $\chi(v) = \frac{1}{\sqrt{2\pi}} \exp(-v^2/2)$ [let us remark that χ has not a compact

support; but (Estivalezes and Villedieu, 1996) have prove that, in that particular case, the positivity of the monodimensionnal cartesian kinetic scheme is still verified; see also (Gressier, Villedieu and Moschetta, 1999)]. In what follows, CFL is a positive constant such that $\Delta t < 4CFL\Delta t_1^n$ where Δt_1^n is given by (9). More over, Δr_i is constant and the number of meshes is equal to 100. We will see on the following numerical tests that the axisymmetrical and spherical kinetic schemes have the same behaviour than the cartesian kinetic scheme i.e. they are very robust (cf. Noh's tube) but with a lack of precision in contact discontinuities (cf. Sod's tube).

Sod's tube: The initial conditions are classical. We take $\gamma = 1.4$ and $CFL = 0.9$. The results (figures 1 and 2) are presented when the time $t = 0.14$.

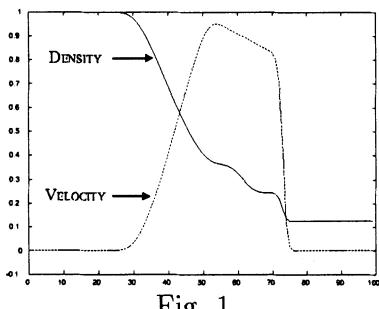


Fig. 1

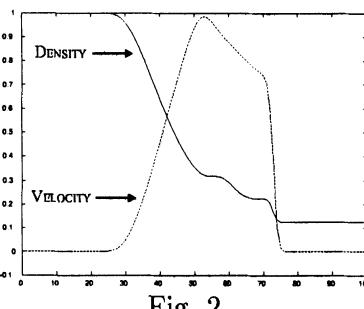


Fig. 2

Sod's tube in axisymmetrical and spherical geometry
Density and Velocity

Noh's tube: The initial conditions are the following: $\rho(r, t = 0) = 1$, $P(r, t = 0) = 0$ and $U(r, t = 0) = -1$. We take $\gamma = 5/3$ and $CFL = 0.5$ (in axisymmetrical geometry, we also show results with $CFL = 0.9$). This numerical test was proposed by W.F. Noh in (Noh, 1987): it is a hard numerical test because it appears a very strong centrifugal shock in the center of the cylinder or of the sphere. The results (figures 3 and 4) are presented when the time $t = 0.75$. The numerical error in the center is a classical phenomena called *wall heating* [cf. (Noh, 1987)].

6. Conclusions

In that paper, we have studied and tested monodimensionnal axisymmetrical and spherical kinetic schemes. In (Dellacherie, 1999), we verify that the bidimensionnal axisymmetrical kinetic scheme deduced from the monodimensionnal axisymmetrical kinetic scheme gives also good numerical results. This suggests that it would be possible to use this scheme to couple a dense area with a kinetic area in bidimensionnal axisymmetrical geometry.

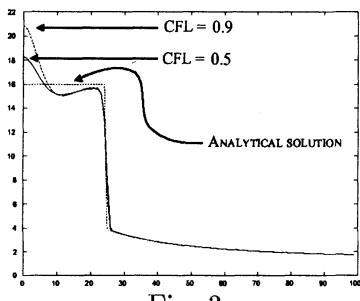


Fig. 3

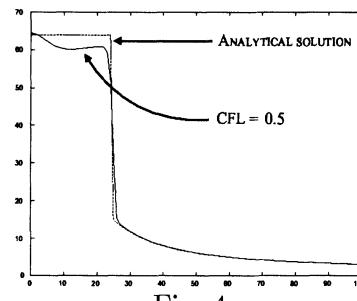


Fig. 4

Noh's tube in axisymmetrical and spherical geometry
Density

Acknowledgments

I wish to thank Bruno Dubroca (CEA-CESTA) who talked me about the crucial convexity property using to prove in a very easy way the positivity of the monodimensionnal axisymmetrical kinetic scheme under a CFL criteria.

References

- Chatain S., Gonella C. and Roblin P. (1997). Experimental results and Monte-Carlo simulations for a gadolinium atomic beam produced with a focused electron gun. *J. Phys. D: Appl. Phys.*, **30**, p. 360-367.
- Qiu Y. (1993). Étude des Équations d'Euler et de Boltzmann et de leur Couplage. Application à la Simulation Numérique d'Écoulements de Gaz Raréfiés. *Ph.D. Thesis* of Pierre and Marie Curie University (Paris VI).
- Dellacherie S. (1998). Contribution à l'analyse et à la simulation numériques des équations cinétiques décrivant un plasma chaud, part 2. *Ph.D. Thesis* of Denis Diderot University (Paris VII).
- Dellacherie S. (1999). Extension des schémas cinétiques en géométries axisymétrique et sphérique. Couplage des équations d'Euler et de Boltzmann. *Internal report CEA, DCC/DPE/99-RT-148*.
- Perthame B. (1992). Second-order Boltzmann Schemes for Compressible Euler Equations in One and Two Space Dimensions. *SIAM Journal of Numerical Analysis*, **29**, n° 1, p. 1-19.
- Estivalezes J.L. and Villedieu P. (1996). Higher order positivity preserving kinetic schemes for the compressible Euler equation: see the theorem 3.1. *SIAM journal of Numerical Analysis*, **33**, n° 5, p. 2050-2067.
- Gressier J., Villedieu P. and Moschetta J.M. (1999). Positivity of flux vector splitting schemes. *Journal of Computational Physics*, **155**, p. 199-220.
- Coquel F., Amine K.El , Godlewski E., Perthame B. and Rascle P. (1997). A Numerical Method Using Upwind Schemes for the Resolution of Two-Phase Flows. *Journal of Computational Physics*, **136**, p. 272-288.
- Mieussens L. (1999). Modèles à vitesses discrètes et méthodes numériques pour l'équation de Boltzmann-BGK, chapter 3. *Ph.D. Thesis* of Bordeaux I University.
- Noh W.F. (1987). Errors for calculations of strong shocks using an artificial viscosity and an artificial heat flux. *Journal of Computational Physics*, **72**, p. 78-120.

LAGRANGIAN SYSTEMS OF CONSERVATION LAWS AND APPROXIMATE RIEMANN SOLVERS

BRUNO DESPRÉS

*Commissariat à l'Energie Atomique, BP 12, 91 680 Bruyères le Chatel, France, and Laboratoire d'analyse numérique, 4 place Jussieu, Université Pierre et Marie Curie, 75252, Paris, France,
e-mail: despres@bruyeres.cea.fr and despres@ann.jussieu.fr*

Abstract. We sum up some new results concerning the mathematical structure of Lagrangian systems of conservation laws. These results include the assumption of Galilean invariance in the hypothesis and extend the classical symmetrization theorem of S. K. Godunov. Based on this representation theorem it is straightforward to derive numerical schemes which are non linearly stable, due to an entropy inequality satisfied by the scheme. We present a numerical application to an ICF-like calculation : the so-called T_i, T_e model for a ionized gas with ionic and electronic temperatures.

1. Introduction and theoretical results

Most of the motivations of that work were inspired by S. K. Godunov's work (as presented and resumed in (Godunov, 1999)).

We present some new results (Després, 1999)-(Després, Submitted to Numer. Math.) concerning the mathematical structure and related numerical issues of 1D Lagrangian systems of conservation laws

$$D_t U + D_x f(U) = 0. \quad (1)$$

The size of the system is m

$$U = \begin{pmatrix} U_1 \\ U_2 \\ \dots \\ U_m \end{pmatrix} \in R^m, \quad \text{and } f(U) = \begin{pmatrix} f_1(U) \\ f_2(U) \\ \dots \\ f_m(U) \end{pmatrix} \in R^m.$$

We assume that there exists an entropy $\mathcal{E}(U) \in R$, where $\mathcal{E}(U)$ is C^2 and strictly convex : i.e. the second derivatives matrix is symmetric and non negative

$$\frac{\partial^2 \mathcal{E}}{\partial U^2} > 0. \quad (2)$$

We add the fundamental hypothesis that the entropy flux is 0

$$\frac{\partial \mathcal{E}^t}{\partial U} \cdot \frac{\partial f}{\partial U} = 0, \quad \forall U \quad (3)$$

from which we deduce that smooth solutions of (1) satisfy

$$D_t \mathcal{E} = 0. \quad (4)$$

For discontinuous solutions (4) becomes

$$D_t \mathcal{E} \leq 0 \quad (5)$$

in the distributional sense. For sake of simplicity we restrict the study to the one dimensional case. However with few modifications, the results we present are true also for the multidimensional case. The simplest example corresponds to the Euler system (non viscous gas dynamics equation) in Lagrange coordinates

$$U = \begin{pmatrix} \tau \\ u \\ e \end{pmatrix}, \quad f(U) = \begin{pmatrix} -u \\ p \\ pu \end{pmatrix}. \quad (6)$$

The entropy is $\mathcal{E} = -S$ where S is the classical physical entropy.

Fluid models of plasma physics written in the Lagrange variable are particular cases of systems of conservation laws satisfying (1-3) : $D_t = \partial_t + u \partial_x$ is the derivative along the stream lines and $D_x = \frac{1}{\rho} \partial_x$ is the derivative in the mass variable. For example two-temperatures models (one ionic temperature and one electronic temperature (Coquel, 1998)-(Zel'dovitch and Raizer, 1967)-(Jaouen, 1997)), one dimensional ideal magneto-hydrodynamics (Beard and Després, 1999)-(Powell, Roe P L and Myong, 1995), some models of radiation hydrodynamics and some elasto-plasticity models may be written in the form (1-3).

One of the first results concerning the structure of systems of conservation laws having an entropy is due to S. K. Godunov (Godunov, 1960),(Godunov, 1999), (Godunov, 1986). From the numerical point of view, the investigation of the structure of hyperbolic systems of conservation laws arising from the physics has always been a very active field of research. The (Von Neumann and Richtmyer, 1950) scheme uses the structure of the

gas dynamics system in non conservative form to design an artificial viscosity. The artificial viscosity allows to capture discontinuous solutions. The Godunov method (Godunov, 1959) uses the exact solution of the Cauchy problem with piecewise constant initial data. However this method has the drawback of being expensive for an arbitrary equation of state or a tabulated equation of state, which is a limitation. The Roe method (Roe, 1981) (which we shall take as a representative of the family of flux-splitting methods (Van Leer, 1982)-(Godlewski and Raviart, 1996)) uses the so-called Roe matrix in order to define a Roe scheme for the numerical solution of any hyperbolic systems. Nevertheless the construction of the Roe matrix is not so easy a task (see (Powell, Roe P L and Myong, 1995)-(Gallice, 1995) for application to magneto-hydrodynamics) : moreover tricky analysis (Harten, Hyman and Lax, 1976) shows that the method may generate some non physical states for Eulerian gas dynamics ; this is also the case for the gas dynamics system written in the Lagrange variable. An up-to-date review of numerical schemes for Lagrangian gas dynamics is (Munz, 1994). New and promising results about relaxation schemes may be found in (Coquel and Perthame, 1998). All these issues turn to be strongly connected to the numerical computation of the solution of the (exact or linearized) Riemann problem.

In the following analysis we use the property that all 1D systems of conservation laws arising from the physics written in the Lagrange variable have a zero entropy flux. This allows to obtain a canonical representation of these systems. We are then able to derive numerical schemes which are entropy consistent under CFL condition. The analysis of both the system and the associated scheme uses ideas coming from thermodynamics : as an example the understanding of the relationships between the scheme and the resolution of the Riemann problem is achieved through the use of the enthalpy of the system. The enthalpy of the system is a generalization of the classical enthalpy given in standard thermodynamics textbooks (Callen, 1985)-(Bazarov I, 1994).

Theorem 1 *Systems of conservation laws with zero entropy flux (1-3), corresponding to fluid models, with galilean invariance and reversibility for smooth solutions are all of the form*

$$D_t U + D_x \begin{pmatrix} B\psi \\ -\frac{1}{2}\psi^t B\psi \end{pmatrix} = 0 \quad (7)$$

where $\psi \in R^{m-1}$ is derived from the primitive variables

$$\psi = \left(\frac{V_1}{V_m}, \frac{V_2}{V_m}, \dots, \frac{V_{m-1}}{V_m} \right)^t, \quad V = \frac{\partial \mathcal{E}}{\partial U} \quad (8)$$

and B is a constant symmetric matrix. The last component of U is the total energy $U_m = e$.

We refer to (Després, 1999) and (Després, Submitted to Numer. Math.) for a detailed presentation of the assumptions of that Theorem. This is a “representation theorem” for the flux. In some sense, the flux is no more viewed as a non-linear function of U . It is now viewed as a linear-quadratic function of ψ .

From the numerical point of view the previous result has the following major consequence. Instead of studying the Jacobian matrix $\frac{\partial f}{\partial U}$ which is a $m \times m$ non constant non symmetric matrix, it is sufficient to study the matrix B which is a constant $m - 1 \times m - 1$ symmetric matrix. Hence we easily derive a family of numerical schemes of order one in time and space

$$\frac{\Delta x}{\Delta t} (U_i^{n+1} - U_i^n) + \left(f(U)_{i+\frac{1}{2}} - f(U)_{i-\frac{1}{2}} \right) = 0 \quad (9)$$

where the numerical flux is obtain from a splitting of the symmetric matrix B in a symmetric positive part $B_{i+\frac{1}{2}}^+ = (B_{i+\frac{1}{2}}^+)^t \geq 0$ and a symmetric negative part $B_{i+\frac{1}{2}}^- = (B_{i+\frac{1}{2}}^-)^t \leq 0$

$$B = B_{i+\frac{1}{2}}^+ + B_{i+\frac{1}{2}}^-, \quad (10)$$

$$f(U)_{i+\frac{1}{2}} = \left(\left(B_{i+\frac{1}{2}}^+ \psi_{i+1}^n + B_{i+\frac{1}{2}}^- \psi_i^n \right), \quad (11) \right.$$

$$\left. \left(-\frac{1}{2} (\psi_{i+1}^n)^t B_{i+\frac{1}{2}}^+ \psi_{i+1}^n - \frac{1}{2} (\psi_i^n)^t B_{i+\frac{1}{2}}^- \psi_i^n \right) \right).$$

Theorem 2 *The scheme (9-11) is entropy consistent under a CFL condition (Després, Submitted to Numer. Math.), i.e. there exists constants $c_i^n > 0$ such that*

$$\text{if } c_i^n \frac{\Delta t}{\Delta x} \leq 1, \text{ then } \mathcal{E}(U_i^{n+1}) \leq \mathcal{E}(U_i^n). \quad (12)$$

This result may be interpreted in two ways.

- Inequality (12) is a non-linear stability estimate. For example for a perfect gas law, $\mathcal{E} = -S = -\log(\varepsilon\tau^{\gamma-1})$. So the decrease $\mathcal{E}(U_i^{n+1}) \leq \mathcal{E}(U_i^n)$ shows that the product $\varepsilon\tau^{\gamma-1}$ is an increasing function in each cell. It can be proved that it implies that both variable ε and τ remain positive in the cell during the time step.
- The inequality expresses the fact that the numerical solution follows a correct thermodynamic path in the thermodynamic diagram. Correct thermodynamic paths are characterised by the increase of the entropy, while wrong thermodynamic paths are characterised by the decrease of the entropy.

2. A simple numerical example

We give here some numerical results as an illustration of Theorems 1 and 2. Let us consider again the Euler equations (6). Even if this example is the simplest one, it contains all the structure common to systems satisfying the hypothesis of Theorem 1. An application to magneto-hydrodynamics is discussed in (Bezard and Després, 1999). Let us emphasize that in the following we do not advocate for one solver against the other : we just show how to apply the previous theory to obtain numerical schemes which are all entropy consistent. Straightforward calculations show that

$$U = \begin{pmatrix} \tau \\ u \\ e \end{pmatrix}, \quad V = \begin{pmatrix} -\frac{p}{T} \\ \frac{u}{T} \\ -\frac{1}{T} \end{pmatrix}, \quad \psi = \begin{pmatrix} p \\ -u \end{pmatrix}, \quad f(U) = \begin{pmatrix} B\psi \\ -\frac{1}{2}\psi^t B\psi \end{pmatrix}$$

with B set to

$$B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Let us choose a coefficient $(\rho^* c^*)_{i+\frac{1}{2}} > 0$. We split B into

$$B_{i+\frac{1}{2}}^+ = \begin{pmatrix} \frac{1}{2(\rho^* c^*)_{i+\frac{1}{2}}} & \frac{1}{2} \\ \frac{1}{2} & \frac{(\rho^* c^*)_{i+\frac{1}{2}}}{2} \end{pmatrix}$$

and

$$B_{i+\frac{1}{2}}^- = \begin{pmatrix} -\frac{1}{2(\rho^* c^*)_{i+\frac{1}{2}}} & \frac{1}{2} \\ \frac{1}{2} & -\frac{(\rho^* c^*)_{i+\frac{1}{2}}}{2} \end{pmatrix}.$$

The difference equations are

$$\frac{\Delta x}{\Delta t} \begin{pmatrix} \tau_i^{n+1} - \tau_i^n \\ u_i^{n+1} - u_i^n \\ e_i^{n+1} - e_i^n \end{pmatrix} + \begin{pmatrix} -u_{i+\frac{1}{2}}^* + u_{i-\frac{1}{2}}^* \\ p_{i+\frac{1}{2}}^* - p_{i-\frac{1}{2}}^* \\ (pu)_{i+\frac{1}{2}}^* - (pu)_{i-\frac{1}{2}}^* \end{pmatrix} = 0$$

where

$$\begin{aligned} \begin{pmatrix} -u_{i+\frac{1}{2}}^* \\ p_{i+\frac{1}{2}}^* \end{pmatrix} &= B_{i+\frac{1}{2}}^+ \psi_{i+1}^n + B_{i+\frac{1}{2}}^- \psi_i^n \\ &= \begin{pmatrix} \frac{1}{2(\rho^* c^*)_{i+\frac{1}{2}}} & \frac{1}{2} \\ \frac{1}{2} & \frac{(\rho^* c^*)_{i+\frac{1}{2}}}{2} \end{pmatrix} \begin{pmatrix} p_{i+1}^n \\ -u_{i+1}^n \end{pmatrix} \end{aligned}$$

$$+ \begin{pmatrix} -\frac{1}{2(\rho^* c^*)_{i+\frac{1}{2}}} & \frac{1}{2} \\ \frac{1}{2} & -\frac{(\rho^* c^*)_{i+\frac{1}{2}}}{2} \end{pmatrix} \begin{pmatrix} p_i^n \\ -u_i^n \end{pmatrix}$$

that is

$$\begin{cases} p_{i+\frac{1}{2}}^* = \frac{1}{2}(p_i^n + p_{i+1}^n) + \frac{(\rho^* c^*)_{i+\frac{1}{2}}}{2}(u_i^n - u_{i+1}^n) \\ u_{i+\frac{1}{2}}^* = \frac{1}{2}(u_i^n + u_{i+1}^n) + \frac{1}{2(\rho^* c^*)_{i+\frac{1}{2}}}(p_i^n - p_{i+1}^n) \end{cases}. \quad (13)$$

Of course this formula (or very similar ones) for the numerical fluxes has been introduced by many authors (Richtmyer and Morton, 1957), (Godunov, 1959), (Munz, 1994), (Després, 1997), (Godlewski and Raviart, 1996), (Godunov, 1999), (Toro, 1997). The difference is that the flux is now based on the general invariance properties of the law of mechanics expressed in Theorem 1. The values of the pressure and the velocity at the interfaces are arithmetic averages of the pressures and velocities on both sides plus a viscous term. For the last equation we check that

$$(pu)_{i+\frac{1}{2}}^* = p_{i+\frac{1}{2}}^* u_{i+\frac{1}{2}}^*$$

The fluxes (13) may be recovered through a standard analysis of the linearized Riemann invariants (Richtmyer and Morton, 1957), (Godunov, 1999). These linearized Riemann invariant may be written in a well known differential form

$$dp \pm (\rho c)du = 0$$

where (ρc) is the product of the density times the velocity of sound $(\rho c)^2 = -\frac{\partial p}{\partial T|E}$. We freeze that (ρc) (that is we assume it is more or less constant $(\rho c) = (\rho^* c^*)_{i+\frac{1}{2}}$). So the solution of the linearized Riemann problem between a left state τ_i^n, u_i^n, e_i^n and a right state $\tau_{i+1}^n, u_{i+1}^n, e_{i+1}^n$ gives an intermediate state referred to as p^*, u^* . The equations for these p^*, u^* are

$$\begin{cases} p^* + (\rho^* c^*)_{i+\frac{1}{2}} u^* = p_i^n + (\rho^* c^*)_{i+\frac{1}{2}} u_i^n \\ p^* - (\rho^* c^*)_{i+\frac{1}{2}} u^* = p_{i+1}^n - (\rho^* c^*)_{i+\frac{1}{2}} u_{i+1}^n \end{cases}$$

whose solution is exactly (13).

However many other splittings are available, some of them being exotic. One may choose to split B into

$$B_{i+\frac{1}{2}}^+ = \begin{pmatrix} \frac{1}{(\rho^* c^*)_{i+\frac{1}{2}}} & 1 \\ 1 & (\rho^* c^*)_{i+\frac{1}{2}} \end{pmatrix}$$

and

$$B_{i+\frac{1}{2}}^- = \begin{pmatrix} -\frac{1}{(\rho^* c^*)_{i+\frac{1}{2}}} & 0 \\ 0 & -(\rho^* c^*)_{i+\frac{1}{2}} \end{pmatrix}.$$

We end up to other formulas for the fluxes : (13) turns into

$$\begin{cases} p_{i+\frac{1}{2}}^* = p_{i+1}^n + (\rho^* c^*)_{i+\frac{1}{2}} (u_i^n - u_{i+1}^n) \\ u_{i+\frac{1}{2}}^* = u_{i+1}^n + \frac{1}{(\rho^* c^*)_{i+\frac{1}{2}}} (p_i^n - p_{i+1}^n) \end{cases}. \quad (14)$$

We know from Theorem 1 that these exotic fluxes provide an entropy consistent scheme which reveals to be non linearly stable (Després, 1997). Numerical results using the fluxes (13) and (14) are provided in figure 1 and 2 for the same test case. The test case is the Sod shock tube with a “perfect gas” equation of state ($\gamma = 1.4$) : the initial left state is $p_L = 1$, $\frac{1}{\tau_L} = \rho_L = 1$, $u_L = 0$ while the initial right state is $p_R = 0.1$, $\frac{1}{\tau_R} = \rho_R = 0.125$, $u_R = 0$. We used 1000 points in order to have “almost converged” solutions. The entropy is $\mathcal{E} = -\log(\varepsilon \tau^{\gamma-1})$. We have chosen not to move the mesh for the visualization of the numerical result, that is we plot the solution in the mass variable.

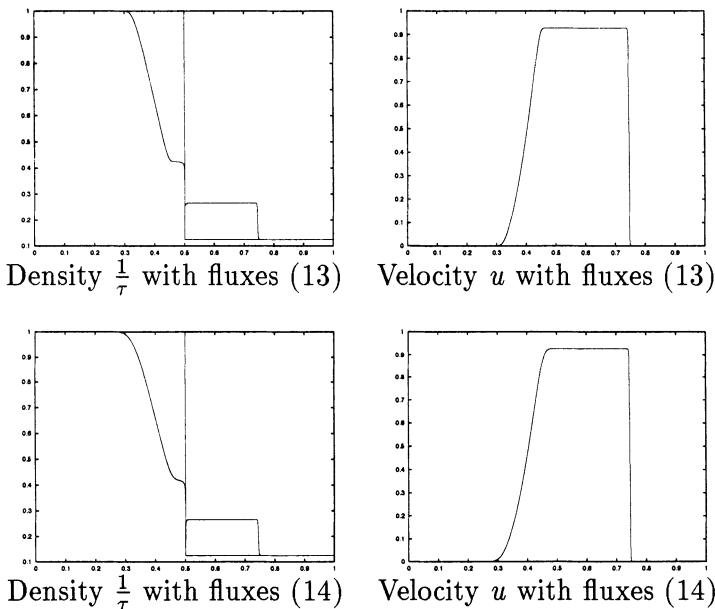


Figure 1. Sod shock tube at time $t = 0.14$

The “wall heating” phenomena (Noh, 1987) is visible at the contact discontinuity. With the exotic fluxes, that discrepancy is greater. However both schemes obtain exactly the same ”almost converged” solution. It is visible that the schemes are of order one in the rarefaction zone.

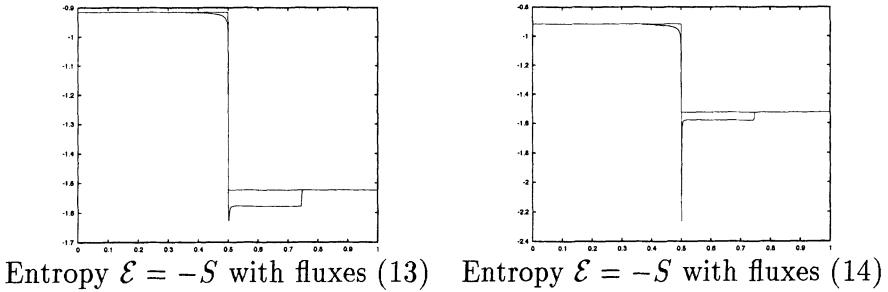


Figure 2. Sod shock tube at time $t = 0.14$

In fact it can be proved (Després, Submitted to Numer. Math.) that the constant c_i^n which appears in Theorem 2 may be approximated by the product of two constants

$$c_i^n \approx C(U_L, U_R, B^+, B^-) \times |\lambda_{\max}|$$

where $|\lambda_{\max}|$ is a local approximate value of the spectral radius of the Jacobian matrix and

$$C(U_L, U_R, B^+, B^-) \geq 1.$$

So it is possible to base the choice of (B^+, B^-) in order to minimize $C(U_L, U_R, B^+, B^-)$. For (13) one may prove (Després, 1999)

$$C(U_L, U_R, B^+, B^-) \approx \max \left(\frac{\rho c}{\rho^* c^*}, \frac{\rho^* c^*}{\rho c} \right) \geq 1.$$

Here ρc is an approximate local value of the product of the density times the velocity of sound, while $\rho^* c^*$ is the parameter which appears in the definition (13). This is another explanation of the optimal choice $(\rho^* c^*)_{i+\frac{1}{2}} = (\rho c)$: it gives $C(U_L, U_R, B^+, B^-) \approx 1$. For (14) the same analysis shows $C(U_L, U_R, B^+, B^-) > 1$ whatever $(\rho^* c^*)_{i+\frac{1}{2}}$ is. From the practical point of view, it seems clear that these non symmetric fluxes are useless : the CFL condition number with (14) is poorer than with (13), which is sufficient to prefer (13).

3. Application to the T_i, T_e model for ICF computation

In view of solving ICF oriented problems we introduce the non-conservative T_i, T_e model for ionized gas, in Euler coordinates

$$\begin{cases} \partial_t \rho + \partial_x \rho u = 0 \\ \partial_t \rho u + \partial_x (\rho u^2 + p_i + p_e) = 0 \\ \partial_t \rho \varepsilon_i + \partial_x \rho u \varepsilon_i + p_i \partial_x u = \frac{1}{\tau_{ei}} (T_e - T_i) \\ \partial_t \rho \varepsilon_e + \partial_x \rho u \varepsilon_e + p_e \partial_x u = \frac{1}{\tau_{ei}} (T_i - T_e) + \partial_x K_e \partial_x T_e \end{cases} \quad (15)$$

The density of internal ionic energy is ε_i and the density of internal electronic energy is ε_e . The total pressure is the sum of the ionic pressure $p_i = p_i(\rho, \varepsilon_i)$ and the electronic pressure $p_e = p_e(\rho, \varepsilon_e)$: $p = p_e + p_i$. For sake of simplicity we consider perfect gas laws $p_i = (\gamma_i - 1)\rho\varepsilon_i$ and $p_e = (\gamma_e - 1)\rho\varepsilon_e$. The ionic and electronic temperatures are $T_i = C_{vi}^{-1}\varepsilon_i$ and $T_e = C_{ve}^{-1}\varepsilon_e$. The relaxation time is $\tau_{ei} = \tau_{ei}(\rho, T_e, T_i) > 0$ and the diffusion coefficient is $K_e = K_e(\rho, T_e, T_i) > 0$.

The density of total energy $e = \varepsilon_i + \varepsilon_e + \frac{1}{2}u^2$ satisfies the conservative equation

$$\partial_t \rho e + \partial_x (\rho u e + p_i u + p_e u) = \partial_x K_e \partial_x T_e.$$

It is necessary to determine what is the meaning for discontinuous solutions of the last equation of (15) in order to get a "correct" solution. Through algebraic combinations of (15) and using the thermodynamic identity $T_e dS_e = d\varepsilon_e + p_e d\tau$, we get the "correct" mathematical expression

$$\partial_t \rho S_e + \partial_x \rho u S_e = \frac{1}{\tau_{ei} T_e} (T_i - T_e) + \frac{1}{T_e} \partial_x K_e \partial_x T_e$$

So the "almost" conservative form of (15) is

$$\begin{cases} \partial_t \rho + \partial_x \rho u = 0 \\ \partial_t \rho u + \partial_x (\rho u^2 + p_i + p_e) = 0 \\ \partial_t \rho S_e + \partial_x \rho u S_e = \frac{1}{\tau_{ei} T_e} (T_i - T_e) + \frac{1}{T_e} \partial_x K_e \partial_x T_e \\ \partial_t \rho e + \partial_x (\rho u e + p_i u + p_e u) = \partial_x K_e \partial_x T_e \end{cases} . \quad (16)$$

Note that $\frac{1}{T_e} \partial_x K_e \partial_x T_e$ makes sense (in the distributional sense) if we assume T_e is piecewise C^1 . We will see that (16) is compatible with the fact that the electronic part of the gas is adiabatic at discontinuities. Note that

In order to solve (16) we split it between its hyperbolic part and the right hand side. We study

$$\begin{cases} \partial_t \rho + \partial_x \rho u = 0 \\ \partial_t \rho u + \partial_x (\rho u^2 + p_i + p_e) = 0 \\ \partial_t \rho S_e + \partial_x \rho u S_e = 0 \\ \partial_t \rho e + \partial_x (\rho u e + p_i u + p_e u) = 0 \end{cases} \quad (17)$$

It is now clear that (17) expresses the fact that electronic part of the gas is adiabatic at discontinuities. We introduce the Lagrangian form of (17). Let us recall that $D_t = \partial_t + u\partial_x$ and $D_x = \frac{1}{\rho}\partial_x$.

$$\begin{cases} D_t \tau - D_x u = 0 \\ D_t u + D_x(p_i + p_e) = 0 \\ D_t S_e = 0 \\ D_t e + D_x(p_i u + p_e u) = 0 \end{cases} \quad (18)$$

The mathematical entropy of the system is $\mathcal{E} = -S_i$ where S_i is the physical ionic entropy (Jaouen, 1997) ($T_i dS_i = d\varepsilon_i + p_i d\tau$). The electronic entropy S_e is an unknown in the system. Theorem 1 applies and straightforward calculations show that

$$\psi = \begin{pmatrix} p = p_i + p_e \\ -T_e \\ -u \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

We deduce that the scheme (9-11) may be applied to (18) during one time step Δt with

$$B_{i+\frac{1}{2}}^+ = \begin{pmatrix} \frac{1}{2(\rho^* c^*)_{i+\frac{1}{2}}} & 0 & \frac{1}{2} \\ 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{(\rho^* c^*)_{i+\frac{1}{2}}}{2} \end{pmatrix}$$

and

$$B_{i+\frac{1}{2}}^- = \begin{pmatrix} -\frac{1}{2(\rho^* c^*)_{i+\frac{1}{2}}} & 0 & \frac{1}{2} \\ 0 & 0 & 0 \\ \frac{1}{2} & 0 & -\frac{(\rho^* c^*)_{i+\frac{1}{2}}}{2} \end{pmatrix}.$$

where $(\rho^* c^*)_{i+\frac{1}{2}}$ is some local average of $c^2 = \frac{\partial p}{\partial \rho|S} = \frac{\partial p_i}{\partial \rho|S_i} + \frac{\partial p_e}{\partial \rho|S_e} = c_i^2 + c_e^2$. Theorem 2 implies that $S_i^{n+1} \geq S_i^n$ in each cell, provided some CFL condition is satisfied.

In some sense solving (18) means that the mesh "moves" with the flow. So we remap the mesh using the interface velocities defined through $B_{i+\frac{1}{2}}^\pm$ (like in (13)). Finally we solve the right hand side of (15) using some implicit solver conservative in the total energy variable, that is we solve

$$\begin{cases} \partial_t \rho = 0 \\ \partial_t \rho u = 0 \\ \partial_t \rho \varepsilon_i = \frac{1}{\tau_{ei}}(T_e - T_i) \\ \partial_t \rho \varepsilon_e = \frac{1}{\tau_{ei}}(T_i - T_e) + \partial_x K_e \partial_x T_e \end{cases} \quad (19)$$

Note that the global scheme is entropic for arbitrary equations of state and constants τ_{ei} and K_e , since the Lagrangian step is entropic, the remap step is entropic and the relaxation-diffusion step (19) is entropic provided we use an implicit solver for (19).

The figure (3) shows that this procedure gives the same almost converged solution than the one produced by a code written in a non conservative formulation with the Von Neumann-Richtmyer pseudo-viscosity technique to handle discontinuities. Equations of state were perfect gas laws, while τ_{ei} and K_e were "real" tabulated coefficients. The calculation simulates the heating of a gas pushed on the right by a "piston". We see that the ionic part of the gas is violently heated, while the electronic part of the gas is better pre-heated by the electronic conduction than by the shock. Continuity of the electronic temperature is indeed satisfied: it is compatible with our remark about the mathematical meaning of $\frac{1}{T_e} \partial_x K_e \partial_x T_e$ in (16). Note that the "wall-heating" discrepancy (local increase of ionic temperature on the right) is much bigger with the Von Neumann-Richtmyer pseudo-viscosity technique.

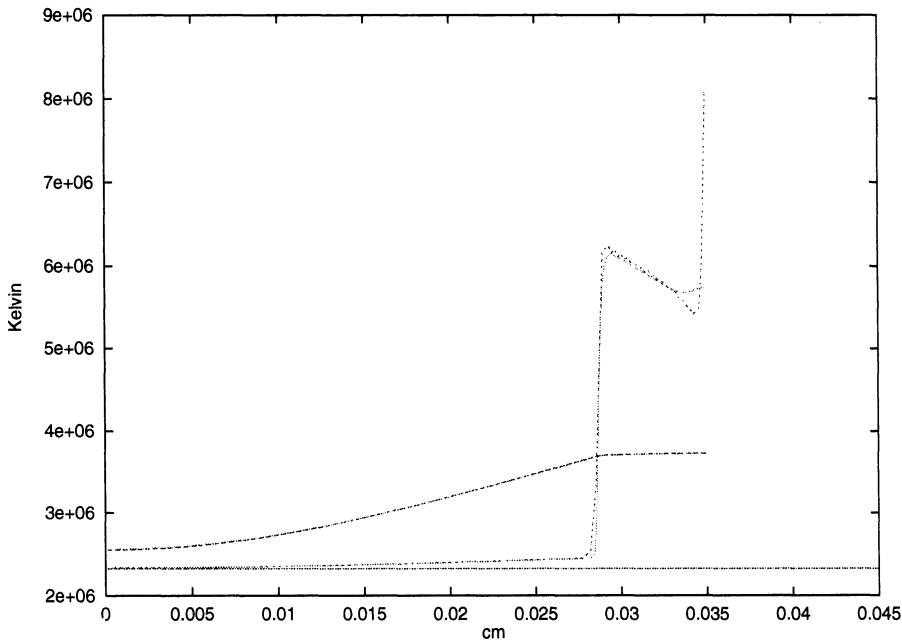


Figure 3. Values of the ionic ("shocked" curves) and electronic ("smooth" curves) temperatures for the VNR method and our method

We are now working on various multi-dimensional and high order ex-

tension of this family of schemes (Després, 1998), (Després, 1999), taking in account extra-physics as in (Després, Submitted to Numer. Math.), (Bezard and Després, 1999).

References

- Bazarov I (1989). *Thermodynamique*. Edition MIR (in french).
- Bezard F and Després B (1999). An entropic solver for ideal Lagrangian Magnetohydrodynamics. *Jour. of Comp. Phys.*
- Callen H B (1985). *Thermodynamics and introduction to thermostatistics*. John WILEY & SONS.
- Coquel F (1998). Schémas hyperboliques. INRIA-France-1998.
- Coquel F and Perthame B (1998). Relaxation of energy and approximate Riemann solvers for general pressure laws in fluid dynamics. *SIAM J. Numer. Anal.*, 35(6):2223–2249.
- Després B (Submitted to Numer. Math.). Structure of lagrangian systems of conservation laws, approximate Riemann solvers and the entropy condition .
- Després B (1997). Inégalité entropique pour un solveur conservatif du système de la dynamique des gaz en coordonnées de Lagrange. *C. R. Acad. Sci., Série I*, 324:1301–1306.
- Després B (1998). Entropy inequality for high order discontinuous Galerkin approximation of Euler equations. VII conference on hyperbolic problems, ETHZ-Zurich.
- Després B (1999). Discontinuous Galerkin Method for the numerical solution of Euler equations in axisymmetric geometry. DGM International Symposium, Newport.
- Després B (1999). Structure des systèmes de lois de conservation en variables Lagrangiennes . *Comptes Rendus de l'Académie des Sciences, Série I*, 328:721–724.
- Gallice G (1995). Résolution numérique des équations de la magnétohydrodynamique idéale. In GdR SPARCH, editor, *Proceedings of the conference*. Méthodes numériques pour la MHD.
- Godlewski E and Raviart P A (1996). *Numerical approximation of hyperbolic systems of conservation laws*. Number 118. Springer Verlag New York. AMS 118.
- Godunov S K (1986). Lois de conservation et intégrales d'énergie des équations hyperboliques. In *Nonlinear hyperbolic problems*, pages 135–149. Springer-Verlag. Lecture notes in mathematics.
- Godunov S K (1999). Reminiscences about difference schemes. *Journal of Computational Physics*, (153):6–25.
- Godunov S K (1959). A difference scheme for numerical computation of discontinuous solutions of equations of fluid dynamics. *Mat. Sb.*, 89:271–306.
- Godunov S K (1960). Sur la notion de solution généralisée. *DAN*, 134:1279–1282.
- Harten A, Hyman J M, and Lax P D (1976). On finite difference approximations and entropy conditions for shocks. *Comm. Pure Appl. Math.*, 29:297–322.
- Jaouen S (1997). Solveur entropique d'ordre élevé pour les équations de l'hydrodynamique à deux températures. Technical report, Université de Bordeaux. Mémoire de fin d'étude.
- Van Leer B (1982). Flux-vector splitting for the euler equations. Technical Report 82-30, ICASE.
- Munz D (1994). On Goudounov type schemes for Lagrangian formulations. *SIAM Journal of Numer. Anal.*, 31:17–42.
- Von Neumann J and Richtmyer R D (1950). A method for the numerical calculations of hydrodynamics shocks. *Journal of Applied Physics*, 21:232.
- Noh W F (1987). Errors for calculations of strong shocks using an artificial viscosity and an artificial flux. *Journal of Computational Physics*, 72:78–120.
- Powell K G, Roe P L, and Myong R S (1995). An upwind scheme for Magnetohydrodynamics. In GdR SPARCH, editor, *Proceedings of the conference*. Méthodes

- numériques pour la MHD.
- Richtmyer R D and Morton K W (1957). *Difference methods for initial-value problems*. Interscience Publishers.
- Roe P L (1981). Approximate Riemann solver, parameter vectors and difference schemes. *Journal of Computational Physics*, 43:357–372.
- Toro E F (1997). *Riemann Solvers and Numerical Methods for Fluid Dynamics*. Springer-Verlag.
- Zel'dovitch Y B and Raizer Y P (1967). *Physics of shock waves and High-temperature hydrodynamic phenomena*. Academic Press.

INTERMEDIATE SHOCKS IN 3D MHD BOW SHOCK FLOWS

H. DE STERCK AND S. POEDTS

*Centre for Plasma Astrophysics,
Celestijnenlaan 200B, 3001 Leuven, Belgium
Emails: hans.desterck@wis.kuleuven.ac.be
stefaan.poedts@wis.kuleuven.ac.be*

Abstract. Numerical simulations are presented of stationary magnetohydrodynamic (MHD) bow shock flows around perfectly conducting rigid spheres. When the upstream magnetic field is strong, several consecutive interacting shock fronts of different MHD shock type are needed to channel the flow around the obstacle. This new bow shock flow topology contains non-classical (overcompressive) intermediate shock segments. This seems to be the first confirmation in 3D of recent theoretical results on the admissibility of intermediate shocks for MHD flows with small dissipation.

1. New MHD bow shock topology for strong upstream magnetic field

Many phenomena in astrophysical and laboratory plasmas may be described by the magnetohydrodynamic (MHD) equations (De Sterck et. al., 1998). The non-strictly hyperbolic MHD system allows for three different wave modes, the fast, the Alfvén and the slow wave, with anisotropic wave speeds c_f , c_A and c_s , respectively. Three types of shocks are described by the MHD equations, connecting plasma states which are traditionally labeled from 1 to 4, with state 1 a super-fast state, state 2 sub-fast but super-Alfvénic, state 3 sub-Alfvénic but super-slow, and state 4 sub-slow. The fast 1–2 shock transition refracts the magnetic field away from the shock normal. A limiting case of the fast 1–2 shock is the 1–2=3 switch-on shock, for which the upstream magnetic field is parallel to the shock normal, while the magnetic field makes a nonzero angle with the shock normal in the downstream state. The tangential component of the magnetic field is thus switched on. Intermediate shocks (1–3, 1–4, 2–3 and 2–4) bring a

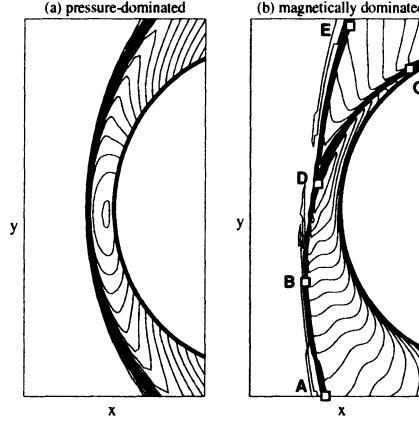


Figure 1. Bow shock flows over a sphere (thick solid). The flow comes in from the left. Density contours (thin solid) in a plane through the sphere center are shown. The incoming magnetic field is aligned with the horizontal x -axis. (a) Pressure-dominated flow: $M_{Ax} = 3.985$, $\beta = 0.4$, $\theta_{vB} = 5^\circ$. (b) Magnetically dominated flow: $M_{Ax} = 1.5$, $\beta = 0.4$, $\theta_{vB} = 3.8^\circ$.

super-Alfvénic upstream plasma to a sub-Alfvénic downstream state, while the magnetic field is flipped over the shock normal — the tangential component of the magnetic field changes sign. The slow 3–4 shock transition refracts the magnetic field towards the shock normal.

In this paper we report on the numerical simulation in three spatial dimensions (3D) of MHD bow shock flows around a perfectly conducting sphere (De Sterck, 1999; De Sterck and Poedts, 1999). A uniform superfast plasma flow falls in on the sphere and a stationary bow shock is formed. This problem has three free parameters, for which we choose the upstream plasma $\beta = 2p/B^2$ — with p the thermal pressure and \vec{B} the magnetic field —, the Alfvénic Mach number M_{Ax} along the upstream magnetic field lines — which are aligned with the x axis, see Fig. 1 —, and the angle θ_{vB} between the upstream velocity field \vec{v} and magnetic field. We simulate the 3D bow shock flows starting from a uniform initial condition and by advancing the time-dependent MHD equations with adiabatic index $\gamma = 5/3$ until a steady state solution is reached. We solve the MHD equations using a conservative finite volume high resolution Godunov shock capturing scheme which is second order accurate in space and time, employing a slope-limiter approach (De Sterck et al., 1998; De Sterck, 1999). The time-integration is explicit. For our present simulations, we use the Lax-Friedrichs numerical flux function which is simple and introduces a well-behaved numerical dissipation. As proposed by Powell

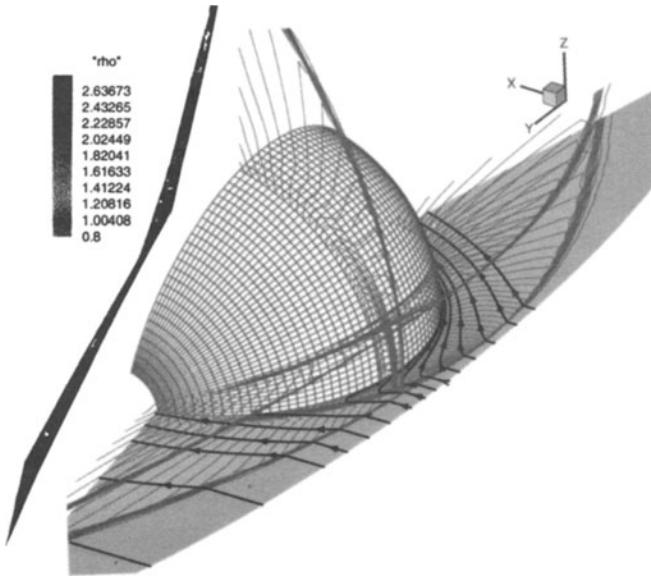


Figure 2. Magnetically dominated 3D bow shock flow around a sphere with inflow $\beta = 0.4$, $M_{Ax} = 1.49$, and $\theta_{vB} = 5^\circ$. Density contours and magnetic field lines are shown in the xy plane, which is a plane of symmetry parallel to the upstream magnetic field and velocity vectors and going through the center of the sphere. Density contours are also shown in two additional planes. In the upstream flow the magnetic field is aligned to the x -axis. ($40 \times 80 \times 40$ grid).

we control the $\nabla \cdot \vec{B}$ constraint by using a source term (Powell et. al., 1995). The simulations are performed on stretched polar-like structured grids.

Fig. 1a shows that for an upstream flow with a weak magnetic field — where it is specified below what weak means exactly — a bow shock flow is obtained with a single shock front. This is the classical bow shock topology which is well-known from hydrodynamic bow shocks, and which until now was believed to arise for all MHD bow shock flows as well. Fig. 1b, however, shows that for an upstream flow with a *strong* magnetic field the leading bow shock front is followed by a secondary shock front. Extensive simulations (De Sterck, 1999; De Sterck and Poedts, 1999) show that this previously unknown complex bow shock topology arises for all upstream flows which satisfy the criterion that

$$B^2 > \gamma p \quad (1)$$

and

$$\rho v_x^2 > B^2 > \rho v_x^2 \frac{\gamma - 1}{\gamma(1 - \beta) + 1}, \quad (2)$$

with v_x the velocity along the magnetic field and ρ the plasma density. These conditions are satisfied when the magnetic field strength is large

and thermal and dynamical pressure effects are dominated by magnetic effects. We call an upstream plasma which satisfies these conditions magnetically dominated, as opposed to pressure-dominated. Fig. 2 gives a 3D visualization of another bow shock flow with upstream parameters satisfying conditions (1) and (2). The leading shock front is clearly followed by a secondary shock front, which extends away from the xy plane.

The new results on MHD bow shock topology for strong upstream magnetic field are expected to have applications in space physics flows with shocks. In (De Sterck, 1999; De Sterck and Poedts, 1999) two applications are described, namely shocks induced by fast solar coronal mass ejections, and the earth's bow shock flow. There is some observational evidence for the occurrence of the new bow shock topology in these space physics flows.

2. Explanation in terms of the geometrical properties of MHD shocks

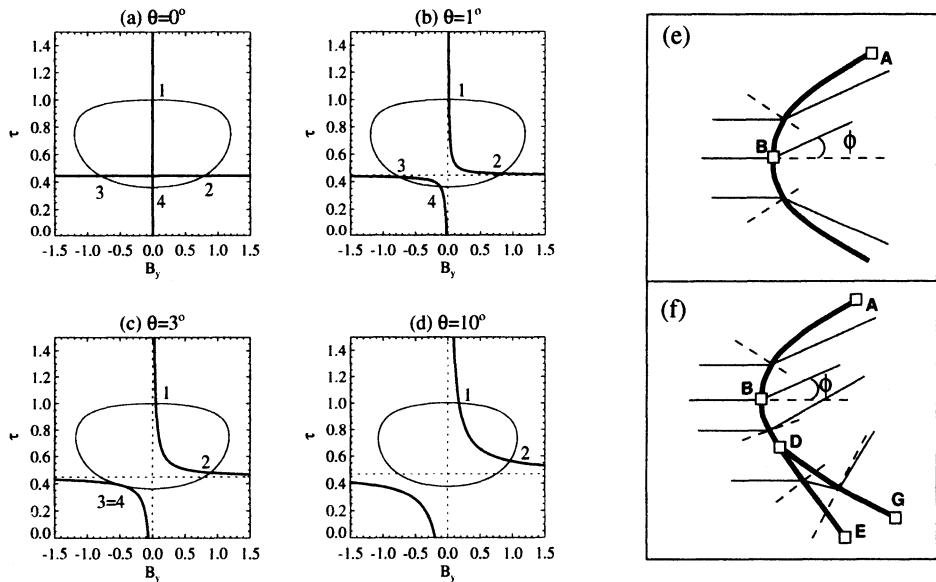


Figure 3. (a-d) Intersection points 1, 2, 3 and 4 are solutions of the MHD Rankine-Hugoniot relations for the upstream parameters of Fig. 1b (state 1) and for varying angle θ between the upstream magnetic field and the shock normal. The maximum angle θ for which 1-3 intermediate shocks can occur is approximately 3° . (e-f) Two proposed topologies for a shock front with an upper segment AB of 1-2 fast type for the case of a magnetically dominated upstream flow. Thick lines are shock fronts, thin lines are magnetic field lines, and shock normals are dashed. (e) The shock front cannot entirely be of the 1-2 fast type. (f) A complex shock topology is necessary to channel the flow.

The substantial difference between pressure-dominated and magnetically dominated MHD bow shock topologies can be explained in terms of the properties of MHD shocks (De Sterck, 1999; De Sterck et. al., 1998). This is illustrated in Fig. 3. We call a point on a shock front where the upstream magnetic field is perpendicular to the shock front a perpendicular point (e.g., point B in Fig. 3ef). Switch-on shocks and intermediate shocks only exist for certain parameter ranges of the upstream plasma. The conditions (1) and (2) are precisely the conditions under which switch-on shocks exist. For magnetically dominated upstream flows the shock at the perpendicular point B bordering the upper 1–2 shock segment AB is necessarily a switch-on shock with upward deflection. In this case the topology of Fig. 3e with a single shock entirely of fast type is impossible, because the lower fast shock segment, which deflects the magnetic field downwards, cannot be linked continuously to the switch-on shock at point B.

Instead, the complex topology of Fig. 3f arises, which comprises two consecutive shock fronts and shock segments of slow and intermediate type. Shock segment BD necessarily has to be of 1–3 intermediate type. The curved 1–3 intermediate shock segment BD can only have a limited extent, because for increasing angle θ between the magnetic field and the shock normal, the 1–3 shock first becomes a 1–3=4 shock and then ceases to exist (see Fig. 3a–d). At this point the leading shock front splits up into two consecutive shock fronts. The leading shock segment DE is of the 1–2 fast type, and the secondary shock segment DG is 2–4 intermediate, evolving into 3–4 slow along the front. This complex topology is obtained in the simulation of Fig. 1b. In contrast, for pressure-dominated flows the magnetic field is not refracted at a perpendicular point: the angle ϕ in Fig. 3e vanishes and the shock front can entirely be of fast shock type. This single-front topology is indeed obtained in the simulation of Fig. 1a.

3. Intermediate shocks in 3D MHD flows

The Mach number plots along cuts perpendicular to shock fronts in Fig. 4 show clearly that the shock segment DG is of 2–4 intermediate type (close to 2=3–4) at cut C1, and that BD is of 1–3 intermediate type (close to 1–3=4) at cut C3. The presence of these stationary intermediate shock fronts in our 3D MHD simulation results is relevant for the ongoing debate on the physical existence of intermediate shocks. In the 60s intermediate shocks were found to be unstable — in a certain peculiar sense, see (Wu, 1991; Freistuehler, 1998; Myong and Roe, 1997; De Sterck, 1999; Falle et. al., 1998) — in the ideal MHD system. However, recently it has been shown that intermediate shocks can be stable in the dissipative MHD system for wide ranges of the dissipation coefficients (Wu, 1991; Freistuehler, 1998;

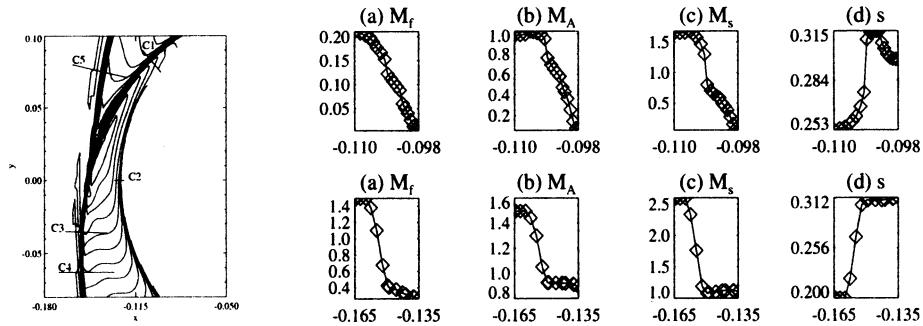


Figure 4. Normal Mach number and entropy plots along cuts perpendicular to various shock fronts for the simulation of Fig. 1b. Top row: cut along C1. Bottom row: cut along C3.

Myong and Roe, 1997). They can be destabilized by non-coplanar perturbations of the magnetic field, but only when the integrated amplitude of the perturbation is sufficiently large (Wu, 1991). The amplitude of the perturbation required for destabilization decreases with decreasing dissipation (Freistuehler, 1998). In our simulations, numerical dissipation plays the role of a small physical dissipation. Intermediate shocks have been found before in one-dimensional simulations of the dissipative MHD equations (Wu, 1991), but it seems that our simulations are the first confirmation in 3D that intermediate shocks can indeed exist and persist for small dissipation MHD in a realistic flow configuration.

References

- De Sterck H (1999). Numerical simulation and analysis of magnetically dominated MHD bow shock flows with applications in space physics. PhD thesis, K.U. Leuven, Belgium, and NCAR, Boulder, Colorado, USA.
- De Sterck H, Low B C, and Poedts S (1998). Complex magnetohydrodynamic bow shock topology in field-aligned low- β flow around a perfectly conducting cylinder. *Phys. Plasmas* **5**(11), pp 4015-4027.
- De Sterck H and Poedts S (1999). Stationary slow shocks in the magnetosheath for solar wind conditions with $\beta < 2/\gamma$: Three-dimensional MHD simulations. *J. Geophys. Res.*, **104**(A10), pp 22,401-22,406.
- Falle S A E G, Komissarov S S, and Joarder P (1998). A multidimensional upwind scheme for magnetohydrodynamics. *Mon. Not. R. Astron. Soc.* **297**, pp 265-277.
- Freistuehler H (1998). Small amplitude intermediate magnetohydrodynamic shock waves. *Phys. Scripta* **T74**, pp 26-29.
- Myong R S and Roe P L (1997). Shock waves and rarefaction waves in magnetohydrodynamics. Part 1. A model system. *J. Plasma Physics* **58**, pp 485-519.
- Powell K G, Roe P L, Myong R S, Gombosi T I, and De Zeeuw D L (1995). An upwind scheme for magnetohydrodynamics. *AIAA Paper 95-1704-CP*.
- Wu C C (1991). New theory of MHD shock waves. In *Viscous profiles and numerical methods for shock waves*, Siam Proceedings Series, pp 209-236.

A SECOND ORDER GODUNOV-TYPE SCHEME FOR NAVAL HYDRODYNAMICS

A. DI MASCIO, R. BROGLIA

*INSEAN, Istituto Nazionale per
Studi ed Esperienze di Architettura Navale,
Via di Vallerano 139, 00128 Roma, Italia
Email: a.dimascio@insean.it, r.broglia@insean.it*

AND

B. FAVINI

*Dipartimento di Meccanica e Aeronautica,
Università di Roma "La Sapienza",
Via Eudossiana 18, 00184 Roma, Italia
Email: favini@caspur.it*

Abstract.

A second order Godunov-type scheme for the simulation of free surface turbulent incompressible flows is presented. The scheme is applied to the RANSE written in pseudo-compressible formulation, whose asymptotic solution is computed by means of a Runge-Kutta time integration coupled with a multigrid algorithm. Examples of application of this scheme to the computation of the flow in a driven-cavity and past a surface-piercing hull are reported and, for the latter, compared with towing tank experiments carried out at INSEAN. Convergence properties when halving the grid size are also shown.

1. Introduction

Flows past ship hulls are characterised by very high Reynolds number regimes (typically $10^6 \sim 10^7$ for towing tank tests and up to 10^9 for full scale problems), complex geometry, and a great variety of physical phenomena like boundary layer thickening and separation, complex vortex structures in the field, evolution and breaking of free surface waves, cavitation, and

so on. Such a rich phenomenology imposes very stiff constraints on the development of numerical methods for the simulation of naval problems. On the basis of our previous experiences in this field, standard techniques, like centred schemes with artificial dissipation, have proved to be rather poor as to stability and convergence rate. Godunov schemes (Godunov, 1959) are eligible alternatives, possibly in ENO formulation (Harten et. al., 1987). Since the non-viscous problem for incompressible flows is not described by a hyperbolic system of equations, the Godunov approach can be adopted by following two different paths: by applying it only to the convection terms in the momentum equation, in the framework of the projection method, or by adopting a pseudo-compressible formulation, and then including also the continuity equation and the pressure term in the momentum equation. The pseudo-compressible approach turns out to be very effective when only the steady state solution has to be computed.

In the present paper, the pseudo-compressibility approach has been considered, and a novel scheme for the solution of the Riemann problem is proposed. ENO-type second order method is used for the convective and pressure part, which is hyperbolic, whereas the viscous terms are approximated by a centred finite volume scheme. Steady state convergence rate is improved by a standard Full-Multigrid technique, and a Runge-Kutta time integration with local time step is used as relaxation scheme.

The proposed scheme is applied to the computation of both 2D and 3D problems. Some simulations of the flow in a driven-cavity and past a ship with bulbous bow and transom stern are discussed.

2. Mathematical model

The free-surface incompressible turbulent flow past a rigid body can be described by the solution of the Reynolds averaged Navier-Stokes equations (RANSE). As well known, the major difficulty in solving these equations lies in the solenoidal constraint on the velocity vector field; when only the average steady state is to be computed, this system of equations can be conveniently replaced by the pseudo-compressible formulation (Chorin, 1967), that reads, in integral form:

$$\begin{aligned} \int_V \frac{\partial p}{\partial t} dV + \beta \int_S u_i n_i dS &= 0 \\ \int_V \frac{\partial u_i}{\partial t} dV + \int_S [u_i u_j n_j + p n_i - \tau_{ij} n_j] dS &= 0 \quad i = 1, 2, 3. \end{aligned} \tag{1}$$

A reference length l and velocity U_∞ have been chosen to make the equations non-dimensional. In the previous equations, u_i is the i -th cartesian component of the velocity vector; p is a new variable related to the pres-

sure P and the acceleration of gravity g by $p = P + z/Fr^2$, $Fr = U_\infty/\sqrt{gl}$ being the Froude number; $\tau_{ij} = \nu_t(u_{i,j} + u_{j,i})$ is the stress tensor, $\nu_t = 1/Re + \nu_T$ is the global kinematic viscosity, and ν_T is the turbulent viscosity. In the present work, the turbulent viscosity was calculated by means of the Baldwin-Lomax algebraic model (Baldwin and Lomax, 1978); β is the pseudo-compressibility factor.

As regards the boundary conditions, on solid walls the velocity is set to zero (whereas no condition on the pressure is required). At the free surface, two conditions are required. The first one is a dynamic boundary condition, which enforces continuity of the stress tensor through the surface; if the effect of air above the free surface and the surface tension are neglected, the dynamic boundary condition is:

$$p + \tau_{ij} n_j n_i = \frac{h}{Fr^2} \quad \text{and} \quad \tau_{ij} n_j t_i^l = 0, \quad (2)$$

where n_i is the normal unit vector, $t_i^l, l = 1, 2$ are two unit tangent vectors and $h = h(t, x, y)$ is the free surface elevation. The second boundary condition is a kinematic constraint:

$$\frac{\partial h}{\partial t} + u \frac{\partial h}{\partial x} + v \frac{\partial h}{\partial y} = w, \quad (3)$$

which states that the moving boundary is a material surface.

3. Integration Technique

The pseudo-compressible Navier-Stokes equations can be split in two parts, viscous terms and eulerian terms, the latter including convection and pressure. The convective operator defines a hyperbolic problem with a complete set of real eigenvalues and eigenvectors. By projecting the equations along the normal to the cell interface, the eulerian flux reads: $(\beta u_n; u_n^2 + p; u_{t^1} u_n; u_{t^2} u_n)^T$, u_n being the normal component of velocity and u_{t^1}, u_{t^2} two tangential components. The four eigenvalues of the Jacobian matrix associated to this flux vector are: $u_n \pm \sqrt{u_n^2 + \beta}$, u_n , u_n and the related characteristic equations are:

$$\partial p + (u_n \pm \sqrt{u_n^2 + \beta}) \partial u_n = 0, \quad \partial u_{t^1} = 0, \quad \partial u_{t^2} = 0. \quad (4)$$

A Riemann problem is associated to these equations, which is the building block of the algorithm. The value of pressure $p(x/t)$ and velocity $u_n(x/t)$ for $x/t = 0$ can be obtained by an approximate solution of this system of equations. The eigenvalues are supposed to be constant and are evaluated by averaging the normal component of the velocities of the right and left

states. Then, the values of $u_n(0)$ and $p(0)$ are calculated by solving an upwind finite difference approximation of the characteristic equations. This solution is used in the following, for evaluating the fluxes at cell interfaces.

The discrete RANSE equations are approximated by a finite volume technique, with pressure and velocity co-located at the cell centre. The residual on each control volume is computed as a flux balance at the cell interfaces, that implies the evaluation of pressure and velocity at the face centre. To this aim, a second order ENO-type scheme (Harten et. al., 1987) has been adopted for the non viscous part of the equations. In the particular case of second order accuracy, it can be shown that the left and right states, to be used for the Riemann problem at each cell face, have to be evaluated as:

$$\begin{aligned} \mathbf{q}_l = \mathbf{q}_{i,j,k} &+ \frac{1}{2} \text{minmod}(\Delta_{i-1/2}, \Delta_{i+1/2}) \\ \mathbf{q}_r = \mathbf{q}_{i+1,j,k} &- \frac{1}{2} \text{minmod}(\Delta_{i+1/2}, \Delta_{i+3/2}), \end{aligned}$$

where $\Delta_{i+1/2} = \mathbf{q}_{i+1,j,k} - \mathbf{q}_{i,j,k}$ and \mathbf{q} is the vector of the state variables. Once pressure and velocity at the cell interfaces are computed, the calculation of non-viscous fluxes easily follows. The computation of the viscous fluxes is made by standard centred finite volume approximation.

The discretization of the kinematic boundary condition (3) is made by an ENO finite difference approximation of the spatial derivatives of $h(t, x, y)$ along the curvilinear coordinate lines, the upwind bias being determined by the sign of the covariant velocities. As regards the boundary conditions for (3), the wave height is set to zero at the upstream boundary, is extrapolated at the outflow boundary, whereas the normal derivative is set to zero at solid boundaries, surface tension effects being neglected.

Time integration of the discrete model is achieved by means of a standard two-stage second order Runge–Kutta scheme (Jameson et. al., 1981). Both the bulk flow (governed by the RANSE) and the free-surface location (according to the kinematic boundary condition) are updated simultaneously at each time step. Steady state convergence is accelerated by means of local time stepping and a multigrid algorithm (for more details and performance of such a technique when applied to a time marching ENO–scheme, see (Favini et. al., 1996)).

In order to fit the actual position of the free surface, the numerical solution is computed on a moving grid, whose vertices run along the coordinate lines of a fixed underlying grid. CPU time is saved by updating the mesh and the metric terms only at the end of each multigrid cycle.

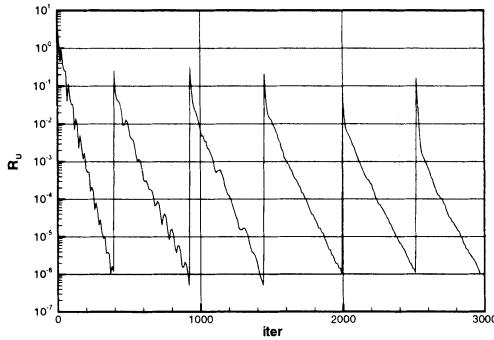


Figure 1. Residual history for u -velocity.

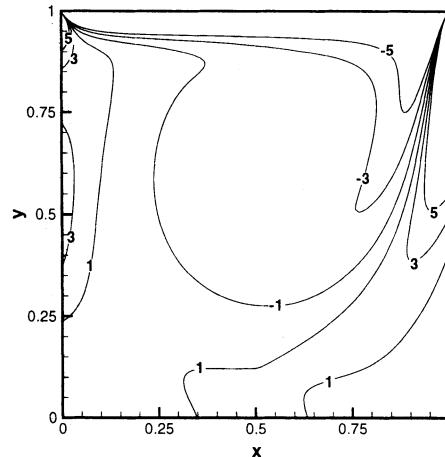


Figure 2. Fine Grid, isolines for vorticity; $\omega_{min} = -5.0$, $\omega_{max} = 5.0$, $\Delta\omega = 2.0$.

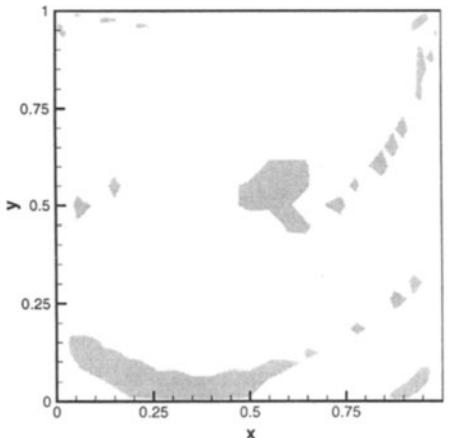


Figure 3. Observed convergence order.

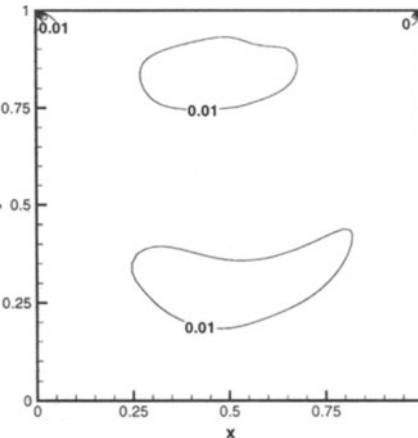


Figure 4. Isolines for the grid convergence index; $GCI_{min} = 0.00$, $GCI_{max} = 0.12$, $\Delta GCI = 0.01$.

4. Results

Convergence properties of the scheme have been measured by numerical experiments on driven cavity flows; some results are reported in figures 1–4 ($Re = 400$, 128×128 grid points for the finest mesh). Good convergence properties towards steady state and multigrid performance are shown in figure 1, where the residual for the x -component of momentum is presented (six grid level, FMG-cycle); as it can be noted, the residual decreases about

six orders of magnitude, with almost the same number of cycles on each grid. In figure 2, the solution on the finest grid in terms of vorticity isolines is presented; grid convergence properties are shown in figure 3 and 4 (see (Roache, 1997) for the definitions of the observed convergence order and of the Grid Convergence Index). It can be seen that the convergence order is close to the theoretical value of two in most of the domain, and that the grid convergence index is extremely small everywhere, except near the upper corners, where the solution is singular.

An example of application of this scheme to practical computation of the flow past a frigate hull with bulbous bow and transom stern (US Navy Combatant DDG 51; see figure 5) is reported in this section.

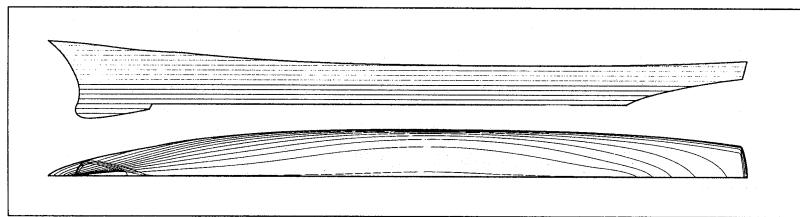


Figure 5. US Navy Combatant DDG 51 (INSEAN Model 2340).

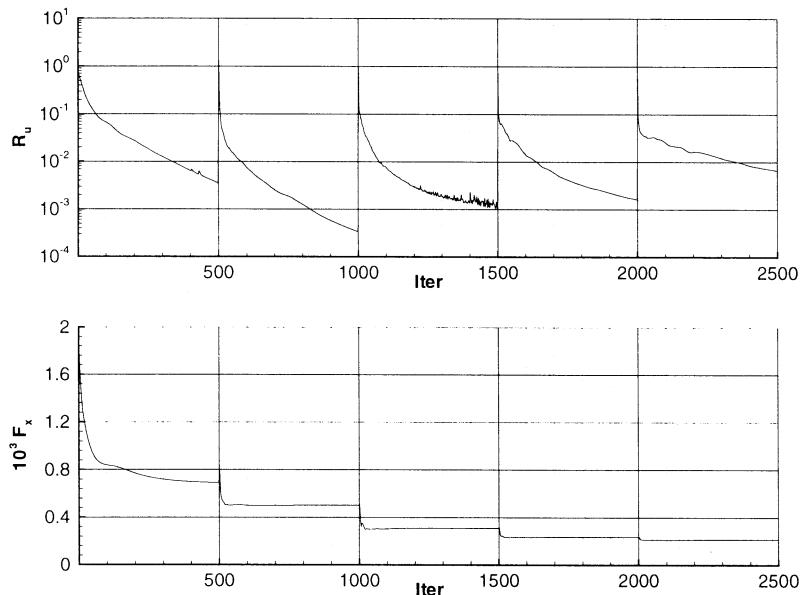


Figure 6. Top: residual history for u -velocity. Bottom: history of the non-dimensional x -force on the hull.

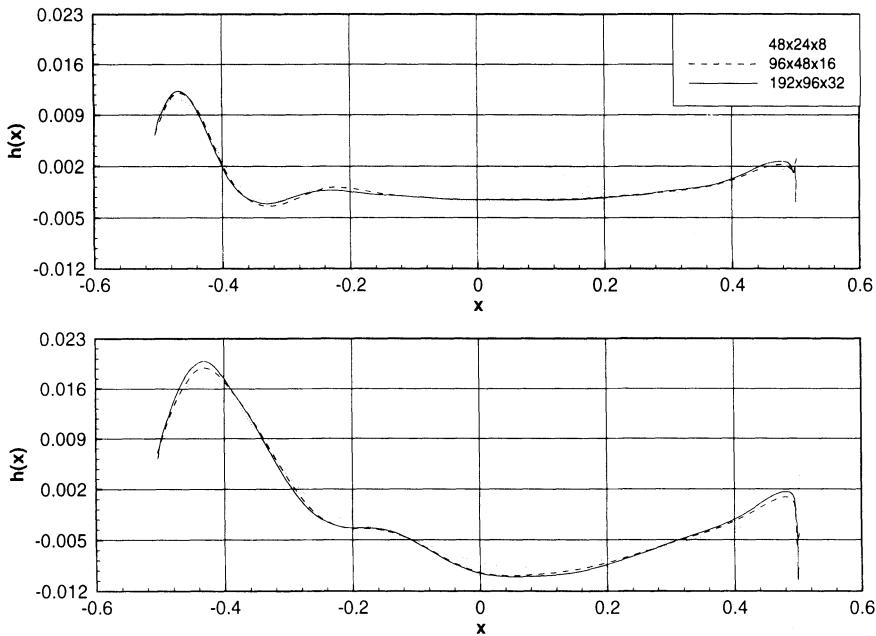


Figure 7. Wave profiles. Top: $Fr = 0.28$, $Re = 1.2 \times 10^7$. Bottom $Fr = 0.41$, $Re = 1.757 \times 10^7$. Grids $192 \times 96 \times 32$, $96 \times 48 \times 16$ and $48 \times 24 \times 8$.

The experiments reported in the following were carried out in the INSEAN towing tank. Two experimental conditions were simulated in the computations: $Fr = 0.28$ - $Re = 1.2 \times 10^7$ and $Fr = 0.41$ - $Re = 1.757 \times 10^7$. In both cases, a $192 \times 96 \times 32$ grid (stream-wise, normal and girth-wise directions, respectively) with O-O topology was used. The convergence history for the case $Fr = 0.28$ is reported in figure 6 in terms of x momentum residual and non dimensional resistance. Five levels were used for the Full Multigrid cycle; it can be seen that the iterative convergence was well attained after 500 V-cycles on all grids.

The wave profile for the two conditions are reported in figure 7 for three grids, while the computed wave patterns are compared with the measurements in figure 8. As it can be inferred from figures 6 and 7, grid convergence is almost reached, the variation of the numerical solution when refining the mesh being monotonic and reasonably small.

The comparison with experimental data in figure 8 is also satisfactory. In fact, the prediction of both wave height and phase is quite good in most of the field; of course, the wave breaking phenomenon, which appear close to the stern in the case $Fr = 0.41$ (highlighted by the clustering of contour lines) cannot be reproduced by the eulerian formulation used in the simulations (the high frequency oscillations in the measured data are due to the phase lag error of the capacity probes).

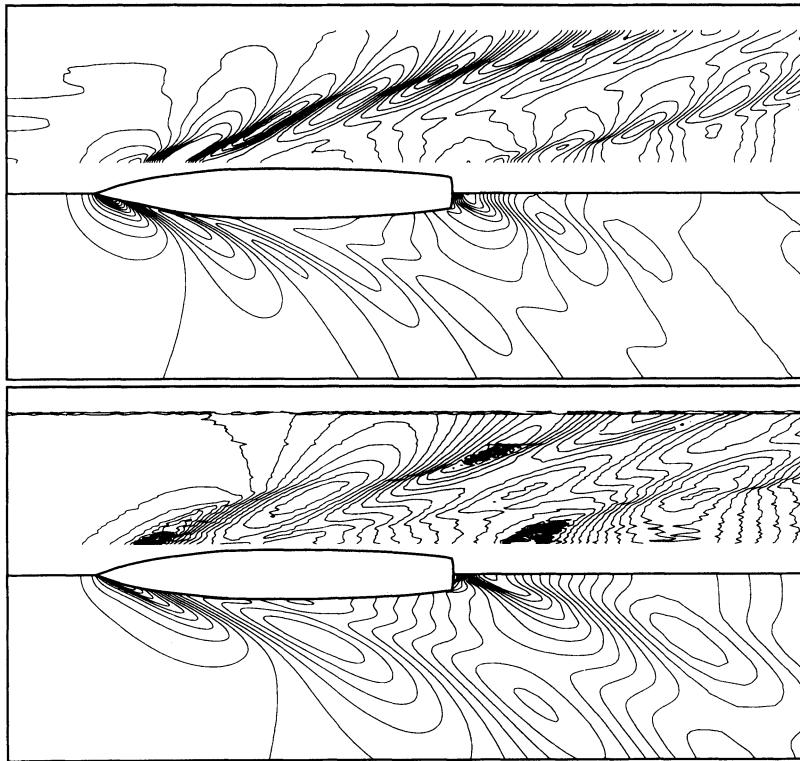


Figure 8. Wave pattern. Grid $192 \times 96 \times 32$. Top: $Fr = 0.28$, $Re = 1.2 \times 10^7$. Bottom: $Fr = 0.41$, $Re = 1.757 \times 10^7$ (Top half: measured data; Bottom half: numerical results).

5. Conclusions and perspectives

An ENO type scheme with multigrid acceleration was developed and applied to the simulation of free surface turbulent flows for both two-dimensional and three-dimensional problems. The proposed technique has proved to be robust and stable, and to have a good convergence rate. Accuracy has also been tested and the results appear satisfactory.

Some aspects will receive more attention in future work: multigrid performances in relation to cell aspect ratio near rigid walls, free surface flows with wave breaking, true time evolution.

Acknowledgements

Special thanks are due to Dr. Roberto Penna for the experimental data. The work was partially supported by the Italian Ministry of Transportation in the frame of INSEAN research plan 1997-99.

References

- Baldwin B S and Lomax H (1978). Thin Layer Approximation and Algebraic Model for Separated Turbulent Flows. *AIAA Paper* 78-257.
- Chorin A (1967). A Numerical Method for Solving Incompompressible Viscous Flow Problems. *J. Comput. Phys.* **2**, pp 12.
- Favini B, Broglia R and Di Mascio A (1996). Multigrid Acceleration of Second Order ENO Schemes from Low Subsonic to High Supersonic Flows. *Int. J. Num. Meth. Fluids* **23**, pp 589-606.
- Godunov S K (1959). A Finite Difference Method for the Computation of Discontinuous Solutions of the Equations of Fluid Dynamics. *Mat. Sb.* **47**, pp 357-393.
- Harten A, Engquist B, Osher S and Chakravarthy S R (1987). Uniformly High-order Accurate Essentially Non-oscillatory Schemes, III. *J. Comput. Phys.* **71**, pp 2-47.
- Jameson A, Schmidt W and Turkel E (1978). Numerical Solutions of the Euler Equations by Finite Volume Methods Using Runge-Kutta Time-Stepping Schemes. *AIAA Paper* 81-1259.
- Roache P J (1997). Quantification of Uncertainty in Computational Fluid Dynamics. *Ann. Rev. Fluid Mech.* **29**, pp 123-160.

UNIFORMLY HIGH ORDER METHODS FOR UNSTEADY INCOMPRESSIBLE FLOWS

D. DRIKAKIS

*Queen Mary and Westfield College
University of London
Department of Engineering
London E1 4NS, United Kingdom
Email: d.drikakis@qmw.ac.uk*

Abstract. The paper presents the development of *uniformly high-order (UHO)* schemes for unsteady incompressible flows. The schemes are designed via high-order (r th-order) polynomial reconstruction of the fluxes at the cell faces. The similarity of the present schemes with the essentially nonoscillatory (ENO) and weighted ENO (WENO) schemes, originally developed for flows with discontinuities, lies in the concept of high-order flux reconstruction. In the present case the combination of all neighbouring fluxes is obtained by an interpolation procedure based on constant weight coefficients. The latter are defined by posing numerical conditions for minimising the numerical dispersion and dissipation. The high-order flux reconstruction can be implemented in conjunction with any first- or second-order Godunov-type method. In the present work, the high-order reconstruction is combined with a characteristic-based (CB) scheme. The latter is a flux averaging procedure according to which the flow variables, and subsequently the fluxes, are calculated at the cell faces by an upwind Godunov-type scheme. For the time integration we employ a fourth-order TVD Runge-Kutta scheme and a non-linear multigrid method in conjunction with dual-time stepping. The primary goal of the paper is to present the development of the schemes. However, their capability to resolve complex unsteady flow features, is also demonstrated for the case of two- and three-dimensional direct numerical simulations of a jet in a doubly period geometry and laminar-to-turbulent transition in a mixing layer.

1. Introduction

There are different computational frameworks within which one can pursue the development of high-order methods. These include the spectral methods, finite element methods, particularly the more recent discontinuous Galerkin approach, finite difference methods and finite volume methods. In the present research we adopt the finite volume approach since so far this approach remains the most broadly employed in the development of both academic and industrial codes.

Existing advanced high-order methods include the essentially nonoscillatory schemes (ENO) originated from the works of Harten, Osher, Shu, and their co-workers (Harten et al., 1987; Liu et al., (1994); Jiang and Shu, (1996)). These methods have been developed over the past 10-15 years mainly in the context of the compressible Euler and Navier-Stokes equations, while very limited experience has also been gained by implementing these approaches in incompressible laminar flows (E and Shu, (1994)). Other approaches for constructing high-order methods are based on centred stencils, as distinct from upwind-biased (see (Jiang and Tadmor, (1998))), as well as compact difference schemes (Tolstykh, (1994)). Obvious advantages of high-order methods are: (i) the desired accuracy is attained with reduced memory and storage requirements, this is of paramount importance for three-dimensional time dependent problems, and (ii) bearing in mind the current computing resources available, very high order methods are the only way of providing reliable solutions for both practical engineering design as well as for more fundamental studies concerned with the simulation of flow phenomena not yet well understood (e.g. transition and turbulence). Low-order methods, even third-order accurate methods, can be completely inadequate without resorting to very fine meshes that are not affordable with current computers.

The aim of the present research is to develop high-order, high-resolution methods, both in space and time, for essentially unsteady two- and three-dimensional incompressible flows. Uniformly high-order (UHO) accuracy is obtained via high-order polynomial reconstruction of the fluxes at each cell face. The idea of polynomial high-order reconstruction originates from the development of essentially nonoscillatory (ENO) (Harten et al., 1987) and weighted ENO (WENO) (Liu et al., (1994); Jiang and Shu, (1996)) schemes for hyperbolic conservation laws. However, in the present case the reconstruction is not based on the “smootherst”-stencil approach employed in the case of ENO or the convex combination of stencils in the case of WENO schemes (Liu et al., (1994); Jiang and Shu, (1996)), both originally developed for flows with discontinuities. Instead, in the present case the polynomial reconstruction has constant weight coefficients which are as-

signed by solving a system of algebraic equations. The latter is derived by posing conditions for minimising the numerical dissipation and dispersion. The UHO approach can be combined with any Godunov-type method, the latter being used for calculating locally the cell face fluxes. In the present work, the characteristic-based (CB) scheme of (Drikakis et al., (1994); Drikakis, (1996)) has been employed to calculate the cell face fluxes which are used in the polynomial reconstruction.

The coupling of the continuity and momentum equations is obtained here via the artificial compressibility approach (Chorin, (1967)). For the time integration we have employed high-order TVD Runge-Kutta schemes as proposed by Shu and Osher (Shu and Osher, (1988)), in conjunction with dual-time stepping and non-linear multigrid methods (Drikakis et al., (1998)). The above are applied in two- and three-dimensional direct numerical simulations of single- and double-periodic mixing layers. The study shows that the schemes provide high resolution, even when coarse grids are used, both at low and high Reynolds number flows.

2. Dual Time Stepping and Artificial Compressibility

Following Chorin's idea (Chorin, (1967)), coupling of the continuity and momentum equations in the case of incompressible flows can be achieved by introducing a pseudotime derivative in the continuity equation. Subsequently, the equations are written in dimensionless form as

$$\begin{aligned} \frac{\partial u_i}{\partial t} + \frac{\partial}{\partial x_j} (u_i u_j + p \delta_{ij}) &= \frac{1}{Re} \Delta u_i \\ \frac{1}{\beta} \frac{\partial p}{\partial t} + \frac{\partial u_j}{\partial x_j} &= 0 \end{aligned} \quad (1)$$

where β is the artificial compressibility parameter, u_i are the velocity components, p is the pressure, ρ is the density, Re is the Reynolds number, and t is the time. The indices $i, j = 1, 2, 3$ refer to the space co-ordinates x, y, z . Temam (Temam, (1968)) has shown that under suitable hypotheses, solutions of the system of Eq. (1) exist and are unique, and that these solutions satisfy the incompressibility constraint in the limit as $1/\beta \rightarrow 0$. The artificial compressibility formulation can easily be extended to time accurate flows by performing subiterations at the the pseudotime level τ for each real time. For unsteady flows the system of equations is written as

$$\frac{\partial u_i}{\partial t} + \frac{\partial u_i}{\partial \tau} + \frac{\partial}{\partial x_j} (u_i u_j + p \delta_{ij}) = \frac{1}{Re} \Delta u_i \quad (2)$$

$$\frac{1}{\beta} \frac{\partial p}{\partial \tau} + \frac{\partial u_j}{\partial x_j} = 0 \quad (3)$$

The above system provides a coupling of the equations with respect to the pseudotime τ , at each real time step t . This approach is also referred to as dual-time stepping.

Since our aim is to develop high-order numerical methods for the Navier-Stokes equations in their most general form, i.e. 3D, unsteady and for arbitrary geometries, we employ the equations in generalised curvilinear co-ordinates. The equations are written in matrix form as

$$(JU)_\tau + (J\tilde{I}U)_t + (E_I)_\xi + (F_I)_\eta + (G_I)_\zeta = (E_V)_\xi + (F_V)_\eta + (G_V)_\zeta \quad (4)$$

The unknown solution vector U and matrix \tilde{I} are defined by

$$U = (p/\beta, u, v, w)^T, \quad \tilde{I} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (5)$$

The inviscid, E_I, F_I, G_I , and viscous, E_V, F_V, G_V , fluxes are written as functions of their corresponding counterparts e.g.

$$E_I = J(\tilde{E}_I \xi_x + \tilde{F}_I \xi_y + \tilde{G}_I \xi_z), \quad (6)$$

where the fluxes $\tilde{E}_I, \tilde{F}_I, \tilde{G}_I$ are the corresponding Cartesian fluxes:

$$\tilde{E}_I = \begin{pmatrix} u \\ u^2 + p \\ uv \\ uw \end{pmatrix}, \quad \tilde{F}_I = \begin{pmatrix} v \\ uv \\ v^2 + p \\ vw \end{pmatrix}, \quad \tilde{G}_I = \begin{pmatrix} w \\ uw \\ vw \\ w^2 + p \end{pmatrix} \quad (7)$$

In the above relations J is the Jacobian of the transformation from Cartesian (x, y, z) to generalised coordinates (ξ, η, ζ) .

3. High-resolution, High-Order Methods

We are interested in developing high-resolution, high-order numerical methods for the system of the Navier-Stokes equations. High-resolution can be achieved via Godunov-type schemes which calculate the fluxes at each cell face using (locally) high-order approximations for the flow variables. However, the global accuracy is limited to second-order due to the flux discretization obtained by $(E_I)_\xi = (E_I)_{i+1/2} - (E_I)_{i-1/2}$. By uniformly high-order schemes we refer to methods which extend the accuracy of the

discretisation by combining the cell face fluxes via high-order polynomial reconstructions. It is possible, though not yet fully exploited, to include in the analysis the viscous terms, as well. However, our experience so far indicates that second-order discretisation for the viscous terms is sufficient for obtaining accurate solutions both at low and high Reynolds number flows. Similar conclusions have also been drawn by others (e.g. (E and Shu, (1994))).

The analysis presented below concerns four points: (i) the basics for the development of the characteristic-based (CB) discretisation, (ii) the derivation of CB-average relations for the case of the three-dimensional flow equations, (iii) the development of UHO flux reconstruction and (iv) the implementation of the above in conjunction with high-order time accurate TVD Runge-Kutta discretisation and non-linear multigrid methods.

3.1. BASIC FORMULATION OF CHARACTERISTIC-BASED DISCRETISATION

We consider the one-dimensional hyperbolic conservation law

$$u_t + f_x(u) = 0 \quad (8)$$

The update of the solution is given by

$$u(t + \Delta t) = u(t) - \int_t^{t+\Delta t} f(u)_x dt \quad (9)$$

In the last equation the space derivative of the flux is only known at the initial time level t which is exactly at the lower bound of the interval of integration. However, the fundamental theorem of integration requires the integrand to be known inside the limits of integration for stable numerical update. Therefore, we need to make use of a device which will allow us to propagate the integrand in time and perform a stable numerical update. Let consider the propagation of the solution from time t to $t + \Delta t$ as shown in Fig. 1. In order to define the solution at the point $(x, t + \Delta t)$, we perform a linear backward Taylor series expansion in the neighbourhood of that point

$$u(x, t + \Delta t) = u(x - \Delta \xi, t) + \Delta \xi \frac{\partial u}{\partial x} + \Delta t \frac{\partial u}{\partial t} \quad (10)$$

where higher order terms have been neglected. By denoting $u_l \equiv u(x - \Delta \xi, t)$ and $\tilde{u} \equiv u(x, t + \Delta t)$, and by introducing the wave speed, $\dot{\xi} = \frac{\partial \xi}{\partial t}$, such that $\Delta \xi = \dot{\xi} \Delta t$, Eq. (10) is written

$$\tilde{u} = u_l + \left(\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} \dot{\xi} \right) \Delta t \quad (11)$$

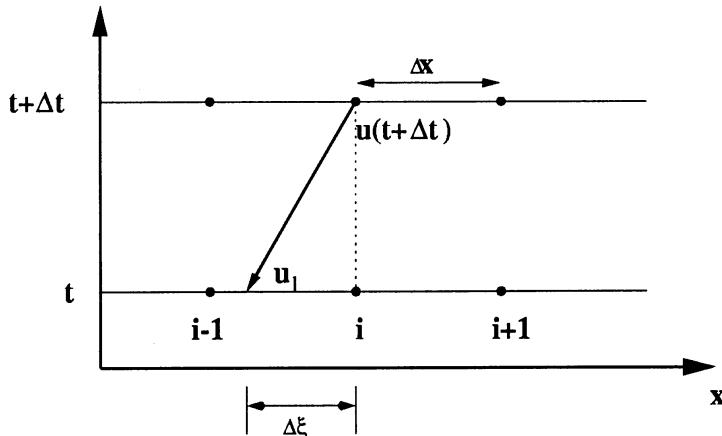


Figure 1. Schematic representation of the definition of variable $u(t+\Delta t) \equiv \tilde{u}$ as function of the characteristic variable u_i .

All terms in the above equation are unknown since neither the spatial and time derivatives nor the value u_i are known. Yet, the position from which the initial value u_i should be taken is also unknown. From the hyperbolic conservation law (Eq. 8) the term $\partial u / \partial t \equiv \dot{u}$ is given by

$$\dot{u} = -f_x(u)$$

Substituting \dot{u} from the last relation into Eq. (11), one obtains

$$\tilde{u} = u_i + u_x (\xi - f_u(u)) \Delta t \quad (12)$$

where we made use of $f_x(u) = f_u(u)u_x$. If, by definition, the wave speed is an eigenvalue i.e. $\xi = f_u(u)$, then the coefficient of the unknown spatial derivative on the right hand side of the above equation can be eliminated, thus, obtaining:

$$\tilde{u} = u_i \quad (13)$$

In other words, the solution \tilde{u} can be calculated by u_i (henceforth called *characteristic variable*) where the latter corresponds to the point with coordinates $(x - \xi \Delta t, t)$ and belongs to the line with slope $1/\xi$ (*characteristic line*). The variable u_i can subsequently be calculated by high-order interpolation as we shall see later. Once the value $\tilde{u} = u_i$ is known, the flux $f(\tilde{u})$ can be calculated. In the case where more than one characteristic lines are involved (multi-dimensional flow problems), then the flux $f(\tilde{u})$ will be defined as $f(g(u_l))$ since the variable \tilde{u} will be a function of all characteristic variables lying on the lines with slopes $1/\xi_l$ ($l = 0 \div 2$). In the next section we discuss how $f(g(u_l))$ can be constructed for the case of the three-dimensional incompressible flow equations.

3.2. CHARACTERISTIC-BASED RECONSTRUCTION

To present the derivation of characteristic-based relations for the flux E_I (Eq. 7) we consider the one-dimensional counterpart of Eq. (4)

$$(J\tilde{U})_\tau + (E(\tilde{U}))_\xi = 0 \quad (14)$$

where \tilde{U} is the vector of the variables for which characteristic-based solutions will be derived, and $E(\tilde{U}) \equiv E_I$ is the inviscid flux which will be defined in terms of \tilde{U} . We should also mention that the analysis below follows similar steps with those used in (Drikakis et al., (1994)) for the derivation of two-dimensional characteristic-based discretisation.

Let consider the non-conservative form of Eq. (14).

$$\begin{aligned} \frac{J}{\beta\sqrt{\xi_x^2 + \xi_y^2 + \xi_z^2}}\tilde{p}_\tau + \tilde{u}_\xi\tilde{x} + \tilde{v}_\xi\tilde{y} + \tilde{w}_\xi\tilde{z} &= 0 \\ \frac{J}{\sqrt{\xi_x^2 + \xi_y^2 + \xi_z^2}}\tilde{u}_\tau + \tilde{u}_\xi(\tilde{u}\tilde{x} + \tilde{v}\tilde{y} + \tilde{w}\tilde{z}) + \tilde{u}(\tilde{u}_\xi\tilde{x} + \tilde{v}_\xi\tilde{y} + \tilde{w}_\xi\tilde{z}) + \tilde{p}_\xi\tilde{x} &= 0 \\ \frac{J}{\sqrt{\xi_x^2 + \xi_y^2 + \xi_z^2}}\tilde{v}_\tau + \tilde{v}_\xi(\tilde{u}\tilde{x} + \tilde{v}\tilde{y} + \tilde{w}\tilde{z}) + \tilde{v}(\tilde{u}_\xi\tilde{x} + \tilde{v}_\xi\tilde{y} + \tilde{w}_\xi\tilde{z}) + \tilde{p}_\xi\tilde{y} &= 0 \\ \frac{J}{\sqrt{\xi_x^2 + \xi_y^2 + \xi_z^2}}\tilde{w}_\tau + \tilde{w}_\xi(\tilde{u}\tilde{x} + \tilde{v}\tilde{y} + \tilde{w}\tilde{z}) + \tilde{w}(\tilde{u}_\xi\tilde{x} + \tilde{v}_\xi\tilde{y} + \tilde{w}_\xi\tilde{z}) + \tilde{p}_\xi\tilde{z} &= 0 \end{aligned} \quad (15)$$

where

$$\tilde{k} = \frac{\xi_k}{\sqrt{\xi_x^2 + \xi_y^2 + \xi_z^2}}, \quad k = x, y, z$$

We apply now the procedure presented in the preceding section, namely we develop $U(\tau + \Delta\tau) \equiv \tilde{U}$ in Taylor series expansion (Eq. 11) around the time level τ .

$$U(\tau + \Delta\tau) = U_l(\tau) + \tilde{\Delta\xi}U_\xi + U_\tau\Delta\tau \quad (16)$$

where $U_l(\tau) \equiv U_l$ ($l = 0, 1, 2$) are the characteristic variables and the interval $\tilde{\Delta\xi}$ is defined by introducing a wave speed λ such that:

$$\tilde{\Delta\xi} = \lambda \frac{\sqrt{\xi_x^2 + \xi_y^2 + \xi_z^2}}{J} \Delta\tau.$$

Thus, Eq. (16) can be solved with respect to U_τ

$$U_\tau = \frac{U - U_l}{\Delta\tau} + \lambda \frac{\sqrt{\xi_x^2 + \xi_y^2 + \xi_z^2}}{J} U_\xi \quad (17)$$

By substituting the last relation into the system of Eqs. (15), we obtain

$$\begin{aligned} \frac{J}{\beta \Delta \tau \sqrt{\xi_x^2 + \xi_y^2 + \xi_z^2}} (\tilde{p} - p_l) + \frac{1}{\beta} \tilde{p}_\xi \lambda + \tilde{u}_\xi \tilde{x} + \tilde{v}_\xi \tilde{y} + \tilde{w}_\xi \tilde{z} &= 0 \\ \frac{J}{\Delta \tau \sqrt{\xi_x^2 + \xi_y^2 + \xi_z^2}} (\tilde{u} - u_l) + \tilde{u}_\xi (\lambda_0 - \lambda) + \tilde{u}(\tilde{u}_\xi \tilde{x} + \tilde{v}_\xi \tilde{y} + \tilde{w}_\xi \tilde{z}) + \tilde{p}_\xi \tilde{x} &= 0 \\ \frac{J}{\Delta \tau \sqrt{\xi_x^2 + \xi_y^2 + \xi_z^2}} (\tilde{v} - v_l) + \tilde{v}_\xi (\lambda_0 - \lambda) + \tilde{v}(\tilde{u}_\xi \tilde{x} + \tilde{v}_\xi \tilde{y} + \tilde{w}_\xi \tilde{z}) + \tilde{p}_\xi \tilde{y} &= 0 \\ \frac{J}{\Delta \tau \sqrt{\xi_x^2 + \xi_y^2 + \xi_z^2}} (\tilde{w} - w_l) + \tilde{w}_\xi (\lambda_0 - \lambda) + \tilde{w}(\tilde{u}_\xi \tilde{x} + \tilde{v}_\xi \tilde{y} + \tilde{w}_\xi \tilde{z}) + \tilde{p}_\xi \tilde{z} &= 0 \end{aligned}$$

where the eigenvalue λ_0 is defined by: $\lambda_0 = \tilde{u}\tilde{x} + \tilde{v}\tilde{y} + \tilde{w}\tilde{z}$

To eliminate the spatial derivatives in the last set of equations, we follow the idea presented in the book of Courant and Hilbert (Courant and Hilbert, (1968)) for eliminating unknowns in a system of linear equations (referred to as Riemann method). According to the latter, we multiply the above equations by arbitrary non-zero coefficients a, b, c and d for each of the four equations, respectively. Summation of the new equations gives

$$\begin{aligned} \frac{J}{\Delta \tau \sqrt{\xi_x^2 + \xi_y^2 + \xi_z^2}} \left(\frac{1}{\beta} a(\tilde{p} - p_l) + b(\tilde{u} - u_l) + c(\tilde{v} - v_l) + d(\tilde{w} - w_l) \right) \\ + \tilde{p}_\xi \left(-\frac{a}{\beta} \lambda + b\tilde{x} + c\tilde{y} + d\tilde{z} \right) + \\ \tilde{u}_\xi \left(a\tilde{x} + b(\lambda_0 - \lambda + \tilde{u}\tilde{x}) + c\tilde{v}\tilde{x} + d\tilde{w}\tilde{x} \right) + \\ \tilde{v}_\xi \left(a\tilde{y} + b\tilde{u}\tilde{y} + c(\lambda_0 - \lambda + \tilde{v}\tilde{y}) + d\tilde{w}\tilde{y} \right) + \\ \tilde{w}_\xi \left(a\tilde{z} + b\tilde{u}\tilde{z} + c\tilde{v}\tilde{z} + d(\lambda_0 - \lambda + \tilde{w}\tilde{z}) \right) &= 0 \end{aligned}$$

An ordinary set of differential equations can consequently be defined by setting the coefficients of the partial spatial derivatives (i.e. the terms inside the brackets) to be zero:

$$\frac{1}{\beta} a(\tilde{p} - p_l) + b(\tilde{u} - u_l) + c(\tilde{v} - v_l) + d(\tilde{w} - w_l) = 0 \quad (18)$$

$$-\frac{a}{\beta} \lambda + b\tilde{x} + c\tilde{y} + d\tilde{z} = 0 \quad (19)$$

$$a\tilde{x} + b(\lambda_0 - \lambda + \tilde{u}\tilde{x}) + c\tilde{v}\tilde{x} + d\tilde{w}\tilde{x} = 0 \quad (20)$$

$$a\tilde{y} + b\tilde{u}\tilde{y} + c(\lambda_0 - \lambda + \tilde{v}\tilde{y}) + d\tilde{w}\tilde{y} = 0 \quad (21)$$

$$a\tilde{z} + b\tilde{u}\tilde{z} + c\tilde{v}\tilde{z} + d(\lambda_0 - \lambda + \tilde{w}\tilde{z}) = 0 \quad (22)$$

The eigenvalues of the system of the above equations are:

$$\lambda_0 = \tilde{u}\tilde{x} + \tilde{v}\tilde{y} + \tilde{w}\tilde{z}, \quad \lambda_1 = \lambda_0 + s, \quad \lambda_2 = \lambda_0 - s \quad (23)$$

where $s = \sqrt{\lambda_0^2 + \beta}$. As seen, the eigenvalues are functions of $\tilde{u}, \tilde{v}, \tilde{w}$ which are not, however, yet known. Therefore, the eigenvalues need to be defined from the previous time level - for time accurate problems this would be the pseudotime level τ - using the values $U(\tau)$. Alternatively, one could consider to perform sub-iterations at this stage in order to obtain new approximations of the eigenvalues after the vector \tilde{U} has been calculated. Using such a procedure an “exact” Godunov method can be approached, but this would require very large computational resources.

A non-trivial solution of the above equations can be found for each of the eigenvalues.

For $\lambda_0 = u\tilde{x} + v\tilde{y} + w\tilde{z}$:

$$\tilde{x}(\tilde{w} - w_0) - \tilde{z}(\tilde{u} - u_0) = 0 \quad (24)$$

$$\tilde{x}(\tilde{v} - v_0) - \tilde{y}(\tilde{u} - u_0) = 0 \quad (25)$$

For $\lambda_1 = \lambda_0 + s$:

$$\tilde{p} = p_1 - \lambda_1 \left(\tilde{x}(\tilde{u} - u_1) + \tilde{y}(\tilde{v} - v_1) + \tilde{z}(\tilde{w} - w_1) \right) \quad (26)$$

For $\lambda_2 = \lambda_0 - s$:

$$\tilde{p} = p_2 - \lambda_2 \left(\tilde{x}(\tilde{u} - u_2) + \tilde{y}(\tilde{v} - v_2) + \tilde{z}(\tilde{w} - w_2) \right). \quad (27)$$

The above four equations can subsequently be solved to obtain the values of $\tilde{p}, \tilde{u}, \tilde{v}, \tilde{w}$ as functions of the characteristics values p_l, u_l, v_l, w_l ($l = 0, 1, 2$). After some algebra we obtain

$$\begin{pmatrix} \tilde{p} \\ \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{pmatrix} = \begin{pmatrix} \frac{1}{2s}(\lambda_1 k_2 - \lambda_2 k_1) \\ R\tilde{x} + u_0(\tilde{y}^2 + \tilde{z}^2) - v_0\tilde{x}\tilde{y} - w_0\tilde{x}\tilde{z} \\ R\tilde{y} + v_0(\tilde{x}^2 + \tilde{z}^2) - w_0\tilde{z}\tilde{y} - u_0\tilde{x}\tilde{y} \\ R\tilde{z} + w_0(\tilde{y}^2 + \tilde{x}^2) - v_0\tilde{z}\tilde{y} - u_0\tilde{x}\tilde{z} \end{pmatrix} \quad (28)$$

where

$$\begin{aligned} R &= \frac{1}{2s} \left(p_1 - p_2 + \tilde{x}(\lambda_1 u_1 - \lambda_2 u_2) + \tilde{y}(\lambda_1 v_1 - \lambda_2 v_2) + \tilde{z}(\lambda_1 w_1 - \lambda_2 w_2) \right) \\ k_1 &= p_1 + \lambda_1(u_1\tilde{x} + v_1\tilde{y} + w_1\tilde{z}) \\ k_2 &= p_2 + \lambda_2(u_2\tilde{x} + v_2\tilde{y} + w_2\tilde{z}) \end{aligned} \quad (29)$$

The variables given by Eq. (28) are *characteristic-based average quantities* which are used in defining the flux $E(\tilde{U}) \equiv E_I$ (see Eq. 7) at the cell faces $(i - 1/2, j, k)$ of the computational volume (i, j, k) .

We are now facing two issues: (i) to examine how the global accuracy of the flux discretization can increase and, (ii) to calculate the characteristic variables $U_l = (p_l, u_l, v_l, w_l)^T$ at the cell faces by high-order interpolation.

4. Uniformly High-Order Reconstruction

We consider again the one dimensional hyperbolic conservation law

$$u_t + f_\xi(u) = 0 \quad (30)$$

and define the spatial operator

$$L = -\frac{1}{\Delta x} (\tilde{f}_{i+1/2}(\tilde{u}) - \tilde{f}_{i-1/2}(\tilde{u})) \quad (31)$$

where \tilde{f} is the characteristic-based approximation of the flux f at the cell faces calculated using the averaging relations of Eq. (28). The flux \tilde{f} can be split into positive and negative parts

$$\tilde{f} = \tilde{f}^+ + \tilde{f}^- \quad (32)$$

where $d\tilde{f}^+ \geq 0$ and $d\tilde{f}^- \leq 0$. The positive and negative parts can subsequently be calculated using any flux splitting scheme, e.g. the Lax-Friedrichs or Roe flux formulae. In the case of the former the positive and negative fluxes are defined by

$$\tilde{f}^\pm = \frac{1}{2} (\tilde{f}(\tilde{u}) \pm \alpha \tilde{u}) \quad (33)$$

where $\alpha = \max |\tilde{f}'(\tilde{u})|$.

Our objective is to find high-order approximations of the fluxes \tilde{f}^\pm . Below we present the analysis only for the positive flux (for sake of simplicity we also drop the "+" sign in the superscript). Let $(i+1/2)$ and $(i-1/2)$ be the right and left faces, respectively, of the computational cell denoted by the index (i) . We define the fluxes $\tilde{f}_{i+1/2}$ and $\tilde{f}_{i-1/2}$ through r -th order polynomial approximations as

$$\tilde{f}_{i+1/2} = \sum_{k=-r+3-n}^{r-2} \alpha_k^r \tilde{f}_{i+k} \quad (34)$$

$$\tilde{f}_{i-1/2} = \sum_{k=-r+2-n}^{r-3} \alpha_{k+1}^r \tilde{f}_{i+k} \quad (35)$$

where

$$n = 0 \quad \forall \quad r > 3 \quad \text{and} \quad n = 1 \quad \text{if} \quad r = 3.$$

α_k^r are weight coefficients which need to be defined. For the case $r > 3$ ($n = 0$), the derivative of the flux at (i) is calculated by

$$(\tilde{f}_\xi)_i = \tilde{f}_{i+1/2} - \tilde{f}_{i-1/2} = \sum_{k=-r+3}^{r-2} \alpha_k^r \tilde{f}_{i+k} - \sum_{k=-r+2}^{r-3} \alpha_{k+1}^r \tilde{f}_{i+k} \quad (36)$$

which can be expanded to give

$$\begin{aligned} (\tilde{f}_\xi)_i = & \alpha_{-r+3}^r \tilde{f}_{i-r+3} + \alpha_{-r+4}^r \tilde{f}_{i-r+4} + \cdots + \alpha_0^r \tilde{f}_i \\ & + \cdots + \alpha_{r-3}^r \tilde{f}_{i+r-3} + \alpha_{r-2}^r \tilde{f}_{i+r-2} \\ & - \alpha_{r-1}^r \tilde{f}_{i-r+2} - \alpha_{r-2}^r \tilde{f}_{i-r+3} - \cdots \\ & - \alpha_1^r \tilde{f}_i - \cdots - \alpha_{r-3}^r \tilde{f}_{i+r-4} - \alpha_{r-2}^r \tilde{f}_{i+r-3} = \\ & - \alpha_{-r+3}^r \tilde{f}_{i-r+2} + (\alpha_{-r+3}^r - \alpha_{-r+4}^r) \tilde{f}_{i-r+3} + \cdots + \\ & (\alpha_0^r - \alpha_1^r) \tilde{f}_i + \cdots + (\alpha_{r-3}^r - \alpha_{r-2}^r) \tilde{f}_{i+r-3} + \alpha_{r-2}^r \tilde{f}_{i+r-2} \end{aligned} \quad (37)$$

By developing the fluxes \tilde{f}_{i+k} ($k = -r+2, \dots, r-2$) in Taylor series expansion around the point i , up to r -th order of accuracy in the computational space (thus the spacing between the cell centers is equal to one), we obtain

$$\begin{aligned} (\tilde{f}_\xi)_i = & -\alpha_{-r+3}^r \left(\tilde{f}_i - |-r+2| \cdot \tilde{f}^{(1)} + \frac{|-r+2|^2}{2!} \cdot \tilde{f}^{(2)} \right. \\ & \left. - \frac{|-r+3|^3}{3!} \cdot \tilde{f}^{(3)} + \cdots + (-1)^r \frac{|-r+2|^r}{r!} \cdot \tilde{f}^{(r)} \right) \\ & + (\alpha_{-r+3}^r - \alpha_{-r+4}^r) \left(\tilde{f}_i - |-r+3| \cdot \tilde{f}^{(1)} + \frac{|-r+3|^2}{2!} \cdot \tilde{f}^{(2)} - \right. \\ & \left. \frac{|-r+3|^3}{3!} \cdot \tilde{f}^{(3)} + \cdots + (-1)^r \frac{|-r+3|^r}{r!} \cdot \tilde{f}^{(r)} \right) + \cdots + (\alpha_0^r - \alpha_1^r) \tilde{f}_i \quad (38) \\ & + \cdots + (\alpha_{r-3}^r - \alpha_{r-2}^r) \left(\tilde{f}_i + (r-3) \cdot \tilde{f}^{(1)} + \frac{(r-3)^2}{2!} \cdot \tilde{f}^{(2)} \right. \\ & \left. + \cdots + \frac{(r-3)^r}{r!} \cdot \tilde{f}^{(r)} \right) + \\ & \alpha_{r-2}^r \left(\tilde{f}_i + (r-2) \cdot \tilde{f}^{(1)} + \frac{(r-2)^2}{2!} \cdot \tilde{f}^{(2)} + \cdots + \frac{(r-2)^r}{r!} \cdot \tilde{f}^{(r)} \right) \end{aligned}$$

where the terms $\tilde{f}^{(\cdot)}$ represent high-order flux derivatives. By consolidating the coefficients in the above relation, we can write the term $(\tilde{f}_\xi)_i$ as function of the high-order derivatives

$$\begin{aligned}
(\tilde{f}_\xi)_i &= \tilde{f}^{(1)} \left(\alpha_{-r+3}^r | -r+2| - (\alpha_{-r+3}^r + \alpha_{-r+4}^r) | -r+3| + \right. \\
&\quad \cdots (\alpha_{r-3}^r - \alpha_{r-2}^r) (r-3) + \alpha_{r-2}^r (r-2) \Big) + \\
&\quad \tilde{f}^{(2)} \left(-\alpha_{-r+3}^r \frac{| -r+2 |^2}{2!} - (\alpha_{-r+3}^r - \alpha_{-r+4}^r) \frac{| -r+3 |^2}{2!} + \cdots \right. \\
&\quad \left. + (\alpha_{r-3}^r - \alpha_{r-2}^r) \frac{(r-3)^2}{2!} + \alpha_{r-2}^r \frac{(r-2)^2}{2!} \right) + \cdots + \\
&\quad \tilde{f}^{(r)} \left((-1)^{r+1} \alpha_{-r+3}^r \frac{| -r+2 |^r}{r!} + (-1)^r (\alpha_{-r+3}^r - \alpha_{-r+4}^r) \frac{| -r+3 |^r}{r!} \right. \\
&\quad \left. + \cdots + \alpha_{r-2}^r \frac{(r-2)^r}{r!} \right)
\end{aligned} \tag{39}$$

The above general relation can be used to determine the unknown coefficients $\alpha_{(.)}^r$, if one sets certain numerical conditions. For example, we know that the even and odd derivatives are responsible for the dissipation and dispersion effects of the numerical scheme, respectively. Therefore, to construct schemes with minimum dissipation and dispersion errors one can set the coefficients of the derivatives $\tilde{f}^{(2)}, \tilde{f}^{(3)}, \dots, \tilde{f}^{(r)}$ equal to zero. Furthermore, the CFL condition requires the coefficient of the derivative $\tilde{f}^{(1)}$ to be equal to 1. Thus, the following system of algebraic equations can be obtained

$$\begin{aligned}
&\alpha_{-r+3}^r | -r+2| - (\alpha_{-r+3}^r + \alpha_{-r+4}^r) | -r+3| + \cdots \\
&+ (\alpha_{r-3}^r - \alpha_{r-2}^r) (r-3) + \alpha_{r-2}^r (r-2) = 1
\end{aligned} \tag{40}$$

$$\begin{aligned}
&-\alpha_{-r+3}^r \frac{| -r+2 |^2}{2!} - (\alpha_{-r+3}^r - \alpha_{-r+4}^r) \frac{| -r+3 |^2}{2!} + \cdots \\
&+ (\alpha_{r-3}^r - \alpha_{r-2}^r) \frac{(r-3)^2}{2!} + \alpha_{r-2}^r \frac{(r-2)^2}{2!} = 0
\end{aligned} \tag{41}$$

.....

$$\begin{aligned}
&(-1)^{r+1} \alpha_{-r+3}^r \frac{| -r+2 |^r}{r!} + (-1)^r (\alpha_{-r+3}^r - \alpha_{-r+4}^r) \frac{| -r+3 |^r}{r!} + \cdots \\
&+ (\alpha_{r-3}^r - \alpha_{r-2}^r) \frac{(r-3)^r}{r!} + \alpha_{r-2}^r \frac{(r-2)^r}{r!} = 0
\end{aligned} \tag{42}$$

Solution of the above system provides the values of the coefficients $\alpha_{(.)}^r$ of the series expansions and, thus, the high-order reconstruction is completed.

Let's see how the above are applied in the case of the fourth order scheme. In this case the algebraic system is written

$$\alpha_{-1} + \alpha_0 + \alpha_1 + \alpha_2 = 1 \tag{43}$$

$$\alpha_1 - \alpha_0 + 3(\alpha_2 - \alpha_{-1}) = 0 \tag{44}$$

$$\alpha_1 + \alpha_0 + 7(\alpha_2 + \alpha_{-1}) = 0 \tag{45}$$

$$\alpha_1 - \alpha_0 + 15(\alpha_2 - \alpha_{-1}) = 0 \tag{46}$$

Solution of the above system gives $\alpha_0 = \alpha_1 = 7/12$, $\alpha_{-1} = 1/12$, and $\alpha_2 = -1/12$. For the case of third-order reconstruction: $\alpha_0 = 5/6$, $\alpha_{-1} = -1/6$, $\alpha_1 = 1/3$ and $\alpha_2 = 0$.

It is also obvious that a similar reconstruction can also be derived for the case of the characteristic flow variables $U_l = (p_l, u_l, v_l, w_l)^T$ in Eq. (28). Specifically, U_l can be calculated at any cell face according to the sign of the local eigenvalue

$$(U_l^{(\pm)})_{i+\frac{1}{2}} = \frac{1}{2} \left([1 + \text{sign}(\lambda_l)] U_{i+\frac{1}{2}}^{(+)} + [1 - \text{sign}(\lambda_l)] U_{i+\frac{1}{2}}^{(-)} \right), \quad (47)$$

If, for example, we employ the third-order reconstruction, then the positive U^+ and negative U^- counterparts of U are given by

$$U^+_{i+\frac{1}{2}} = \frac{1}{6}(5U_i - U_{i-1} + 2U_{i+1}), \quad U^-_{i+\frac{1}{2}} = \frac{1}{6}(5U_{i+1} - U_{i+2} + 2U_i) \quad (48)$$

5. High-Order Time Discretisation and Non-Linear Multigrid Implementation

In this section we discuss the high-order time discretisation of the system of equations (4). Eq. (4) can also be written in the form:

$$(Q)_\tau + N(Q) = 0 \quad (49)$$

where $Q = JU$ and the operator $N(Q)$ contains the discretised inviscid and viscous fluxes, as well as the term $(\tilde{I}Q)_t$, i.e.

$$N(Q) = (\tilde{I}Q)_t + (E_I)_\xi + (F_I)_\eta + (G_I)_\zeta - (E_V)_\xi - (F_V)_\eta - (G_V)_\zeta \quad (50)$$

The time integration of Eq. (49) requires iterations to be performed at the pseudotime level τ . Let ν and n denote the pseudo-iterations and real time steps, respectively. In order to forward the solution from n to $n+1$ we consider the $(m+1)$ th TVD Runge-Kutta discretisation of Eq. (49) as proposed by Shu and Osher (Shu and Osher, (1988)). The general discretisation form is:

$$Q^{(i)} = \sum_{k=0}^{i-1} \left(\tilde{a}_{ik} Q^{(k)} + b_{ik} \Delta t N(Q^{(k)}) \right), \quad i = 1, 2, \dots, m \quad (51)$$

with

$$Q^{(0)} = Q^{(\nu)}, \quad Q^{(m)} = Q^{(\nu+1)} \quad (52)$$

$$\lim_{Q_\tau \rightarrow 0} Q^{(\nu+1)} \longrightarrow Q^{(n+1)}$$

The above formulation provides $(m + 1)$ th order methods for $m \leq 3$, m th order methods for $m = 4, 5, 6$, and $(m - 1)$ th order methods for $m = 7, 8$ (Shu and Osher, (1988)). We have considered here the fourth order version of the above scheme which is written in a consolidated form as

$$\begin{aligned}
 Q^{(0)} &= Q^{(\nu)} \\
 Q^{(1)} &= Q^{(\nu)} - \frac{\Delta\tau}{2} N(Q^{(0)}) \\
 Q^{(2)} &= Q^{(\nu)} - \frac{\Delta\tau}{2} N(Q^{(1)}) \\
 Q^{(3)} &= Q^{(\nu)} - \Delta\tau N(Q^{(2)}) \\
 Q^{(\nu+1)} &= Q^{(\nu)} - \\
 &\quad \frac{\Delta\tau}{6} [N(U^{(0)}) + 2N(U^{(1)}) + 2N(U^{(2)}) + N(U^{(3)})]
 \end{aligned} \quad (53)$$

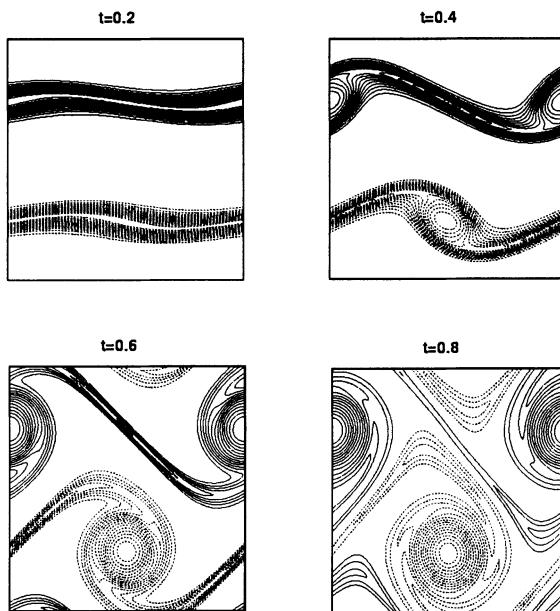


Figure 2. Evolution of a periodic jet (double mixing layer) in a double period geometry as predicted by the third-order UHO scheme ($Re=5000$; grid 128×128 , dashed lines denote negative vorticity).

To achieve high numerical efficiencies the schemes described above have been implemented in conjunction with a non-linear multigrid method (Drikakis et al., (1998)). The artificial compressibility and dual time-stepping formulations allow us to implement the multigrid approach directly to the Navier-Stokes equations, in contrast to pressure-based (pressure-correction) approaches where the benefits from using multigrid emerge indirectly from the acceleration of the Poisson equation for the pressure and/or from the acceleration of the SIMPLE-type smoothing procedure. Recently, Drikakis et al. (Drikakis et al., (1998)) developed a non-linear multigrid method based on the full multigrid (FMG) – full approximation storage (FAS) algorithms, which is realized via an unsteady-type procedure according to which the equations are not solved exactly on the coarsest grid, but some pseudo-time iterations are performed on the finer grids and some on the coarsest grid. The results in (Drikakis et al., (1998)) showed significant acceleration to be achieved in steady flow problems. In time accurate computations the multigrid solution of the equations is obtained by performing V-cycles at each real time step, starting directly from the fine grid.

6. Results

The above schemes are currently investigated in a variety of unsteady flow problems including laminar, transitional and turbulent flows. In the present paper we will present results from two- and three-dimensional direct numerical simulations of a developing jet in a double period geometry as well as of a developing mixing layer which features laminar-to-turbulent transition. Both cases include complex flow features such as large gradients, vortical structures as well as transition to unstable modes.

The first case has been proposed by Bell et al. (Bell et al., (1989)) to validate a second-order projection method. The same case has also been employed by E and Shu (E and Shu, (1994)) to investigate ENO schemes in incompressible flows. The flow problem is defined by two horizontal shear layers of finite thickness, perturbed by a small amplitude vertical velocity

$$u = \begin{cases} \tanh((y - 0.25)/\delta) & \text{if } y \leq 0.5 \\ \tanh((0.75 - y)/\delta) & \text{if } y > 0.5 \end{cases} \quad (54)$$

$$v = v' \sin(2\pi x) \quad (55)$$

where δ determines the shear layer thickness and v' is the perturbation amplitude in the normal direction. All computations reported below have been performed using the three-dimensional code into which the schemes have been implemented. However, in order to be able to compare, wherever possible, with Bell's et al. results we have not imposed random perturba-

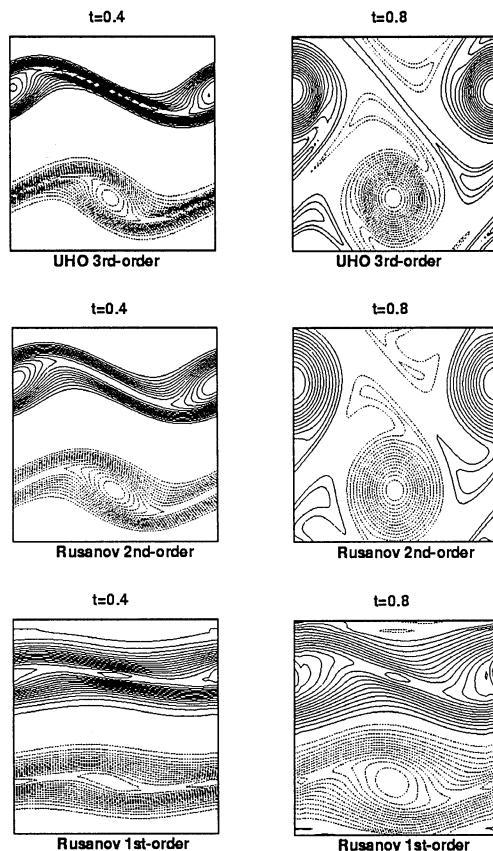


Figure 3. Comparison of 3rd-order UHO with the corresponding results obtained by first- and second-order Rusanov upwind fluxes ($Re=5000$; grid 64×64).

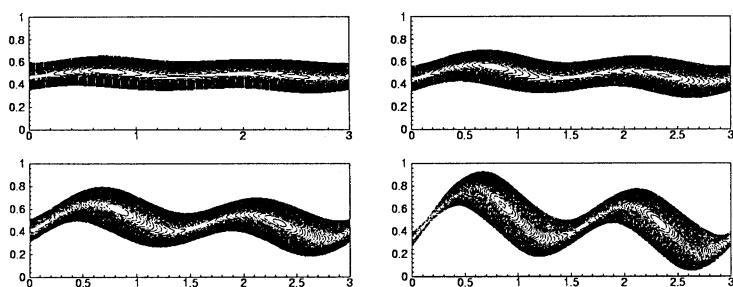


Figure 4. Evolution of the transitional mixing layer at $Re\delta = 220$ (grid 80^3 , $x-y$ middle plane).

tions in the spanwise direction and, thus, three-dimensional instabilities do not initiate in this case.

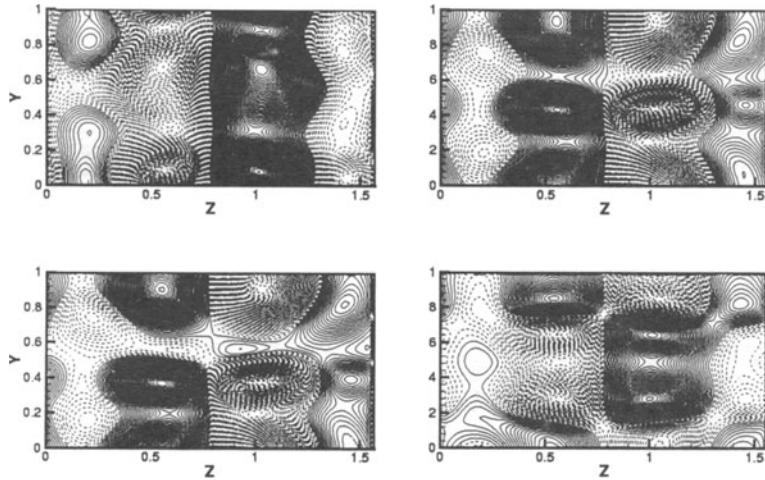


Figure 5. Streamwise vorticity (ω_x) contours in the $z - y$ plane at four different streamwise positions ($t^* = t\Delta U/(2\delta) = 96.6$; streamwise positions increase from left to right and from top to bottom).

The evolution of the flow field for the case of $Re = 5000$ ($\delta = 1/30, v' = 0.05$) is shown in Figure (2) using the third-order version of the UHO scheme. As seen, the flow evolves into large vortices and as time goes on the shear layers are becoming thin due to the large straining field. The results of Fig. (2) have been obtained on a grid 128×128 . In order to show the effects of numerical discretisation on the flow resolution, we have also employed a coarser grid (64×64) and compare the results obtained by the third-order UHO with the corresponding results obtained by first- and second-order Rusanov fluxes (Fig. 3). The objective of these comparisons is to show how the large dissipation exhibited by lower-order schemes can lead to misrepresentation of the flow structures. From Figs. (2) and (3), one can see that using uniformly-high order discretisation many of the flow features are still captured even on the coarse grid. Thus, higher-order discretisation does not only provide a better accuracy than lower order schemes when the same grid resolution is used, but it can also provide more accurate results on coarser grids, thus, resulting in significant savings of CPU time. In all

computations, the code converged to the machine zero (the L_2 norm of the residuals reached below 10^{-8} on a Dell workstation using single precision).

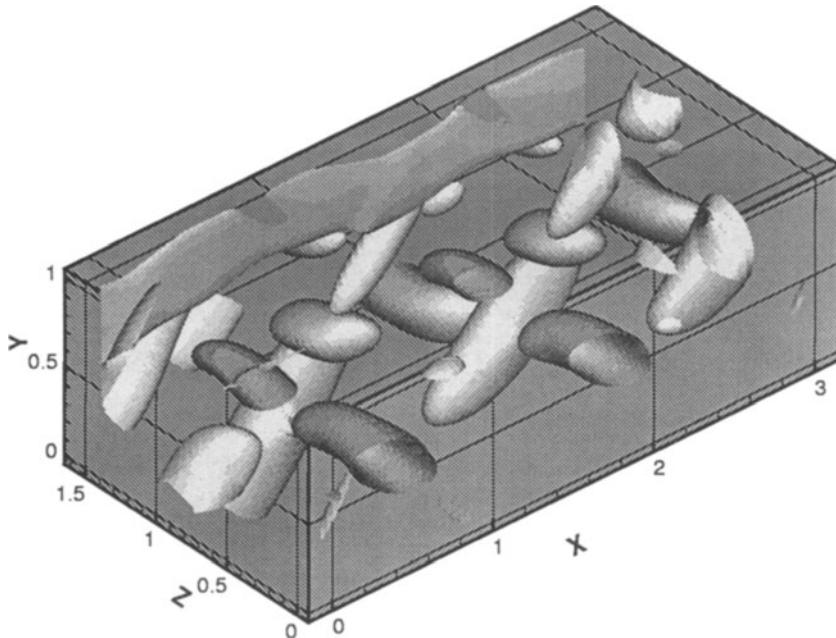


Figure 6. Iso-surfaces of constant vorticity ω_y at $t^* = 96.6$.

The second problem considered here is the the direct numerical simulation of a transitional mixing layer. This problem was chosen in order to examine the capability of the UHO approach in simulating complex three-dimensional unsteady flows featuring large gradients and transition to turbulence. Due to the limited length of the manuscript, we present only indicative results from these simulations and more detailed analysis will be presented in a future paper. The Reynolds number is $Re = 220$ based on the initial vorticity thickness, δ_i , and the half velocity difference across the layer (ΔU). Investigation of similar mixing layers have been presented by Metcalfe et al. (Metcalfe et al., (1987)), Rogers & Moser (Rogers and Moser, (1992)), and Ansari (Ansari, (1997)). In all these studies the simulations were obtained using spectral methods. Comparisons with the above studies, wherever possible, with respect to the instability development and formation of three-dimensional coherent structures, can be made only on a qualitative basis since the initialization of the random disturbance field, Reynolds number and box dimensions in all the above works are not exactly the same. The box employed here had dimensions $L_x : L_y : L_z = 3 : 1 : 1.57$

and the grid had 80^3 points. A tangential profile $u(y)$ in conjunction with three-dimensional random perturbations was used to initiate the computations. The development of the shear layer is shown in Figure (4) by plotting the spanwise vorticity component ω_z . The development of the secondary instability and transition to turbulence can, however, be better observed by plotting the streamwise vorticity component in the (z, y) -plane (Fig. 5). The results shown in this figure are in accord with the observations made by Metcalfe et al. (Metcalfe et al., (1987)) and Roger & Moser (Rogers and Moser, (1992)). Specifically, there is strong spanwise coherence and the secondary instabilities are characterized by streamwise counter rotating vortices that tend to form in the braids. The formation of the pairs of positive-negative vorticity rollers formed in the (z, y) planes have been similarly predicted in (Rogers and Moser, (1992)) for a higher Reynolds number. The existence of large scale structures is also shown in Fig. (6) in which surfaces of constant ω_y (negative) vorticity have been plotted. These results are in agreement with the observations reported in (Metcalfe et al., (1987); Rogers and Moser, (1992)).

Finally, the growth of the mean vorticity thickness (Riley et al., (1986)) is shown in Fig. (7). As seen, initially there is a laminar regime in which the growth is predicted to be linear. The laminar regime is followed by a rapid growth indicating transition to turbulence. The above agree well with the similarity theory and previous simulations (e.g. (Riley et al., (1986); Ansari, (1997))).

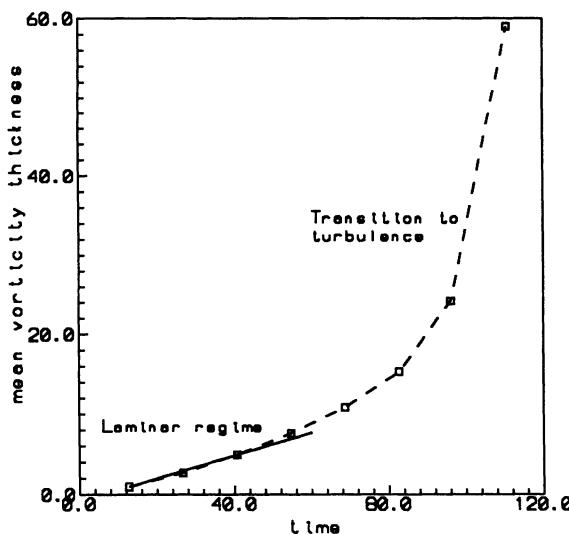


Figure 7. Growth of mean vorticity thickness.

7. Conclusions

A new family of uniformly high-order methods was developed. In the present work, the third-order version of the UHO approach was investigated in conjunction with a Godunov-type scheme, a fourth-order TVD Runge-Kutta method for the time integration and a non-linear multigrid method for the acceleration of the convergence. The results showed that the UHO approach provides high resolution of complex flow physics even when coarse grids are employed. Moreover, the method was found to be highly robust up to high Reynolds numbers. Extensive tests (not presented here) have shown that the method remains stable, converging down to the machine zero, even in the Euler limit ($\frac{1}{Re} \rightarrow 0$). The combination of UHO with the non-linear multigrid method of (Drikakis et al., (1998)) allowed us to perform direct numerical simulation of a transitional mixing layer using 512,000 grid points on a Pentium II at 400 MHz. This computation required about 30 hrs. The results showed that using the present schemes details of the secondary instability and formation of large scale coherent structures can be captured satisfactorily even on coarse grids. Investigation of higher-order of accuracy ($r \geq 4$) versions of the UHO approach for three-dimensional unsteady flows is under way and results will be presented in a future work.

References

- Harten A, Engquist B, Osher S and Chakravarthy S-R (1987). Uniformly high-order accurate essentially non-oscillatory schemes, III. *J. Comput. Phys.* **71**, pp 2-47.
- Liu X-D, Osher S and Chan T (1994). Weighted essentially non-oscillatory schemes. *J. Comput. Phys.* **115**, pp 200-212.
- Jiang G-S and Shu C-W (1996). Efficient implementation of weighted ENO schemes. *J. Comput. Phys.* **126**, pp 202-228.
- E W and Shu C-W (1994). A numerical resolution study of high order essentially non-oscillatory schemes applied to incompressible flow. *J. Comput. Phys.* **110**, pp 39-46.
- Jiang G-S and Tadmor E (1998). Non-oscillatory central schemes for multidimensional hyperbolic conservation laws. *SIAM J. of Scientific Computing* **19**, pp 1892-1917.
- Tolstykh A I (1994). High Accuracy Non-Centred Compact Finite Difference Schemes for Fluid Dynamics Applications. World Scientific Publishing.
- Drikakis D, Govatsos P A and Papantonis D E (1994). A characteristic-based method for incompressible flows. *Int. J. Num. Meth. in Fluids* **19**, pp 667-685.
- Drikakis D (1996). A parallel multiblock characteristic-based method for three-dimensional incompressible flows. *Advances in Engineering Software* **26**, pp 111-119.
- Chorin A J (1967). A numerical method for solving incompressible viscous flow problems. *J. Comput. Phys.* **2**, pp 12-26.
- Shu C-W and Osher S (1988). Efficient implementation of essentially non-oscillatory shock-capturing schemes. *J. Comput. Phys.* **77**, pp 439-471.
- Drikakis D, Iliev O P and Vassileva D P (1998). A nonlinear multigrid method for the three-dimensional incompressible Navier-Stokes equations. *J. Comput. Phys.* **146**, pp 310-321.
- Temam, R (1968). Bulletin Soc. Math. Fr. **96**, pp 115.
- Courant R and Hilbert D (1968). Methoden der Mathematischen Physik. Springer Verlag.
- Bell J B, Colella P and Glaz H M (1989). A second-order projection method for the

- incompressible Navier-Stokes equations. *Comput. Phys.* **85**, pp 257-283.
- Riley J J, Metcalfe R W and Orszag S A (1986). Direct numerical simulations of chemically reacting turbulent mixing layers. *Phys. Fluids* **29**, pp 406-422.
- Metcalfe R W, Orszag S A, Brachet M E, Menon S and Riley J (1987). Secondary instability of a temporally growing mixing layer. *J. Fluid Mech.* **184**, pp 207-243.
- Ansari A (1997). Self-similarity and mixing characteristics of turbulent mixing layers starting from laminar flow conditions. *Phys. Fluids* **9(6)**, pp 1714-1718.
- Rogers M M and Moser R D (1992). The three-dimensional evolution of a plane mixing layer: the Kelvin-Helmholtz rollup. *J. Fluid Mech.* **243**, pp 183-226.

APPLICATION OF THE FINITE VOLUME METHOD WITH OSHER SCHEME & SPLIT TECHNIQUE FOR DIFFERENT TYPES OF FLOW IN A CHANNEL

K.S. ERDURAN

*Department of Civil Engineering,
University of Newcastle upon Tyne,
Newcastle upon Tyne, NE1 7RU, U.K.
Email: Kutsi.Erduran@ncl.ac.uk*

AND

V. KUTIJA

*Department of Civil Engineering,
University of Newcastle upon Tyne,
Newcastle upon Tyne, NE1 7RU, U.K.
Email: Vedrana.Kutija@ncl.ac.uk*

Abstract.

A two dimensional unsteady flow model, FVSSWE, has been applied to different types of flow such as discontinuous flow, unsteady flow and supercritical flow in a channel. The model is based on the *finite volume method (FVM)* using the *Osher scheme* to solve the *homogeneous* part of the shallow water equations. In order to include the contribution of *source* and *sink terms*, conservative variables are updated by solving the *non-homogenous* part of the equations consisting of a bottom elevation and friction terms. This can be solved by a *fourth-order Runge Kutta method* since it is a set of ODEs. The results obtained are compared with previously published results and with other well-known methods and are shown to exhibit good agreement with them.

1. Introduction

The finite volume method (FVM) has been widely used in many fields for numerical solution of partial differential equations (PDEs) such as those de-

scribed by the shallow water equations. Certain formulations of the *FVM* have been used to model problems involving rapid change such as shock waves. These include *Osher*, HLL, and Roe schemes. The published work deals mainly with dam break problems (Alcrudo and Garcia-Navarro, 1993), (Zhao et al., 1994), and (Mingham and Causon, 1998), flows in rivers and flood plains that are initially dry (Zhao et al., 1994). There are also a few papers on different aspects of hydraulics such as the application of the method to coastal flow and simulation of hydraulic jumps (Chippada et al., 1998), and use of the method to model the propagation of a tidal wave (Bermúdez et al., 1998). The *FVM* with *shock capturing schemes* seems to be one of the best methods used in computational hydraulics because it can handle simulation of a wide variety of types of flows. However, there are some difficulties. First of all, almost all of the previously used shock-capturing schemes are *explicit*. This means that time step limitations must be applied in order to avoid stability problems. On the other hand, implicit schemes can be used but they require the solution of a system of simultaneous equations at each time step, which turns to be computationally demanding. The second problem is caused by the *source* and *sink terms* that make the shallow water equations *non-homogenous*. The *Riemann problem* is approximately solved by using *shock-capturing schemes* based on characteristic wave theory and the idea of flux balances through cell interfaces. The solution of the *Riemann problem* at cell interfaces can be achieved easily if the equations are *homogenous*. Otherwise, the solution may require balancing the *source/sink terms* with the fluxes, which is non-trivial. There has been some work done on the treatment of the *source terms*. For example, one approach given in the literature is to *upwind* the *source terms* (Bermúdez et al., 1998) and (Vazquez-Cendon, 1999). Another more commonly used method is to use the *splitting technique* (Hu et al., 1998) and (Zoppou and Roberts, 1999).

2. Description of FVSSWE and Solution Methods

The model, FVSSWE (Finite Volume Solution of the Shallow Water Equations), solves the two-dimensional unsteady problem given by equation[1]. The code is written in Delphi, which is an object-oriented programming language. The main features of the model are that it is capable of 1D and 2D flow simulation, simulation of subcritical, supercritical, steady, unsteady, and discontinuous flows. It can also handle initially dry areas and complex topography, is suitable for both structured and unstructured grids, and can handle different boundary conditions such as rating curve, inflow hydrograph, discharge, closed and open boundaries.

2.1. SOLUTION METHODS

The two dimensional conservative form of the shallow water equations can be written as

$$\frac{\delta[q]}{\delta t} + \frac{\delta f[q]}{\delta x} + \frac{\delta g[q]}{\delta y} = b[q] \quad (1)$$

where q is called the conserved physical vector, $f[q]$ is the flux vector in the x direction and $g[q]$ is the flux vector in the y direction, $b[q]$ denotes source/sink terms. Denoting $F[q] = (f[q], g[q])$ then the last form of the equation before integration, which is a key element in the *FVM*, can be written as;

$$\frac{\delta q}{\delta t} + \nabla F[q] = b[q] \quad (2)$$

Equation [2] is integrated over each cell (finite volume) covering the domain. Using the divergence theorem and the rotational invariance property between $f[q]$ and $g[q]$ on each side of the cell and resolving $F[q]$ in the direction of the normal vector n which allows us to treat the 2D problem as a series of 1D local problems , the discretized form of the *FVM* can be written as follows;

$$A \frac{\delta q}{\delta t} = - \sum_{j=1}^m T^{-1}(\theta) f(\bar{q}) L^j + Ab[q] \quad (3)$$

where A is the area of the element, m shows the total number of sides of an element, $T^{-1}(\theta)$ is the inverse transformation matrix and can be obtained after rotating the co-ordinate axes, L is the length of a side of an element, j is an index that represents the side. θ is the angle between the outward normal vector n and the x axis. \bar{q} means the vector q is multiplied by the transformation matrix and has the velocities in the normal and tangential directions given below. $\bar{u}=u\cos\theta+v\sin\theta$, $\bar{v}=-u\sin\theta+v\cos\theta$

2.2. SOLUTION OF THE INITIAL VALUE RIEMANN PROBLEM

Since the 2D problem is converted into 1D local *Riemann problems*, the next step is to solve this initial value problem;

$\bar{q}_t + [f(\bar{q})]_{\bar{x}} = 0$ where \bar{x} is a new local axis in the direction of normal vector n , with its origin located at the midpoint of the corresponding side.

In the solution of this problem, *Osher's scheme* is used. Osher's scheme is based on characteristic theory and the integration path that connects two states defining the left and the right Riemann interface. Solution of the *Riemann problem* using the *Osher scheme* results in 16 different cases that are used for the estimation of the flux through that interface. Although some

of these cases are unlikely to be seen in the field of hydraulics, others are used to determine the normal fluxes under the occurrence of the subcritical, supercritical, shock wave, and critical flows. For detailed information on the solution of the *Riemann problem* using the *Osher scheme*, the reader can consult the references (Toro, 1997) and (Zhao et al., 1994).

2.3. SPLIT TECHNIQUE FOR THE SOURCE AND SINK TERMS

Toro (1997, p.507) claims that "at the present time there seems to be no clear alternative to *splitting*" and there is still no agreement as to how best to treat the *source terms*. The most widely used approach, the *splitting technique*, is used in the FVSSWE model. Equation [3] is split into its *homogenous* and *non-homogenous* parts. While the *homogenous* part of equation [3] is solved using the *Osher scheme*, a set of ordinary differential equations(ODEs) given as $q_t = b[q]$ and having bottom slope and friction terms only is solved to update the variables (h, u, v) for each time step. For the solution of these ODEs, a *fourth-order Runge Kutta method* is used. It is noted that in order to solve the *non-homogenous* part, values obtained from the solution of the *homogenous* part are required to use as initial values. The advantage of the *splitting technique* is that it is simple and easy to implement. The solution method for the split part is independent of the solution of homogenous part. Therefore, any solution method can be chosen for the split part.

3. Applications of The Model

Supercritical and Subcritical Flow Tests: The purpose of this test is to confirm that the model can simulate *subcritical* and *supercritical* flow in a channel. To do this, the computational domain is divided into 130 cells in the x direction and 8 cells in the y direction. Each cell has a length of 5m in the x and y directions. While Manning's coefficient, n , is kept constant as 0.02, four different constant slopes in the x direction are selected to create *subcritical* and *supercritical* flow in the same channel. Unit discharge of $15m^2/s$ is applied in the x direction as the upstream boundary condition and downstream boundary condition is given as an open boundary. For each case, the model is run until the steady state condition is achieved in the channel. Then, a comparison is made between the model's results and the results obtained from Manning's equation for a rectangular channel. The results are given in Table 1.

Sustained Hydraulic Jump Test: The aim of this test is to simulate a *hydraulic jump* that is *sustained* in a channel. The computational domain is the same as that used for the previous tests. It has been assumed that

Slope in the x direction	Normal water depth(m) (Manning's Formula)	Normal Water Depth (m) (Manning's Formula with Wide Channel Approximation)	Normal Water depth(m) (Model)	Froude Number (computed)	Froude Number (Model)
0.0008	4.47	4.13	4.18	0.507	0.703
0.002	3.33	3.13	3.13	0.787	0.864
0.006	2.36	2.25	2.25	1.324	1.418
0.01	2.01	1.93	1.93	1.682	1.781

TABLE 1. Comparison of the results

there is no friction or bottom slope. Flow conditions are chosen in such a way that a *hydraulic jump* occurs and does not move in the channel. The following formula is used in order to prepare the required conditions. $y_i = \frac{y_s}{2} (\sqrt{1 + 8F_s^2} - 1)$, $F_s = \frac{V_s}{\sqrt{gy_s}}$ where y_i is the water depth at which the jump starts and y_s is the downstream water depth (sequent depth), F_s is the Froude number downstream, V_s is the velocity downstream, g is gravitational acceleration. A water depth of 5.5m is used as a downstream boundary condition. A constant unit discharge of $20 m^2/s$ in the x direction and 1.987m water depth are applied as an upstream boundary condition. The value of the upstream water depth is computed using the above formula for a *hydraulic jump* in a channel. In fact, it is a required depth at which *hydraulic jump* starts for given conditions. Figure 1 shows the initial flow conditions, the water profile after 500 and 4000s respectively. As can be seen, the *jump* is sustained between 320 and 330m as expected.

Backwater Tests: The aim of this test is to reproduce steady non-uniform flow in a channel with mild slope. As the depth at the downstream boundary is greater than the uniform flow depth, the expected flow profile is a *backwater* curve. The computational domain is divided into 130 cells in the x direction and 8 cells in the y direction. Each cell has a length of 10m in the x and y directions. While Manning's coefficient, n , is kept constant as 0.02, a constant slope in the x direction only is taken to be 0.002. Initial flow condition is shown in Figure 2a. A constant unit discharge of $15 m^2/s$ in the x direction and 3.4m water depth are applied as upstream and downstream boundary conditions respectively. Figure 2b shows the occurrence of a *backwater* curve.

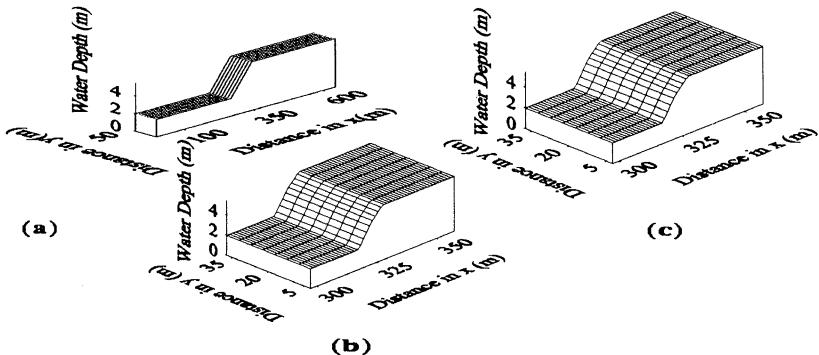


Figure 1. (a) Initial flow condition and (b-c) Water Profile after 500 and 4000s respectively.

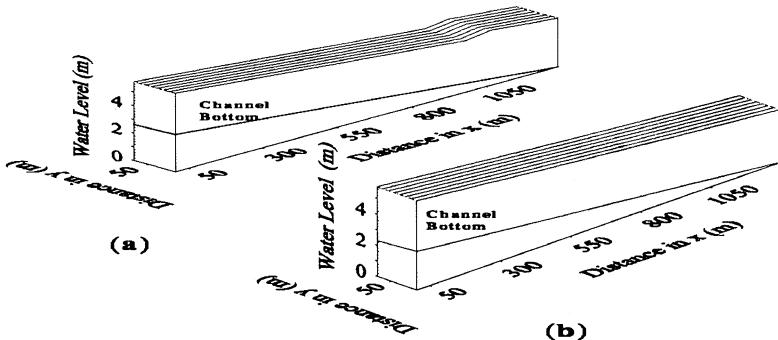


Figure 2. (a) Initial Flow Condition and (b) Backwater profile in the channel

4. Conclusion

It has been shown that the model can be used for a variety of flow types. The results have shown the model's applicability to real engineering problems. Although the results presented show steady flow conditions, they were achieved through unsteady flow simulation with arbitrary initial conditions. These tests were chosen because their results could easily be checked. Since the model uses an *explicit scheme*, small time steps are required resulting

in long simulations. In order to eliminate memory overload, the results for each time step are stored in a file and only one previous time step is kept in the memory. This also extends the simulation time. Refining the computational grid gives better results but also increases simulation times. The differences between the results of the model and the results obtained using Manning's formula increase if the ratio of water depth to channel width increases due to the absence of side friction effects in the model. However, it is clear that the wider the channel, the smaller the differences.

Although the *split technique* works well in these examples, the authors observed that in the case of steady or static flows errors are generated when the channel has a variable bottom topography. These errors become significant when there is a significant variation in the bottom slope especially at the boundaries. The failure of the split technique for these cases is well explained in the literature (LeVeque, 1998). Briefly, it is caused by an imbalance between the *source terms* and the fluxes.

References

- Alcrudo F and Garcia-Navarro P (1993). A High-Resolution Godunov-Type Scheme in Finite Volumes For the 2D Shallow-Water Equations. *International Journal For Numerical Methods in Fluids*, **16**, pp.489-505.
- Bermúdez A, Dervieux A, Desideri J-A and Vázquez-Cendón M E (1998). Upwind Schemes For The two-dimensional Shallow Water Equations With Variable Depth Using Unstructured Meshes. *Computer Methods In Applied Mechanics And Engineering*, **155**:1-2, pp.49-72.
- Chippada S, Dawson C N, Martinez M L and Wheler M F (1998). A Godunov-Type Finite Volume Method For The System Of Shallow Water Equations. *Computer Methods in Applied Mechanics and Engineering*, **151**:1-2, pp.105-129.
- Hu K, Mingham C G and Causon D M (1998). A Bore-Capturing Finite Volume Method For Open-Channel Flows. *International Journal For Numerical Methods in Fluids*, **28**, pp.1241-1261.
- LeVeque R J (1998). Balancing Source Terms and Flux Gradients in High-resolution Godunov Methods: the Quasi-steady Wave-propagation Algorithm. *J. Comput. Phys.* **146**, pp 346-365.
- Mingham C G and Causon D M (1998). High - Resolution Finite - Volume Method For Shallow Water Flows. *Journal of Hydraulic Engineering*, **124**:6, pp.605-614.
- Toro E F (1997). Riemann Solvers and Numerical Methods for Fluid Dynamics. First Edition, Springer-Verlag.
- Vázquez-Cendón M E (1999). Improved Treatment of Source Terms in Upwind Schemes for the Shallow Water Equations with Irregular Geometry. *J. Comput. Phys.* **148**, pp 497-526.
- Zhao D H, Shen H W, Tabios III G Q, Lai J S and Tan W Y (1994). Finite - Volume Two - Dimensional Unsteady - Flow Model For River Basins. *Journal of Hydraulic Engineering*, **120**:7, pp.863-882.
- Zoppou C and Roberts S (1999). Catastrophic Collapse of Water Supply Reservoirs in Urban Areas. *Journal of Hydraulic Engineering*, **125**:7, pp.686-695.

A-PRIORI ESTIMATES FOR A SEMI-LAGRANGIAN SCHEME FOR THE WAVE EQUATION

M. FALCONE

*Dipartimento di Matematica,
Università di Roma "La Sapienza",
P. Aldo Moro 2, 00185 Roma - Italy
Email: falcone@caspur.it*

AND

R. FERRETTI

*Dipartimento di Matematica,
Università di Roma "Tor Vergata",
Via della Ricerca Scientifica, 00133 Roma - Italy
Email: ferretti@mat.uniroma2.it*

Abstract. We present some *a-priori* estimates for a class of semi-Lagrangian approximation schemes for the wave equation. The wave equation is written in the form of a hyperbolic system of the first order and the approximation is based on this representation. The algorithm can work on structured and unstructured grids.

1. Introduction

Several approximation schemes for the wave equation have been proposed using different numerical techniques (mainly finite differences, finite elements, spectral methods, see e.g. (Quarteroni and Valli, 1994)). In some typical applications of the wave equation (e.g. in geophysics) the problem becomes quite difficult since the coefficients can be non differentiable or even discontinuous, and this feature considerably reduces the accuracy of the above mentioned techniques. In those situations it could be convenient to write the wave equation as a first order system in order to apply one of the schemes which have been developed for first order problems and

have shown to be efficient for the approximation of discontinuous solutions. Although the above equivalence is classical, a careful choice of the new variables is necessary to reduce the size of the system, particularly for problems in dimension 2 and 3. In (Hoar and Vogel, 1995) that approach has been already used for the numerical approximation of elastic waves. Their finite difference approximation is based on a local ENO interpolation (cfr. (Harten, Osher, Engquist and Chakravarthy, 1986)) and has proved to be quite effective.

Our goal here is twofold. First we want to sketch a semi-Lagrangian scheme for the hyperbolic system related to the wave equation. Then, we want to extend to that system the error analysis developed for the linear advection equation in (Falcone and Ferretti, 1998), showing that in general this approach allows larger time steps and may reduce numerical viscosity. A more detailed analysis of that scheme as well a discussion of some technical issues, such as the (accurate) approximation of boundary conditions, will be developed in a forthcoming paper (Falcone and Ferretti, 2000). It should be noted that the scheme presented here can also be applied to non linear equations of Hamilton-Jacobi type (see f.e. (Falcone and Ferretti, 1994)) and it can be interpreted as a discrete version of the Hopf-Lax formula for the exact (viscosity) solution (Barles, 1994). In this respect the scheme belongs to the class of generalized Godunov schemes for hyperbolic problems.

2. The model problem

Let us examine the relationship between the wave equation and the hyperbolic system. For simplicity we limit ourselves to the one dimensional case, for the results in higher dimension see (Falcone and Ferretti, 2000).

Let us start with the Cauchy problem for the wave equation in \mathbb{R} ,

$$\begin{cases} \rho(x)u_{tt} - \frac{\partial}{\partial x}(k(x)u_x) = f, & -\infty < x < +\infty, t > 0 \\ u(x, 0) = u_0(x), & -\infty < x < +\infty \\ u_t(x, 0) = u_1(x), & -\infty < x < +\infty \end{cases} \quad (1)$$

where u and f depend on (x, t) . A large amount of results for the existence and uniqueness of solutions is available provided the coefficients $k(x)$ and $\rho(x)$ are regular. However, as we mentioned the above equation is also a good model for applications where the physical parameters (the density of the media and the speed of propagation of the wave) have abrupt changes. This situation occurs e.g. in geophysics and requires the use of generalized solutions (e.g. in the viscosity sense, cfr (Crandall, Ishii and Lions, 1992) and (Barles, 1994)).

It is well known (cfr. (Whitham, 1974)) that the wave equation is equivalent to the following first order system

$$\begin{cases} \frac{\partial}{\partial t}v_1 + \lambda_1(x)\frac{\partial}{\partial x}v_1 = f_1(x, t; \mathbf{v}), & x \in \mathbf{R}, \quad t \in (0, T) \\ \frac{\partial}{\partial t}v_2 + \lambda_2(x)\frac{\partial}{\partial x}v_2 = f_2(x, t; \mathbf{v}), & x \in \mathbf{R}, \quad t \in (0, T) \\ \mathbf{v}(x, 0) = \boldsymbol{\varphi}(x), & x \in \mathbf{R} \end{cases} \quad (2)$$

where $\mathbf{v}(x, t) = (v_1(x, t), v_2(x, t))^t$, $\boldsymbol{\varphi}(x) = (\varphi_1(x), \varphi_2(x))^t$ is the initial data and $f_i(x, t; \mathbf{v}(x, t)) = a_i(x)v_1(x, t) + b_i(x)v_2(x, t) + c_i(x, t)$. For completeness, let us mention the relationships between the coefficients appearing in (2) and the functions $\rho(\cdot)$, $k(\cdot)$ e $f(\cdot, \cdot)$. The system corresponding to (1) can be written in matrix form as

$$B(x)\frac{\partial}{\partial t}\mathbf{z}(x, t) + A(x)\frac{\partial}{\partial x}\mathbf{z}(x, t) + C(x)\mathbf{z}(x, t) = \mathbf{d}(x, t) \quad (3)$$

where $\mathbf{z}(x, t) = (u_x(x, t), u_t(x, t))^t$, $\mathbf{d}(x, t) = (0, f(x, t))^t$ and

$$B(x) = \begin{pmatrix} 1 & 0 \\ 0 & \rho(x) \end{pmatrix}, \quad A(x) = \begin{pmatrix} 0 & -1 \\ -k(x) & 0 \end{pmatrix}, \quad C(x) = \begin{pmatrix} 0 & 0 \\ -k'(x) & 0 \end{pmatrix}$$

It can be shown that the solution of (3) is $\mathbf{z}(x, t) = H^{-1}(x)\mathbf{v}(x, t)$ where

$$H = H(x) \equiv \begin{pmatrix} -\sqrt{k(x)} & \sqrt{\rho(x)} \\ \sqrt{k(x)} & \sqrt{\rho(x)} \end{pmatrix}$$

We can write it componentwise obtaining

$$\begin{cases} u_x(x, t) = \frac{1}{2\sqrt{k(x)}}(-v_1(x, t) + v_2(x, t)) \\ u_t(x, t) = \frac{1}{2\sqrt{\rho(x)}}(v_1(x, t) + v_2(x, t)) \end{cases} \quad (4)$$

where v_1 e v_2 are solutions of (2).

3. The semi-Lagrangian scheme

The discretization will be constructed in two steps. First we discretize time and integrate along characteristics, then we project the semi-discrete scheme on a space grid.

In order to simplify notations we will drop out in the sequel the sub-index i , i.e. we will write v instead of v_i (the same for λ_i , f_i , φ_i , etc.) and a relation valid for v (respectively λ , f , φ , etc.) is intended to be valid for each component v_i .

3.1. TIME DISCRETIZATION AND ESTIMATES

Let us assume that the functions $\lambda : \mathbb{R} \rightarrow \mathbb{R}$ and $f : \mathbb{R} \times (0, T) \times \mathbb{R}^2 \rightarrow \mathbb{R}$ are bounded and Lipschitz continuous, that is:

$$\|\lambda\|_\infty \leq M_\lambda \quad (5)$$

$$|\lambda(x) - \lambda(y)| \leq L_\lambda |x - y|, \quad \forall x, y \in \mathbb{R} \quad (6)$$

$$\|f(\cdot, t; \mathbf{v})\|_\infty \leq M_f \quad (7)$$

$$|f(x, t; \mathbf{v}) - f(y, t; \mathbf{z})| \leq L_f(|x - y| + |\mathbf{v} - \mathbf{z}|), \quad \forall x, y \in \mathbb{R}, \quad \forall \mathbf{v}, \mathbf{z} \in \mathbb{R}^2 \quad (8)$$

We define $\gamma(s) \equiv v(y(\Delta t - s, s))$. For the interpretation of the scheme it will be useful to consider the system of characteristics

$$\begin{cases} \dot{y}(s) = -\lambda(y(s)) \equiv \mu(y(s)) \\ \dot{\gamma}(s) = f(y(\Delta t - s), s; \mathbf{v}(y(\Delta t - s), s)) \\ y(0) = x, \quad \gamma(0) = \varphi(y(\Delta t)) \end{cases} \quad (9)$$

Note that the above Cauchy problem has a unique global solution since we have assumed (5)-(8).

Let us define an approximation for $y(\Delta t)$ and $\gamma(\Delta t)$, which we will denote by $\hat{y}(\Delta t)$ and $\hat{\gamma}(\Delta t)$. That approximation simply correspond to a one-step approximation scheme for the Cauchy problem (9):

$$\begin{cases} \hat{y}(\Delta t) = x + \Delta t \Phi_\mu(x) \\ \hat{\gamma}(\Delta t) = \varphi(\hat{y}(\Delta t)) + \Delta t \Phi_f(\hat{y}(\Delta t), 0, \varphi(\hat{y}(\Delta t))) \end{cases} \quad (10)$$

where Φ_μ e Φ_f are the Henrici functions of the one-step scheme.

In order to simplify notations, we will denote by $z \equiv x + \Delta t \Phi_\mu(x)$ the foot of the characteristic $y(t)$ passing through the point $(x, \Delta t)$.

Substituting the first equation in the second, we get an analytic expression for $\hat{\gamma}(\Delta t)$:

$$\hat{\gamma}(\Delta t) = \varphi(z) + \Delta t \Phi_f(z, 0, \varphi(z))$$

By the definition of $\hat{\gamma}(\Delta t)$ we get the following approximation of v :

$$\hat{v}(x, \Delta t) = \varphi(z) + \Delta t \Phi_f(z, 0, \varphi(z)) \quad (11)$$

In general, at time $t_n = n\Delta t$, we have

$$\begin{cases} \hat{v}(x, t_n) = \hat{v}(z, t_{n-1}) + \Delta t \Phi_f(z, t_{n-1}, \hat{v}(z, t_{n-1})) \\ \hat{v}(x, 0) = \varphi(x) \end{cases} \quad (12)$$

In the sequel, we will always assume that the one-step scheme is (at least) consistent and that Φ_f is Lipschitz continuous, that is:

$$\lim_{\Delta t \rightarrow 0} \Phi_\mu(x, \Delta t) = \mu(x) \quad (13)$$

$$\lim_{\Delta t \rightarrow 0} \Phi_f(x, t, \mathbf{v}, \Delta t) = f(x, t, \mathbf{v}) \quad (14)$$

$$|\Phi_f(x, t, \mathbf{v}) - \Phi_f(y, t, \mathbf{z})| \leq L_\Phi(|x - y| + |\mathbf{v} - \mathbf{z}|) \quad (15)$$

Moreover, we are interested in high-order methods so we will assume that the numerical schemes for (12) are of order p , which means that:

$$|y(\Delta t) - x - \Delta t \Phi_\mu(x)| \leq C \Delta t^{p+1} \quad (16)$$

$$\left| \int_0^{\Delta t} f(y(s), t - s; \mathbf{v}(y(s), t - s)) ds + \right. \\ \left. - \Delta t \Phi_f(y(\Delta t), t - \Delta t, \mathbf{v}(y(\Delta t), t - \Delta t)) \right| \leq C \Delta t^{p+1} \quad (17)$$

It is worthwhile to note that the class of semi-Lagrangian schemes we are considering includes any explicit one-step method used to follow the characteristics, some typical examples of the resulting schemes for a single advection equation can be found in (Falcone and Ferretti, 1998).

Let us prove the convergence of our scheme in the $L^\infty(I)$ norm, where $I \subseteq \mathbb{R}$. Such norm will be denoted by $\|\cdot\|_\infty$.

Theorem 1. *Let us assume that (5)–(8), (13)–(17) hold true and let $\varphi \in L^\infty(\mathbb{R})$. Then, for any $\bar{t} \in (0, t)$ and Δt sufficiently small, there exists a positive constant C such that:*

$$\|\hat{\mathbf{v}}(\bar{t}) - \mathbf{v}(\bar{t})\|_\infty \leq C \Delta t^p \quad (18)$$

Proof. Let us start comparing term by term the equation and its discrete time approximation. We get,

$$v(x, t_n) - \hat{v}(x, t_n) = v(y(\Delta t), t_{n-1}) - \hat{v}(z, t_{n-1}) + \quad (19) \\ + \int_0^{\Delta t} f(y(s), t_n - s; \mathbf{v}(y(s), t_n - s)) ds - \Delta t \Phi_f(z, t_{n-1}, \hat{\mathbf{v}}(z, t_{n-1})) ds$$

Obviously $v(x, 0) - \hat{v}(x, 0) = 0$. The hypotheses (7), (8) and the fact that φ is bounded guarantee that v is Lipschitz continuous (we will denote by L_v its constant). We can write then,

$$|v(x, t_n) - \hat{v}(x, t_n)| \leq |v(y(\Delta t), t_{n-1}) - \hat{v}(z, t_{n-1})| + \quad (20) \\ + \left| \int_0^{\Delta t} f(y(s), t_n - s; \mathbf{v}(y(s), t_n - s)) ds - \Delta t \Phi_f(z, t_{n-1}, \hat{\mathbf{v}}(z, t_{n-1})) ds \right|$$

Let us find a bound for the integral term.

$$\begin{aligned}
& \left| \int_0^{\Delta t} f(y(s), t_n - s; \mathbf{v}(y(s), t_n - s)) ds - \Delta t \Phi_f(z, t_{n-1}, \hat{\mathbf{v}}(z, t_{n-1})) \right| \leq \\
& \leq \left| \int_0^{\Delta t} f(y(s), t_n - s; \mathbf{v}(y(s), t_n - s)) ds + \right. \\
& \quad \left. - \Delta t \Phi_f(y(\Delta t), t_{n-1}, \mathbf{v}(y(\Delta t), t_{n-1})) \right| + \\
& \quad + \Delta t |\Phi_f(y(\Delta t), t_{n-1}, \mathbf{v}(y(\Delta t), t_{n-1})) - \Phi_f(z, t_{n-1}, \hat{\mathbf{v}}(z, t_{n-1}))| \leq \\
& \leq C \Delta t^{p+1} + \Delta t L_{\Phi_f}(|y(\Delta t) - z| + |\mathbf{v}(y(\Delta t), t_{n-1}) - \hat{\mathbf{v}}(z, t_{n-1})|) \leq \\
& \leq C \Delta t^{p+1} + \Delta t L_{\Phi_f} |\mathbf{v}(y(\Delta t), t_{n-1}) - \hat{\mathbf{v}}(z, t_{n-1})|.
\end{aligned} \tag{21}$$

Since $|v(x, t) - \hat{v}(x, t)| \leq |\mathbf{v}(x, t) - \hat{\mathbf{v}}(x, t)|$, we obtain then:

$$\begin{aligned}
& |v(x, t_n) - \hat{v}(x, t_n)| \leq C \Delta t^{p+1} + \\
& \quad + (1 + \Delta t L_{\Phi}) |\mathbf{v}(y(\Delta t), t_{n-1}) - \hat{\mathbf{v}}(z, t_{n-1})| \leq \\
& \leq C \Delta t^{p+1} + (1 + \Delta t L_{\Phi}) |\mathbf{v}(y(\Delta t), t_{n-1}) - \mathbf{v}(z, t_{n-1})| + \\
& \quad + (1 + \Delta t L_{\Phi}) |\mathbf{v}(z, t_{n-1}) - \hat{\mathbf{v}}(z, t_{n-1})| \leq \\
& \leq C \Delta t^{p+1} + (1 + \Delta t L_{\Phi}) L_{\mathbf{v}} |y(\Delta t) - z| + \\
& \quad + (1 + \Delta t L_{\Phi}) \|\mathbf{v}(t_{n-1}) - \hat{\mathbf{v}}(t_{n-1})\|_{\infty} \leq \\
& \leq C \Delta t^{p+1} + (1 + \Delta t L_{\Phi}) \|\mathbf{v}(t_{n-1}) - \hat{\mathbf{v}}(t_{n-1})\|_{\infty}.
\end{aligned} \tag{22}$$

Taking the supremum, we obtain:

$$\begin{aligned}
& \|\mathbf{v}(t_n) - \hat{\mathbf{v}}(t_n)\|_{\infty} \leq C \Delta t^{p+1} + \|\mathbf{v}(t_{n-1}) - \hat{\mathbf{v}}(t_{n-1})\|_{\infty} \leq \\
& \leq 2C \Delta t^{p+1} + (1 + \Delta t L_{\Phi})^2 \|\mathbf{v}(t_{n-2}) - \hat{\mathbf{v}}(t_{n-2})\|_{\infty} \leq \dots \leq \\
& \leq nC \Delta t^{p+1}.
\end{aligned} \tag{23}$$

Setting $\bar{t} = n \Delta t$, we get the estimate (18).

3.2. SPACE DISCRETIZATION AND ESTIMATES

We will describe the space discretization in an infinite strip, $(-\infty, +\infty) \times (0, T)$ assuming that the initial data have compact support. This allows us to avoid the treatment of boundary conditions.

We want to write w has a linear combination of some basis functions, e.g.

$$w(x, t_n) = \sum_{j \in \mathbb{Z}} \hat{v}(x_j, t_n) \psi_j^n(x) \tag{24}$$

Here $\{\psi_j(x)\}_{j \in Q}$ is a basis of bounded functions and we will assume that, for any $j \in \mathbb{Z}$,

$$\psi_j(x_k) = \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases} \quad (25)$$

Although the sum is done over the indices belonging to \mathbb{Z} it includes only a finite number of nonzero terms because the solution has compact support. Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a generic function, we can use the basis $\{\psi_j(x)\}_{j \in Q}$ to define the projection operator P_h

$$P_h g(x) \equiv \sum_{j \in \mathbb{Z}} g(x_j) \psi_j(x) \quad (26)$$

Naturally substituting g by its projection we introduce an error which depends on the regularity of g and on the actual definition of the projection operator P_h . We will assume that for any $g \in W^{1,\infty}(\mathbb{R})$,

$$\lim_{\Delta x \rightarrow 0} \|g - P_h g\|_\infty = 0 \quad (27)$$

We can finally write the fully discrete scheme:

$$\begin{cases} w_j^n = \sum_{m \in \mathbb{Z}} w_m^{n-1} \psi_m(z_j) + \Delta t \Phi_f(z_j, t_{n-1}, \mathbf{w}(z_j, t_{n-1})) \\ v_j^0 = \varphi(x_j) \end{cases} \quad (28)$$

where $z_j \equiv x_j + \Delta t \Phi_\mu(x_j)$.

We just remark that f depends on v_1 e v_2 , so that in $\Phi_f(x, t, \mathbf{w})$ we need to reconstruct also the values of $w_i(x, t)$, $i = 1, 2$, by P_h . At this stage, we can choose among several local reconstructions and this choice gives rise to several semi-Lagrangian schemes belonging to the same class. The list includes naturally polynomial interpolations (as for the finite element spaces P_1, P_2) but does not include ENO interpolation because ENO schemes use an adaptive stencil (which depends on the values of the divided differences).

For computational purposes it is useful to write the scheme in matrix form. Let M be the total number of time steps, i.e. our time horizon will be $T = M \Delta t$. Define $Q \equiv \{1, \dots, N\}$ as:

$$Q \equiv \cup_{n=1}^M Q_n$$

where Q_n is the set of nodes which the scheme uses to compute the solution at time t_n . Let us introduce the following notations:

$$W^n \equiv (w_1^n, \dots, w_N^n)^t, \quad (29)$$

$$F^n(W^n) \equiv (\Delta t \Phi_f(z_1, t_n, \mathbf{w}), \dots, \Delta t \Phi_f(z_N, t_n, \mathbf{w}))^t, \quad (30)$$

$$\Psi \equiv \begin{pmatrix} \psi_1(x_1 + \Delta t \Phi_\mu(x_1)) & \dots & \psi_N(x_1 + \Delta t \Phi_\mu(x_1)) \\ \vdots & & \vdots \\ \psi_1(x_N + \Delta t \Phi_\mu(x_N)) & \dots & \psi_N(x_N + \Delta t \Phi_\mu(x_N)) \end{pmatrix} \quad (31)$$

Then we can write (28) in a more compact form, as

$$\mathbf{W}^{n+1} = \mathbf{F}^n(\mathbf{W}^n) + \Upsilon \mathbf{W}^n \quad (32)$$

where $\mathbf{W}^n = (W_1^n, W_2^n)^t$, $\mathbf{F}^n = (F_1^n, F_2^n)^t$ and Υ is the block matrix

$$\Upsilon = \begin{pmatrix} \Psi_1 & 0 \\ 0 & \Psi_2 \end{pmatrix}.$$

The following error estimate for the fully discrete scheme is proved in (Falcone and Ferretti, 2000).

Theorem 2. *Let us assume that (5)–(8) and (13)–(17) hold true. Moreover, let $\varphi \in W^{1,\infty}(a,b)$ and $\|v(t) - P_h v(t)\|_\infty \leq E(\Delta x)$ for any $t \in [0, \bar{t}]$ and $\sum_{j \in Q} |\psi_j(x)| \leq 1$. Then, for any Δt sufficiently small there exists a positive constant C such that,*

$$\|\mathbf{w}(\bar{t}) - \mathbf{v}(\bar{t})\|_\infty \leq C \left(\Delta t^p + \frac{E(\Delta x)}{\Delta t} \right) \quad (33)$$

Acknowledgements

This work has been supported by the MURST National Project "Metologie Numeriche Avanzate per il Calcolo Scientifico".

References

- Crandall M.G., Ishii H. and Lions P.L. (1992). User's guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc.* **27**, pp.1-67.
- Falcone M. and Ferretti R. (1994). Discrete time high-order schemes for viscosity solutions of Hamilton–Jacobi–Bellman equations. *Numer. Math.* **67**, pp. 315–344.
- Falcone M. and Ferretti R. (1998). Convergence analysis for a class of semi-lagrangian advection schemes. *SIAM J. Num. Anal.* **35**, pp. 909–940.
- Falcone M. and Ferretti R. (2000). Analysis of a class of semi-Lagrangian schemes for the wave equation. In preparation.
- Harten A., Osher S., Engquist B. and Chakravarthy S. (1986). Some results on uniformly high-order accurate essentially non-oscillatory schemes. *Appl. Numer. Math.* **2**, pp. 347–377.
- Hoar R.H. and Vogel C.R. (1995). An adaptive stencil finite differencing scheme for linear first order hyperbolic systems – a preliminary report. Computation and control, IV (Bozeman, MT, 1994), 169–183, Progr. Systems Control Theory, 20, Birkhäuser Boston, Boston, MA, 1995.
- Barles G. (1994). Solutions de viscosité des équations de Hamilton-Jacobi. Springer-Verlag, Paris.
- Quarteroni A. and Valli A. (1994). Numerical approximation of partial differential equations. Springer-Verlag.
- Whitham G.B. (1974). Linear and non linear waves. Wiley-Interscience, New York.

INTERSTELLAR SHOCK STRUCTURES IN WEAKLY IONISED GASES

S.A.E.G. FALLE

Department of Applied Mathematics

University of Leeds

Leeds LS2 9JT, UK

Email: sam@amsta.leeds.ac.uk

Abstract. It is well known that relaxation effects can ensure a smooth shock structure in which viscosity is negligible and that the existence, or otherwise, of such a structure depends only upon the Mach numbers in the upstream and downstream state (Whitham, 1959), (Whitham, 1974). However, this analysis is restricted to cases in which there is only one relaxation length scale, whereas in this paper we show that if there is more than one such scale, then it is possible for there to be a viscous subshock with a downstream sonic point in cases for which the Whitham criterion would predict a smooth relaxation shock structure. This is of relevance to shocks in dense interstellar clouds.

1. Introduction

In dense interstellar clouds the degree of ionization is often extremely small, but, despite this, the charged component of the plasma cannot be ignored because it couples strongly to a magnetic field that is generally strong enough to be dynamically significant. Under these conditions, the collision rate between the neutral and charged components is sufficiently small for it to be necessary to consider these as distinct, but interpenetrating fluids (Draine and McKee, 1993). In the simplest case, the neutral fluid obeys the ordinary gas dynamic equations, whereas the charged fluid behaves like an ideal conducting fluid. These two fluids are coupled together by source terms that describe the exchange of momentum and energy.

It is well known (Marshall, 1955), (Whitham, 1959), (Whitham, 1974), (Liubarski, 1961), that in such cases there exist shock structures in which

viscosity is negligible and the dissipation is entirely due relaxation effects. (Whitham, 1974) considers a number of different cases and shows that the existence, or otherwise of a viscous sub-shock depends only upon the upstream and downstream Mach numbers with respect to the wavespeeds of the decoupled system. However, in all these cases there was only a single length scale for relaxation, whereas in interstellar shocks there are several different length scales. We shall see that, if the length scale for radiative cooling is neither very large nor very small compared to the other length scales, a sub-shock can occur in cases for which the Whitham criterion would predict that it should not. This is simply because the addition of a cooling term allows a smooth subsonic to supersonic transition within the shock structure.

2. Equations of Two Fluid Magnetohydrodynamics

For our purposes it is sufficient to consider the one dimensional equations for the two fluids. These are

Neutral Fluid:

$$\frac{\partial \mathbf{U}_n}{\partial t} + \frac{\partial \mathbf{F}_n}{\partial x} = \mathbf{S}_n ,$$

where

$$\mathbf{U}_n = \begin{bmatrix} \rho_n \\ \rho_n v_n \\ e_n \end{bmatrix} \quad \mathbf{F}_n = \begin{bmatrix} \rho_n v_n \\ p_n + \rho_n v_n^2 \\ (e_n + p_n)v_n \end{bmatrix} \quad \mathbf{S}_n = \begin{bmatrix} 0 \\ f \\ -H + fv_n - L_n \end{bmatrix}$$

$$e_n = \frac{p_n}{(\gamma - 1)} + \frac{1}{2}\rho_n v_n^2 .$$

Conducting Fluid

$$\frac{\partial \mathbf{U}_c}{\partial t} + \frac{\partial \mathbf{F}_c}{\partial x} = \mathbf{S}_c$$

where

$$\mathbf{U}_c = \begin{bmatrix} \rho_c \\ \rho_c v_c \\ e_c \\ B_y \end{bmatrix} \quad \mathbf{F}_c = \begin{bmatrix} \rho_c v_c \\ p_c + B_y^2 + \rho_c v_c^2 \\ (e_c + p_c + B_y^2)v_c \\ v_c B_y \end{bmatrix} \quad \mathbf{S}_c = \begin{bmatrix} 0 \\ -f \\ H - fv_c - L_c \\ 0 \end{bmatrix}$$

$$e_c = \frac{p_c}{(\gamma - 1)} + \frac{B_y^2}{2} + \frac{1}{2}\rho_c v_c^2 .$$

Here f is the interaction force given by

$$f = K_m \rho_n \rho_c (v_n - v_c) ,$$

the heat transfer rate between the two fluids, H , is given by

$$H = K_t \rho_n \rho_c (T_n - T_c)$$

and the radiative cooling is given by

$$L = K_r \rho^2 (T^4 - T_e^4) .$$

K_m , K_t and K_r are constants, T_n , T_c are the temperatures of the fluids and T_e is a constant equilibrium temperature.

For disturbances with very short wavelengths, the source terms are negligible and we therefore have a system in which the two fluids are decoupled

$$\frac{\partial \mathbf{U}_n}{\partial t} + \frac{\partial \mathbf{F}_n}{\partial x} = 0 , \quad \frac{\partial \mathbf{U}_c}{\partial t} + \frac{\partial \mathbf{F}_c}{\partial x} = 0 ,$$

which we will call the frozen system.

The wavespeeds of this system are the eigenvalues of the Jacobian matrices

$$\mathbf{J}_n = \frac{\partial \mathbf{F}_n}{\partial \mathbf{U}_n} , \quad \mathbf{J}_c = \frac{\partial \mathbf{F}_c}{\partial \mathbf{U}_c} ,$$

which are

$$\text{NeutralFluid } (\mathbf{J}_n) \quad \lambda_1 = v_n , \quad \lambda_{2,3} = v_n \mp a_n , \quad a_n^2 = (\gamma p_n)/\rho_n ,$$

$$\text{ConductingFluid } (\mathbf{J}_c) \quad \lambda_{1,2} = v_c , \quad \lambda_{3,4} = v_c \mp c_c , \quad c_c^2 = a_c^2 + B_y^2/\rho_c .$$

For long wavelength disturbances, the system is in equilibrium with $\mathbf{S} = 0$. In that case we have $v_n = v_c = v$, $T_n = T_c = T_e$ and the system reduces to

$$\frac{\partial \mathbf{U}_e}{\partial t} + \frac{\partial \mathbf{F}_e}{\partial x} = 0 ,$$

where

$$\mathbf{U}_e = \begin{bmatrix} \rho_n \\ \rho_c \\ \rho v \\ B_y \end{bmatrix} , \quad \mathbf{F}_e = \begin{bmatrix} \rho_n v \\ \rho_c v \\ p + B_y^2/2 + \rho v^2 \\ v B_y \end{bmatrix} , \quad (1)$$

$$\rho = \rho_n + \rho_c , \quad p = a_e^2 \rho .$$

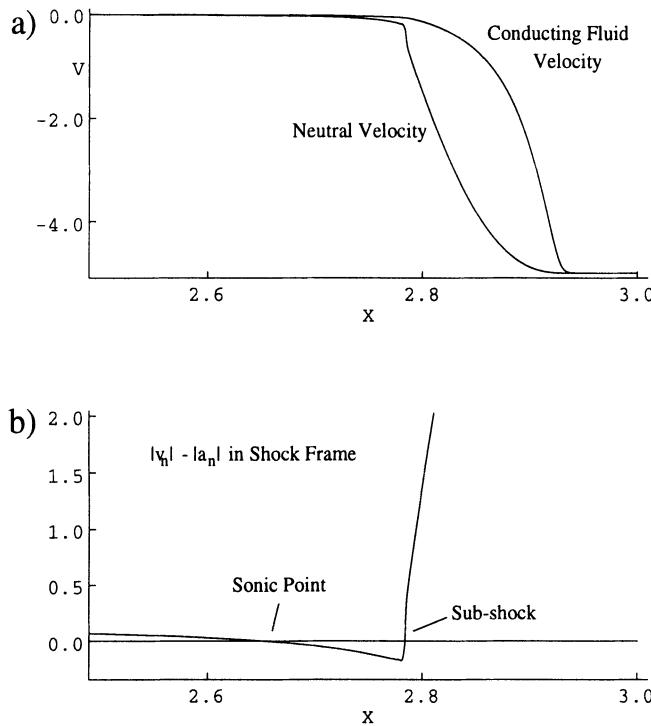


Figure 1. Shock structure with slow cooling and a viscous sub-shock in the neutral fluid. a) Neutral and conducting fluid velocities in the piston frame, b) wavespeed of neutral sound waves in the shock frame.

The wavespeeds of this system are

$$\lambda_1 = v, \quad \lambda_{2,3} = v \mp c_e, \quad c_e^2 = a_e^2 + B_y^2/\rho.$$

Whitham (Whitham, 1974) shows that linear waves are stable if the wavespeeds of the frozen and equilibrium systems interleave, which requires

$$c_c > c_e > a_n. \quad (2)$$

3. Relaxation Shock Structure

It is clear that, on the large scale, the system reduces to the equilibrium system given by (1). Suppose we have a shock associated with one of the non-linear fields, λ_2 or λ_3 , of this system and wish to know whether such a shock can have a continuous structure in which viscous effects are unimportant. The equations for the shock structure are obtained by transforming

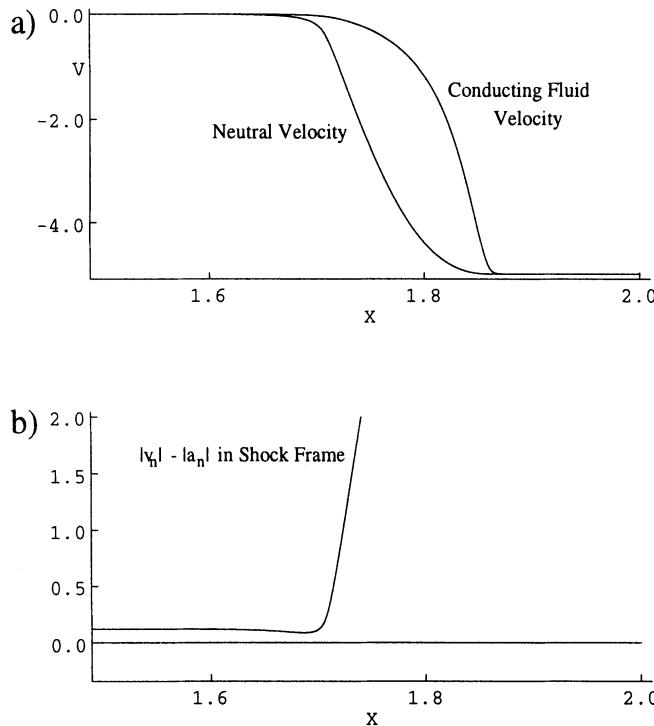


Figure 2. Shock structure with fast cooling and no viscous sub-shock in the neutral fluid. a) Neutral and conducting fluid velocities in the piston frame, b) wavespeed of neutral sound waves in the shock frame.

to a frame moving with the shock and discarding the time derivatives to get

$$\frac{d\mathbf{F}}{dx} = \mathbf{J} \frac{d\mathbf{U}}{dx} = \mathbf{S}, \quad (3)$$

where

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_n \\ \mathbf{U}_c \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} \mathbf{F}_n \\ \mathbf{F}_c \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} \mathbf{S}_n \\ \mathbf{S}_c \end{bmatrix}.$$

A steady shock structure is a solution to (3) that satisfies the boundary conditions

$$\mathbf{U} \rightarrow \begin{cases} \mathbf{U}_{er} & \text{as } x \rightarrow \infty \\ \mathbf{U}_{el} & \text{as } x \rightarrow -\infty \end{cases},$$

where \mathbf{U}_{er} , \mathbf{U}_{el} satisfy the equilibrium shock relations

$$\mathbf{F}_e(\mathbf{U}_{er}) = \mathbf{F}_e(\mathbf{U}_{el}).$$

A necessary condition for the existence of a shock structure of this type can be obtained from the following argument. The shock must satisfy the evolutionary condition for the equilibrium system. If the shock is moving to the right relative to the fluid, then this condition is

$$|v| \leq c_{ef} \text{ in the left state, } |v| \geq c_{ef} \text{ in the right state.} \quad (4)$$

In addition, since the eigenvalues of the matrix, J in (3), are those of the frozen system, these equations will, in general, become singular if any of the frozen wavespeeds change sign within the shock structure. Finally, we must have $|v| \leq c_c$ in the upstream state, otherwise there will be a sub-shock in the conducting fluid.

Together with the interleaving condition (2), this gives the Whitham criterion

Downstream	Shock	Upstream
$c_c > c_e > v > a_n$	$c_c > v > c_e > a_n$	

If this is satisfied, then it is possible to have a relaxation shock structure in which viscosity and heat conduction are unimportant, whereas if it is not, then there must be a viscous sub-shock. This condition is clearly necessary because, if it is violated, then interleaving condition and the evolutionary condition together imply that there must be a transition from super to subsonic with respect to one of the frozen wavespeeds, which can only occur in a sub-shock. Whitham argues that it is also sufficient on the grounds that, there were a sub-shock, then the flow must remain subsonic downstream of the sub-shock.

However, this assumes that there cannot be a smooth transition from sub to supersonic within the shock structure, whereas Liubarskii (Liubarski, 1961) points out that such a transition is possible provided that $J^{-1}S$ is regular at the sonic point. The Whitham criterion is sufficient if there is only one relaxation scale since in that case the only dimensionless quantities are the upstream and downstream Mach numbers. But in our case there are additional dimensionless parameters, the most important of which is, α , the ratio of the length scale for cooling to that for momentum exchange between the two fluids. Since $\alpha \propto K_m/K_r$, it does not depend only upon the upstream and downstream states.

In fact we find that if α is large, but not infinite, then the shock structure is of the form

Downstream	Sonic Point	Sub - shock	Upstream	
As before	$ v > a_n$	$ v < a_n$	$ v > a_n$	As before

which satisfies Whitham criterion, but contains sub-shock in neutral fluid.

Figure 1 shows a shock structure for which α is such that there is a sub-shock in the neutral fluid, downstream of which there is a smooth transition from sub to supersonic with respect to the neutral sound speed. For a lower value of α there is no sub-shock (figure 2). These solutions were obtained by producing a shock with a constant speed piston and integrating the time dependent equations until the structure became steady using the second order Godunov scheme described in (Falle, 1991).

References

- Draine B T and McKee C F (1993). Theory of Interstellar Shocks. *Ann. Rev. Astron. Astrophys.* **31**, pp. 373-432..
- Falle S A E G (1991). Self-similar Jets. *Mon. Not. R. Ast. Soc.* **250**, pp 581-596.
- Liubarskii G Ia (1961). On the Structure of Shock Waves, *PPM* **25**, pp 1041-1049.
- Marshall W (1955) The Structure of Magnetohydrodynamic Shock Waves. *Proc. Roy. Soc. A.* **233**, pp 367-376.
- Whitham G B (1959). Some Comments on Wave Propagation and Shock-wave structure with Application to Magnetohydrodynamics, *Comm. Pure Appl. Math.* **12**, pp 113-158.
- Whitham G B (1974). Linear and Nonlinear Waves. Wiley.

THE GHOST FLUID METHOD FOR NUMERICAL TREATMENT OF DISCONTINUITIES AND INTERFACES

RONALD P. FEDKIW
UCLA Mathematics Department
e-mail: rfedkiw@math.ucla.edu

Abstract.

The Ghost Fluid Method (GFM) has recently been developed to handle interfaces in a robust and efficient fashion leading to a general class of “boundary condition capturing” techniques based on the identification of “continuous” and “discontinuous” variables and the subsequent treatment of these variables to allow seamless finite differencing across the interface.

1. Introduction

The most commonly used computational methods for interfaces are front tracking, volume of fluid, and level set methods. While each of these methods have well known strengths and weaknesses, only the level set method will be addressed in this paper, since the GFM has been developed using level set techniques for the interface. However, it should be noted that the GFM is not a level set specific method and could be extended to front tracking or volume of fluid formulations in a straightforward way. Level Set methods were first presented in (Osher and Sethian, 1988) where the time dependent level set equation

$$\phi_t + \vec{W} \cdot \nabla \phi = 0 \quad (1)$$

was used to keep track of the interface location as the set of points where $\phi = 0$. Although ϕ is initialized as a smooth signed distance function, it usually loses this desirable property as the interface deforms under the velocity \vec{W} making it necessary to apply reinitialization in order to keep $|\nabla \phi| \approx 1$ (Sussman et. al., 1994). Another useful equation

$$I_\tau + \vec{N} \cdot \nabla I = 0 \quad (2)$$

can be used to extrapolate the quantity I in the normal direction (Fedkiw et. al., 1999a).

2. Conservation

Traditionally, Eulerian based numerical methods for compressible flow are based on the Lax-Wendroff Theorem (Lax and Wendroff, 1960) which dictates that numerical methods should be fully conservative, and it is well known that nonconservative methods produce shocks with incorrect speeds and strengths. However, (Karni, 1996) advocates nonconservative form at contact discontinuities which are lower dimensional sets (e.g. one dimensional in a two dimensional calculation) that move with the local fluid velocity. In (Karni, 1996), full conservation was applied away from interfaces and a nonconservative method was applied near interfaces without adversely effecting the shock speeds or strengths. Since shocks *do not* move at the local interface velocity, any portion of a shock is only in contact with an interface, and thus the nonconservative method, on a set of measure zero in space and time minimizing the accumulation of errors.

While it is true that others have used nonconservative discretizations, there is no doubt that the work in (Karni, 1996) is responsible for markedly increasing their popularity in the shock capturing community where traditionally schemes were thought to require conservation at all cost. In part, this is because (Karni, 1996) identified and fixed large numerical oscillations introduced at interfaces by the fully conservative scheme presented in (Mulder et. al., 1992). It is interesting to note that many front tracking and volume of fluid schemes are actually nonconservative, i.e. they do not satisfy the strict flux differencing conservation form usually thought to be required by the Lax-Wendroff Theorem. In this sense, many of these schemes share similar properties with the scheme in (Karni, 1996). For example, consider the front tracking approach in (Pember et. al., 1995) where a high order Godonuv method is used to obtain a nonconservative update near the tracked interface and a fully conservative update away from the tracked interface. All flow features including shock speeds and strengths as well as the speed of the tracked front are correctly determined as is ensured by the solutions of the appropriate Riemann problems. Note that the authors go one step further and correct the lack of conservation at the interface using a redistribution procedure (Chern and Colella, 1987) which is presumably not necessary for obtaining a grid resolved solution, but is only used to maintain *exact* conservation. In fact, the nature of this redistribution procedure does not allow strict application of the Lax-Wendroff Theorem, and one has to believe that the correct solutions are obtained because the numerical method is fully conservative except at the lower di-

mensional tracked interface which is updated correctly based on solutions of the appropriate Riemann problems. Similar loss of exact conservation occurs in volume of fluid methods where nonphysical overshoots may occur in the volume fraction equation (Puckett et. al., 1997). These overshoots can be ignored violating conservation, or redistributed in a manner similar to (Chern and Colella, 1987) to preserve exact conservation.

3. Isobaric Fix

The well known “overheating effect” occurs when a shock reflects off of a solid wall boundary causing overshoots in temperature and density, while pressure and velocity remain constant. In one spatial dimension, a solid wall boundary condition can be applied with the aid of ghost cells by constructing a symmetric pressure and density reflection and an asymmetric normal velocity reflection about the solid wall. Then a shock wave impinging on the wall will collide with a shock in the ghost cells that has equal strength traveling in the opposite direction producing the desired shock reflection. In (Menikoff, 1994) and (Noh, 1978), the authors showed that overheating errors are a symptom of smeared out shock profiles and that sharper shocks usually produce less overheating. In addition, they showed that the pressure and velocity equilibrate quickly, while errors in the temperature and density persist. In order to dissipate these errors in temperature and density, (Noh, 1978) proposed adding artificial heat conduction to the numerical method in a form similar to artificial viscosity. Later, (Donat and Marquina, 1996) proposed a flux splitting method with a built in heat conduction mechanism that dissipates these errors throughout the fluid.

At this point, it is instructive to consider the one dimensional Euler equations and the associated Rankine-Hugoniot jump conditions for a discontinuity moving at speed D . Since a solid wall moves at the local flow velocity, $D = V_N$ and Rankine-Hugoniot jump conditions of $[V_N] = 0$, $[p] = 0$, and $0 = 0$ describe the relationship between the external flowfield and the internal one, i.e. both the normal velocity and the pressure must be continuous across the solid wall boundary extending into the ghost cells. Since these jump conditions are inherently part of the equations and thus part of any consistent numerical method, jumps in pressure and velocity are hard to maintain for any duration of time at a solid wall boundary, i.e. jumps between the fluid values and the ghost cell values are quickly dissipated. In this sense, one can think of pressure and velocity equilibration at a solid wall boundary as an intrinsic action of the boundary conditions. Note that there is no such condition for the temperature or the density. In the case of a complete equation of state (Davis, 1985), only one variable in the linearly degenerate field need be defined and all other variables can

be determined from the equation of state relations. In this sense, one can state that there is no boundary condition for the linearly degenerate field as is emphasized by the trivially satisfied jump condition $0 = 0$. Since a solid wall boundary is an initial boundary value problem, the value of the temperature at the wall must come from initial data as one can see from

$$S_t + \vec{V} \cdot \nabla S = 0 \quad (3)$$

which states that entropy is advected along streamlines of the fluid implying that the entropy near the wall stays near the wall since the wall moves with the local fluid velocity. We stress that this equation is only valid for smooth flow and is not true for streamlines that cross shock waves, i.e. entropy jumps across a shock wave. However, shock waves do not move at the same speed as solid wall boundaries so this equation is true near the wall most of time, i.e. except for a lower dimensional subset of space and time.

In (Fedkiw et. al., 1999c), equation 3 was used to develop the Isobaric Fix which is a boundary condition type of treatment for the linearly degenerate field at a solid wall boundary. The Isobaric Fix modifies the linearly degenerate field at a solid wall without changing the values of the pressure or the normal velocity. Noting that entropy is advected along streamlines and that streamlines are continuous, the entropy errors at the wall are repaired using new values of entropy extrapolated from the surrounding flow. For example, replacing the entropy at the wall with the entropy of the neighboring cell gives a first order accurate value of the entropy at the wall for smooth entropy profiles. Higher order accurate extrapolation can be used as well, but this has been found to be quite dangerous in practice due to the presence of discontinuous shock waves that cause large overshoots when extrapolating. In multiple spatial dimensions, the solid wall can be represented as the zero level of a level set function and moved rigidly using equation 1 where \vec{W} is the spatially constant wall velocity or deformed with equation 1 and a spatially varying \vec{W} . Then the Isobaric Fix can be applied using equation 2 with $I = S$.

4. Ghost Fluid Method

Similar to a solid wall boundary, a level set function can be used to track a contact discontinuity as the set of points where $\phi = 0$ separating two different fluids that each satisfy the Euler equations with different equations of state. Since the equation of state properties are discontinuous across the interface, the discretization techniques are employed in a similar fashion as for a solid wall boundary, except that they are applied twice, i.e. once for each fluid. Conceptually, each grid point corresponds to one fluid or the other and ghost cells can be defined at every point in the computational

domain so that each grid point contains the mass, momentum, and energy for the real fluid that exists at that point (according to the sign of the level set function) and a ghost mass, momentum, and energy for the other fluid that does not really exist at that grid point (the fluid from the other side of the interface). Once the ghost cells are defined, standard one phase numerical methods can be used on the entire domain for each fluid, i.e. we now have two separate single fluid problems. After each fluid is advanced in time, the level set function is updated using equation 1 with $\vec{W} = \vec{V}$ (the local fluid velocity), and the sign of the level set function is used to determine the appropriate real fluid values at each grid point. Note that ghost cells are defined everywhere for exposition, but only a band of 3 to 5 ghost cells is actually used in practice.

Since contact discontinuities move at the local fluid velocity, the Rankine-Hugoniot jump conditions for a contact discontinuity are the same as those for a solid wall boundary, i.e. $[V_N] = 0$, $[p] = 0$, and $0 = 0$. In multiple spatial dimensions, the $0 = 0$ jump condition is repeated since the tangential velocities are also governed by the linearly degenerate field, e.g. in three spatial dimensions one can only determine the pressure and normal velocities from the boundary conditions, while the entropy and both tangential velocities remain undetermined. Note that in the case of the full viscous Navier-Stokes equations, the physical viscosity imposes continuity of the tangential velocities, and thermal conductivity imposes continuity of the temperature. Since certain properties are discontinuous across the interface, one should be careful when applying finite difference methods *across* the interface, since differencing discontinuous quantities leads erroneously to terms of the form $\frac{1}{\Delta x}$ that increase without bound as the grid is refined. Therefore, the layer of ghost cells should be introduced so that there is continuity with the neighboring fluid that needs to be discretized. For variables that are already continuous across the interface, e.g. pressure and normal velocity, the ghost fluid values can be set equal to the real fluid values at each grid point implicitly capturing the correct interface values of these variables. This is the key mechanism in *coupling* the two distinct sets of Euler equations. On the other hand, the discontinuous variables move with the speed of the interface (see equation 3), and information in these variables does not cross the interface and is not coupled to the corresponding information on the other side of the interface. Moreover, in order to avoid numerical smearing or spurious oscillations these discontinuous variables should not be nonphysically coupled together or forced to be continuous across the interface. The most obvious way of defining the discontinuous variables in the ghost cells is by extrapolating that information from the neighboring real fluid nodes, e.g. the entropy can be extrapolated into the ghost cells using equation 2 in exactly the same way as it was when apply-

ing the Isobaric Fix producing a continuous entropy profile. Since entropy is characteristic of the equation of state information and the fluid itself, we denoted this method the Ghost Fluid Method (Fedkiw et. al., 1999a), i.e. ghost cells that are physically located in one fluid are filled with entropy from the neighboring fluid changing the *kind* of fluid in these cells without changing the way these cells behave, i.e. without changing the pressure and normal velocity. Note that, similar to the Isobaric Fix, one does not have to deal directly with the entropy but can choose any variable in the linearly degenerate field, e.g. density or temperature. Since the tangential velocities are discontinuous as well, a similar extrapolation procedure is used to treat these variables making use of a basis free projection method (Fedkiw et. al., 1999a).

5. Other Discontinuities

For a simple contact discontinuity, the variables were separated into two sets based on their continuity across the interface. The continuous variables were copied into the ghost fluid in a node by node fashion capturing the correct interface values, while the discontinuous variables were extrapolated in a one-sided fashion to avoid errors due to numerical dissipation. In order to apply this idea to a general interface moving at speed D in the normal direction, one needs to *correctly* determine the continuous and discontinuous variables for a general interface problem. For example, consider a shock wave where all variables are discontinuous, and extrapolation of all variables for both the pre-shock and post-shock fluids obviously gives the wrong answer since the physical coupling is ignored. We state, “*For each degree of freedom that is coupled across a discontinuity, one can define a variable which is continuous across the discontinuity, and all remaining degrees of freedom can be expressed as discontinuous variables which can be extrapolated across the interface in a one-sided fashion.*” as the key to extending the GFM. In the case of the Euler equations, conservation of mass, momentum, and energy can be applied to any discontinuity in order to abstract continuous variables, i.e. the Rankine-Hugoniot jump conditions always dictate the coupling between the pre-discontinuity and post-discontinuity fluids. In (Fedkiw et. al., 1999b), the Rankine-Hugoniot relations were used in three spatial dimensions to define F_ρ , $F_{\rho V_N}$, $\vec{F}_{\rho V_T}$, and F_E as continuous variables across a discontinuity which has speed $D \neq \vec{V}_N$, i.e. when the discontinuity is not a contact discontinuity. This allowed us to develop a GFM that *implicitly* captures the interface values of these continuous quantities at shocks, detonations, and deflagrations, i.e. the method *implicitly* captures the Rankine-Hugoniot jump conditions without numerical smearing.

When the GFM is used for general discontinuities, one needs to accurately find the interface speed D for equation 1 with $\vec{W} = D\vec{N}$. For shock waves and detonation waves, D can be found by solving an appropriate Riemann problem in a node by node fashion (Fedkiw et. al., 1999b). In fact, there is no reason one cannot solve a Riemann problem in the case of a contact discontinuity as well using $\vec{W} = D\vec{N}$ in equation 1 as opposed to $\vec{W} = \vec{V}$ (the less accurate local fluid velocity) (Fedkiw et. al., 1999b). Note that a combination of ghost cells and Riemann problems is commonly used in front tracking algorithms, see e.g. (Glimm et. al., 1980) and (Glimm et. al., 1999) where a Riemann problem is solved at the interface and the results are extrapolated into the ghost cells. The difference between the GFM and typical front tracking is in the order of operations, i.e. front tracking algorithms first solve a Riemann problem and then extrapolate while the GFM extrapolates first and then solves the Riemann problem in a node by node fashion removing some of the complications due to geometry. For a deflagration wave, the Riemann problem is not well posed unless the speed of the deflagration, i.e. D , is already given. Luckily, the G-equation for flame discontinuities (first proposed in (Markstein, 1964)) represents a flame front as a discontinuity in the same fashion as the level set method so that one can easily consult the abundant literature on the G-equation to obtain deflagration speeds.

6. Incompressible Flow

In multiphase incompressible flow calculations, a variable coefficient Poisson equation needs to be solved since the density is usually different (although constant) in each phase. This equation is not straightforward to solve especially when $[p] \neq 0$ which is typical for any multiphase flow problem where the viscosity jumps across the interface or surface tension is present. The most notable method for solving the Poisson equation is probably the “immersed boundary” method (Peskin, 1977) which uses a δ -function formulation to smear out the solution on a thin finite band about the interface. However, this numerical smearing has an adverse effect on the solution forcing continuity at the interface regardless of the appropriate interface boundary conditions, i.e. the nonzero jump in the pressure is not accurately represented. This failing has been overcome by a number of authors who solve the Poisson equation with $[p] = 0$ and then add new source terms to the momentum equations, see e.g. (Brackbill et. al., 1992), (Unverdi and Tryggvason, 1992) and (Sussman et. al., 1994). In the interest of solving the Navier-Stokes equations directly, i.e. without the addition of source terms, a new GFM was designed for the variable coefficient Poisson equation (Liu et. al., 2000) allowing one to solve this equation with both $[p]$ and $[\frac{p_n}{\rho}]$ as given

nonzero jumps, and ρ discontinuous. This new method used the given jump conditions to define continuous variables for the finite differencing similar to the way that the Rankine-Hugoniot jump conditions were used for multiphase compressible flow. It is notable that the resulting system of linear equations is completely symmetric allowing for straightforward application of standard linear system solvers. This new numerical method does not suffer from the numerical smearing prevalent in the “immersed boundary” method and was used to solve the multiphase Navier-Stokes equations in (Kang et. al., 2000) without the need for additional source terms.

Acknowledgements

Research supported in part by ONR N00014-97-1-0027.

References

- Brackbill J, Kothe D and Zemach C (1992). A continuum method for modeling surface tension. *J.C.P.* **100**, p 335.
- Chern I-L and Colella P (1987). A conservative front tracking method for hyperbolic conservation laws. *UCRL-97200 LLNL*.
- Davis W (1985). Equation of state for detonation products. *Proceedings Eighth Symposium on Detonation*, p 785.
- Donat R and Marquina A (1996). Capturing shock reflections: an improved flux formula. *J.C.P.* **125**, p 42.
- Fedkiw R, Aslam T, Merriman B and Osher S (1999). A non-oscillatory Eulerian approach to interfaces in multimaterial flows (the ghost fluid method). *J.C.P.* **152**, p 457.
- Fedkiw R, Aslam T and Xu S (1999). The ghost fluid method for deflagration and detonation discontinuities. *J.C.P.* **154**, p 393.
- Fedkiw R, Marquina A and Merriman B (1999). An isobaric fix for the overheating problem in multimaterial compressible flows. *J.C.P.* **148**, p 545.
- Glimm J, Grove J, Li X and Zhao N (1999). Simple front tracking. *J. Nonlinear Partial Differential Equations, Contemporary Mathematics* **238**.
- Glimm J, Marchesin D and McBryan O (1980). Subgrid resolution of fluid discontinuities, II. *J.C.P.* **37**, p 336.
- Kang M, Fedkiw R and Liu X-D (in review). A boundary condition capturing method for multiphase incompressible flow. *J.C.P.*
- Karni S (1996). Hybrid multifluid algorithms. *SIAM J. Sci. Comput.* **17**, p 1019.
- Lax PD and Wendroff B (1960). Systems of conservation laws. *Comm. Pure Appl. Math* **13**, p 217.
- Liu X-D, Fedkiw R and Kang M (to appear). A boundary condition capturing method for Poisson's equation on irregular domains. *J.C.P.*
- Markstein G (1964). Nonsteady flame propagation. *Pergamon Press, Oxford*.
- Menikoff R (1994). Errors when shock waves interact due to numerical shock width. *SIAM J. Sci. Comput.* **15**, p 1227.
- Mulder W, Osher S and Sethian J (1992). Computing interface motion in compressible gas dynamics. *J.C.P.* **100**, p 209.
- Noh W (1978). Errors for calculations of strong shocks using an artificial viscosity and an artificial heat flux. *J.C.P.* **72**, p 78.
- Osher S and Sethian J (1988). Fronts propagating with curvature dependent speed: algorithms based on Hamilton-Jacobi formulations. *J.C.P.* **79**, p 12.

- Pember R, Bell J, Colella P, Crutchfield W and Welcome M (1995). An adaptive cartesian grid method for unsteady compressible flow in irregular regions. *J.C.P.* **120**, p 278.
- Peskin C (1977). Numerical analysis of blood flow in the heart. *J.C.P.* **25**, p 220.
- Puckett EG, Almgren A, Bell J, Marcus D, Rider W (1997). A high order projection method for tracking fluid interfaces in variable density incompressible flows. *J.C.P.* **130**, p 269.
- Sussman M, Smereka P and Osher S (1994). A level set approach for computing solutions to incompressible two-phase flow. *J.C.P.* **114**, p 146.
- Unverdi S and Tryggvason G (1992). A front-tracking method for viscous, incompressible, multi-fluid flows. *J.C.P.* **100**, p 25.

A HYBRID PRIMITIVE-CONSERVATIVE UPWIND SCHEME FOR THE DRIFT FLUX MODEL

KJELL KÅRE FJELDE

RF-Rogaland Research,

Thormølensgt. 55,

N-5008 Bergen.

e-mail: Kjell-Kaare.Fjelde@rf.no

AND

KENNETH HVISTENDAHL KARLSEN

Department of Mathematics,

University of Bergen,

Johs. Brunsgt. 12,

N-5008 Bergen, Norway.

e-mail: kennethk@mi.uib.no

Abstract

We consider a drift flux model describing one-dimensional isothermal gas-liquid flow in a long pipeline. The model consists of equations for the conservation of mass for each of the phases and the conservation of momentum of the mixture. In addition, an equation — the so-called gas slip relation — is supplied which relates the velocities of the two phases at any point. We present and demonstrate a high-resolution hybrid primitive-conservative upwind scheme, which utilises a simple non-conservative MUSCL scheme in shock free flow regions and a fully numerical, conservative Roe scheme in shocked flow regions. The overall conclusion of the numerical tests is that the hybrid upwind scheme is as accurate as the conservative Roe scheme. The hybrid scheme is, however, more efficient.

1. Introduction

This work is concerned with *shock-capturing* schemes of *upwind* type for solving a system of conservation laws modelling one-dimensional two-phase flow in a pipeline. The so-called *drift flux* model considered in this paper is a two-phase flow model consisting of mass conservation equations for the two phases and one equation for the conservation of the momentum of the mixture. Mass transfer between the phases is not considered. More precisely, the one-dimensional drift flux model takes the following form:

The mass conservation equation for the liquid phase

$$(\alpha_l \rho_l)_t + (\alpha_l \rho_l v_l)_x = 0. \quad (1)$$

The mass conservation equation for the gas phase

$$(\alpha_g \rho_g)_t + (\alpha_g \rho_g v_g)_x = 0. \quad (2)$$

The mixture momentum equation

$$(\alpha_l \rho_l v_l + \alpha_g \rho_g v_g)_t + (\alpha_l \rho_l v_l^2 + \alpha_g \rho_g v_g^2 + p)_x = -q. \quad (3)$$

In (1)-(3), t is time; x is the flow direction (space coordinate); p is the pressure; α_ℓ is the volume fraction of phase $\ell = g, l$; ρ_ℓ is the density of phase $\ell = g, l$; v_ℓ is the velocity of phase $\ell = g, l$; and q represents the wall friction. The volume fractions α_l and α_g are related by $\alpha_l + \alpha_g = 1$. The conserved quantities in (1)-(3) represent the liquid mass, gas mass and mixture momentum. In addition, several closure laws must be given which involve density models for each phase, a gas slip relation and a model for wall friction. These models are usually quite complex and often given in tabular form based on experimental data. For the purpose of numerical testing, however, we have used simplified models given in purely algebraic form. Due to space limitation we do not state these models here, but refer instead to (Fjelde and Karlsen, 1999) for details. Some of the mathematical properties of the system (1)-(3) are studied in, e.g., (Benzoni-Gavage, 1991). The model was shown to be hyperbolic with two nonlinear sonic waves and a contact discontinuity corresponding to transport of the gas volume fractions.

Note that (1)-(3) can be written as a system of conservation laws $U_t + F_x = Q$, where U, F, Q should have obvious meaning. In general, it is however difficult to express the flux vector F in terms of the conservative state vector U . The closure laws are formulated in terms of physical variables and often given in tabular form. Consequently, there is often no algebraic expression of the flux F as a function of U . This fact implies that upwind schemes based on an algebraically given Riemann solver or a flux vector

splitting are not easy to apply to the drift flux model. Fortunately, it is possible to construct fully numerical upwind schemes which avoid analytical computations. Although these schemes are accurate, they tend to be very time consuming. This is mainly due to the need for repeatedly computing (numerically) the Jacobian of the system with respect to the conservative variables. Moreover, a conservative scheme will produce conservative variables at each new time level which must be transformed into primitive variables, since the submodels involved are formulated in terms of primitive variables.

As is discussed in (Fjelde and Karlsen, 1999), primitive formulations of the drift flux model have much simpler form than the conservative formulation. Therefore one expects that numerical schemes based on a primitive formulation is more efficient. However, as is well known, primitive schemes fail to produce correct solutions in the presence of shock waves. To deal with this latter problem, we are lead to designing hybrid upwind schemes which use a cheap upwind scheme based on a primitive formulation in smooth regions and switch to a fully numerical and conservative upwind scheme near shocks. The purpose of this short communication is to present (very briefly) and demonstrate one such hybrid scheme for the drift flux model. For further details and other hybrid schemes, we refer to our paper (Fjelde and Karlsen, 1999). This communication as well as (Fjelde and Karlsen, 1999) rely on ideas from (Toro, 1998).

2. Numerical algorithm

To approximate the solution of (1)-(3), we introduce a mesh in the (x, t) plane where the spatial grid points are denoted by x_j and the time levels by t_n . We denote the spacing in the x and t variables by Δx and Δt , respectively. Approximate solutions in primitive and conservative variables are denoted respectively by W_j^n and U_j^n . We next briefly describe our hybrid scheme, which is based on the use of a non-conservative MUSCL scheme in shock free flow regions and a conservative Roe scheme in shocked flow regions. We first describe the non-conservative (primitive) scheme and the conservative scheme. Then we describe a simple strategy for switching from the primitive scheme to a conservative scheme in the vicinity of shock waves. We refer to (Fjelde and Karlsen, 1999) for further details on the numerical algorithm.

2.1. NON-CONSERVATIVE UPDATING

We shall use a primitive formulation of (1)-(3) of the type

$$W_t + A(W)W_x = \tilde{Q},$$

where the primitive state vector W is taken as $W = (\alpha_g, v_l, p)^\top$. Although it is easy to determine the corresponding coefficient matrix $A(W)$ explicitly, we choose not to display $A(W)$ here due to space limitation, see instead (Fjelde and Karlsen, 1999).

We now consider a non-conservative MUSCL scheme of the form (Toro, 1998; Fjelde and Karlsen, 1999)

$$W_j^{n+1} = W_j^n - \frac{\Delta t}{\Delta x} \bar{A}_j [W_{j+1/2}^{n+1/2} - W_{j-1/2}^{n+1/2}] + \tilde{Q}_j^n, \quad (4)$$

where the intermediate states $W_{j\pm 1/2}^{n+1/2}$ and the coefficient matrix \bar{A}_j will be determined next. The basis of any MUSCL scheme is the reconstruction of piecewise constant data $\{W_j^n\}$ into a piecewise linear distribution of the data and then to extrapolate the data to the edges of each cell, yielding the extrapolated values W_j^l, W_j^r . The extrapolated values W_j^l and W_j^r are then evolved half a time step according to the non-conservative formulas $\hat{W}_j^{l,r} = W_j^{l,r} - \frac{1}{2} \frac{\Delta t}{\Delta x} A_j^n [W_j^r - W_j^l]$, where $A_j^n = A(W_j^n)$. For later use, we denote by $\hat{U}_j^{l,r}$ the conservative values corresponding to $\hat{W}_j^{l,r}$. Finally, we need to find the self-similar solution $W_{j+1/2}(\xi)$, $\xi = (x - x_{j+1/2})/t$, of the linear Riemann problem

$$W_t + \hat{A} W_x = 0, \quad \hat{A} = A\left(\frac{\hat{W}_j^r + \hat{W}_{j+1}^l}{2}\right), \quad W(x, 0) = \begin{cases} \hat{W}_j^r, & x \leq x_{j+1/2}, \\ \hat{W}_{j+1}^l, & x > x_{j+1/2}. \end{cases}$$

We then set $W_{j+1/2}^{n+1/2} = W_{j+1/2}(0)$ and $\bar{A}_j = A\left(\frac{W_{j-1/2}^{n+1/2} + W_{j+1/2}^{n+1/2}}{2}\right)$.

2.2. CONSERVATIVE UPDATING

Based on the conservative formulation $U_t + F_x = Q$, we consider the conservative-form scheme

$$U_j^{n+1} = U_j^n - \frac{\Delta t}{\Delta x} [\mathcal{F}_{j+1/2} - \mathcal{F}_{j-1/2}] + Q_j^n, \quad (5)$$

where U_j^n is an average of the conserved variables on $(x_{j-1/2}, x_{j+1/2})$ and $\mathcal{F}_{j+1/2}$ is the numerical flux of Roe

$$\mathcal{F}_{j+1/2}^{\text{Roe}} = \frac{1}{2} (F(\hat{U}_j^r) + F(\hat{U}_{j+1}^l)) - \frac{1}{2} |\hat{J}(\hat{U}_j^r, \hat{U}_{j+1}^l)| (\hat{U}_{j+1}^l - \hat{U}_j^r).$$

Note that we use the values $\hat{U}_j^{l,r}$ (see §2.1) to obtain high-resolution in space and time. The matrix $\hat{J} = \hat{J}(\hat{U}_j^r, \hat{U}_{j+1}^l)$ is the so-called Roe matrix. This matrix must satisfy the three usual conditions implying (i) hyperbolicity,

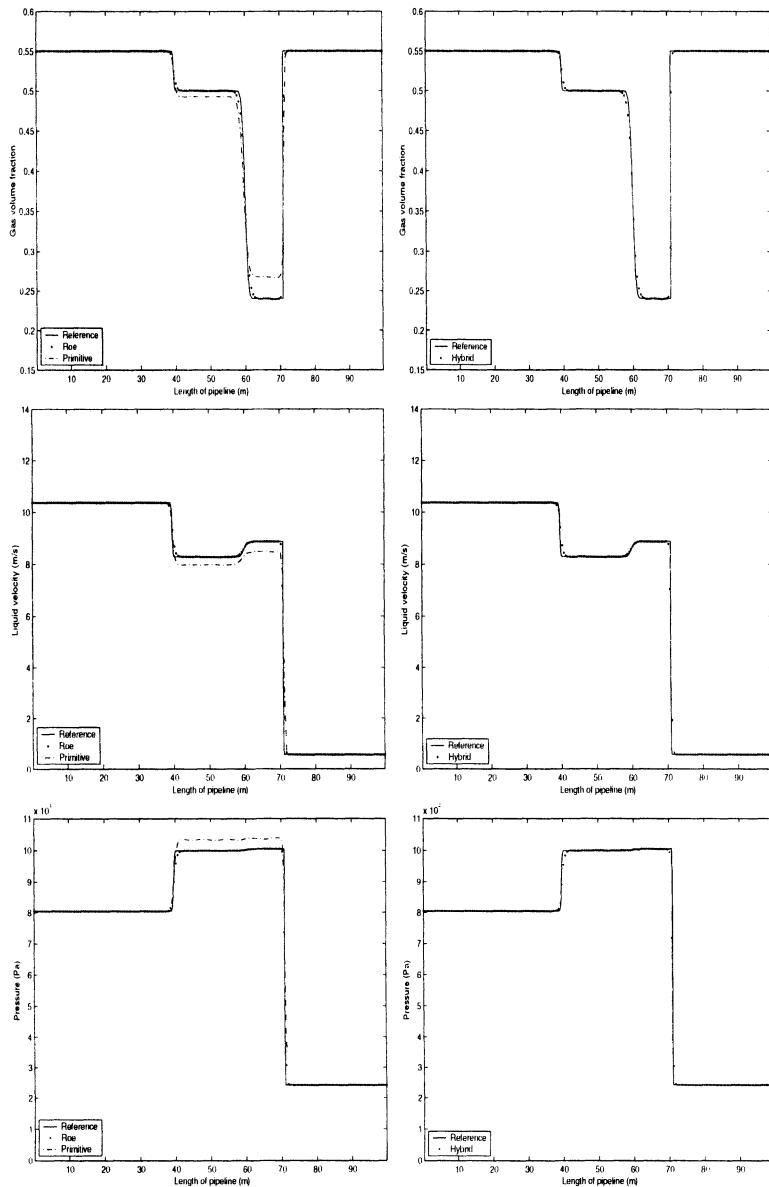


Figure 1. Left: Primitive scheme VS Roe scheme. Right: Hybrid scheme.

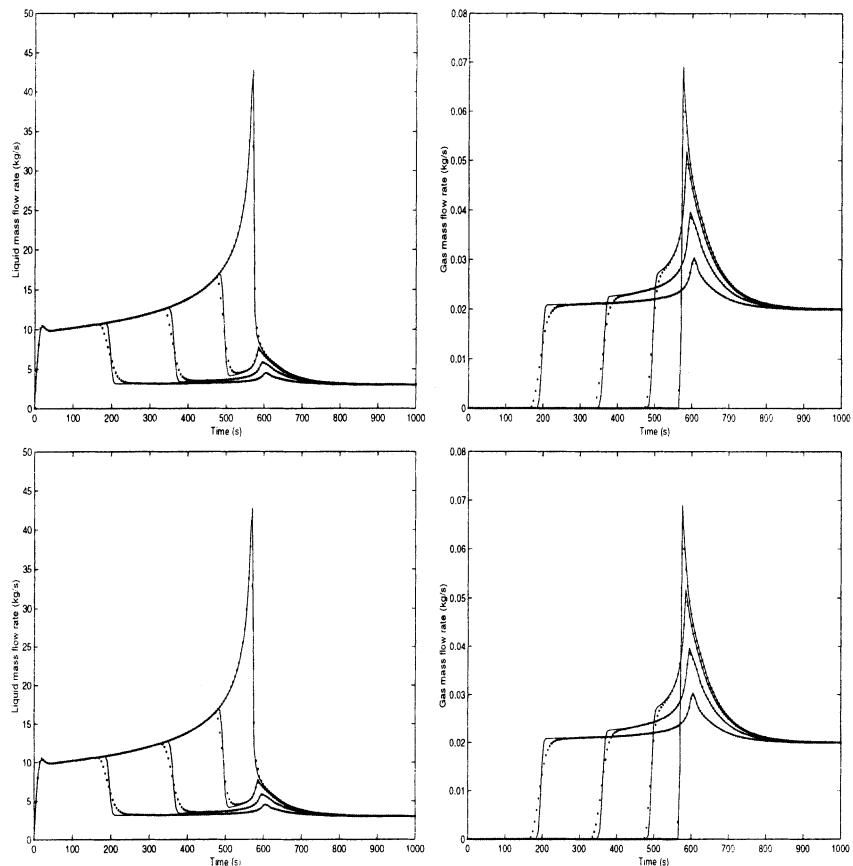


Figure 2. Mass flow rates for positions $x = 250$ [m], 500 [m], 750 [m], 1000 [m] (from left to right in the plots). Upper row: Roe scheme. Lower row: Hybrid scheme. Reference solution and approximate solutions are shown respectively as solid line and thick dots.

(ii) consistency with the conservation laws, and finally (iii) conservation and correct recognition of isolated discontinuities. Thanks to condition (iii), the Roe scheme can be written in conservative form (5). To apply Roe's scheme, one must analytically construct a Roe matrix. In particular, it is difficult to find a matrix \hat{J} satisfying condition (iii). Following (Ramate, 1998), we shall here employ a fully numerical Roe solver that avoids the analytical part of the original Roe scheme. Instead of trying to find an average state $\tilde{U}_{j+1/2}$ so that the Jacobian $\tilde{J} = J(\tilde{U}_{j+1/2})$ satisfies the Roe condition (iii), one can first simply construct the matrix $\bar{J} = J\left(\frac{\hat{W}_j^r + \hat{W}_{j+1}^l}{2}\right)$. The matrix \bar{J} automatically satisfies (i) and (ii) but not (iii). To deal with the latter, we first decompose the Jacobian \bar{J} as $\bar{J} = \bar{R}\bar{\Lambda}\bar{R}^{-1}$, where \bar{R} is the matrix of which the columns are the right eigenvectors of \bar{J} , and $\bar{\Lambda}$ is the diagonal matrix having the eigenvalues of \bar{J} as diagonal entries. Next, we determine a diagonal matrix $\hat{\Lambda}$ such that $\bar{R}\hat{\Lambda}\bar{R}^{-1}\Delta U = \Delta F$, where $\Delta U = \hat{U}_{j+1}^l - \hat{U}_j^r$ and $\Delta F = F(\hat{U}_{j+1}^l) - F(\hat{U}_j^r)$. Then we define $\hat{J} = \bar{R}\hat{\Lambda}\bar{R}^{-1}$. It is easily checked that the matrix \hat{J} now satisfies Roe's condition (iii). Furthermore, conditions (i) and (ii) are also satisfied. Hence, \hat{J} is a true Roe matrix satisfying all three conditions (i) - (iii).

2.3. SWITCHING STRATEGY

Consider a grid point $x = x_j$ and two neighboring Riemann problems $W_{j-1/2}(\xi)$ and $W_{j+1/2}(\xi)$. Denote by $p_{j-1/2}$ and $p_{j+1/2}$ the solutions for the pressure along the interfaces $x = x_{j-1/2}$ and $x = x_{j+1/2}$, respectively. Let $\varepsilon > 0$ be a small positive number. Then the updating of the solution at $x = x_j$ is done with the conservative scheme (5) if $\left|\frac{p_{j\pm 1/2}}{p_j^n} - 1\right| > \varepsilon$, otherwise the updating is done with the primitive scheme (4). The choice of ε is not too important. Experience indicates that any value in the range $(0, 0.1)$ gives good results. This is also consistent with the experience in (Toro, 1998) with the Euler equations (where this switching strategy was first used). We refer to (Fjelde and Karlsen, 1999) for further details.

3. Numerical results

3.1. SHOCK TUBE PROBLEM

We consider a two-phase shock tube problem (similar to the one in (Benzoni-Gavage, 1991)). For a precise description of the parameters involved, we refer to (Fjelde and Karlsen, 1999). Let us only say that the exact solution is composed of a 1-shock, a 2-contact discontinuity and a 3-shock. The results given by the schemes at time $t = 1.0$ [s] are shown in Fig. 1. We have chosen to present plots of the gas volume fraction, liquid velocity and

pressure. As anticipated, the primitive scheme produces shock waves with wrong strength. The shock positions are incorrect, but not very different from those predicted by the conservative scheme. The hybrid scheme used conservative updating both for the left- and right-going shocks and the solution was corrected.

The schemes used a CFL number of 0.8 and space discretization $\Delta x = 0.5$ [m]. The MINMOD limiter was used to obtain second order accuracy in space.

3.2. TRANSPORT OF GAS AND LIQUID

We consider a 1000 [m] long horizontal pipeline with diameter 10 [cm]. Initially, the pipe is filled with stagnant liquid and we are interested in modelling the transient behaviour induced by injecting gas and liquid at the inlet. The gas and liquid mass flow rates are increased to respectively 0.02 [kg/s] and 3.0 [kg/s] in 10 [s]. At the outlet boundary, the pressure is kept constant equal to 1 [bar]. For more details about the physical setup, we refer again to (Fjelde and Karlsen, 1999). The schemes applied the compressive SUPERBEE limiter, a CFL number of 0.5, and space discretization, $\Delta x = 20$ [m]. The Jacobian in the Roe scheme was calculated numerically.

The mass flow rates as function of time at different locations given by the Roe scheme and the hybrid scheme are shown in Fig. 2. The pressure in the pipeline decreases from the inlet towards the outlet and is mainly determined by the frictional source term. Hence, the gas will expand and quite large maximum flow rates are obtained especially at the outlet.

The schemes produced similar results. However, the CPU-time for the conservative Roe scheme was 1088.5 [s] compared to 670.6 [s] for the hybrid Roe scheme. For more details, we refer again to (Fjelde and Karlsen, 1999).

References

- Benzoni-Gavage S (1991). *Analyse numérique des modèles hydrodynamiques d'écoulements diphasiques instationnaires dans les réseaux de production pétrolière*. Thèse, ENS Lyon France.
- Fjelde K K and Karlsen K H (1999). High-resolution hybrid primitive-conservative upwind schemes for the drift flux model. Preprint. Submitted to *Comput. & Fluids*.
- Romate J E (1998). An approximate Riemann solver for a two-phase flow model with numerically given slip relation. *Comput. & Fluids*, **27**(4):455–477.
- Toro E F (1998). Primitive, conservative and adaptive schemes for hyperbolic conservation laws. In *Numerical methods for wave propagation (Manchester, 1995)*, pages 323–385. Kluwer Acad. Publ., Dordrecht.

NUMERICAL SIMULATIONS OF RELATIVISTIC WIND ACCRETION ONTO BLACK HOLES USING GODUNOV-TYPE METHODS

JOSÉ A. FONT

*Max-Planck-Institut für Astrophysik
Karl-Schwarzschild-Str. 1, D-85740 Garching, Germany
Email: font@mpa-garching.mpg.de*

JOSÉ M^A. IBÁÑEZ

*Departamento de Astronomía y Astrofísica
Universidad de Valencia, 46100 Burjassot (Valencia), Spain
Email: ibanez@godunov.daa.uv.es*

AND

PHILIPPOS PAPADOPOULOS

*School of Computer Science and Mathematics
University of Portsmouth, PO1 2EG, Portsmouth, U.K.
Email: Philippos.Papadopoulos@port.ac.uk*

Abstract. We have studied numerically the so-called Bondi-Hoyle (wind) accretion onto a rotating black hole in general relativity. We have used the Kerr-Schild form of the Kerr metric, free of coordinate singularities at the black hole horizon. The ‘test-fluid’ approximation has been adopted, assuming no dynamical evolution of the gravitational field. We have used a formulation of the relativistic hydrodynamic equations which casts them into a first-order hyperbolic system of conservation laws. Our studies were performed using a Godunov-type scheme based on Marquina’s flux-formula.

We find that regardless of the value of the black hole spin the final accretion pattern is always stable, leading to constant accretion rates of mass and momentum. The flow is characterized by a strong tail shock which is increasingly wrapped around the central black hole as the hole angular momentum increases. The rotation induced asymmetry in the pressure field implies that besides the well known drag, the black hole will experience also a *lift* normal to the flow direction.

1. Introduction

The term “wind” or hydrodynamic accretion refers to the capture of matter by a moving object under the effect of the underlying gravitational field. The canonical astrophysical scenario in which matter is accreted in such a non-spherical way was suggested originally by Bondi and Hoyle (Bondi and Hoyle, 1944), who studied, using Newtonian gravity, the accretion on to a gravitating point mass moving with constant velocity through a non-relativistic gas of uniform density. Such process applies to describe mass transfer and accretion in compact X-ray binaries, in particular in the case in which the donor (giant) star lies inside its Roche lobe and loses mass via a stellar wind. This wind impacts on the orbiting compact star forming a bow-shaped shock front around it.

The problem was first numerically investigated in the early 70’s. Since then, contributions of a large number of authors using highly developed Godunov-type methods extended the simplified analytic models. See (Ruffert and Arnett, 1994; Benensohn et al., 1997) and references therein. These Newtonian investigations helped to develop a thorough understanding of the hydrodynamic accretion scenario, in its fully three-dimensional character, revealing the formation of accretion disks and the appearance of non-trivial phenomena such as shock waves or *flip-flop* instabilities.

We have recently considered hydrodynamic accretion on to a moving black hole using relativistic gravity and the “test fluid” approximation (Font and Ibáñez, 1998a; Font and Ibáñez, 1998b; Font et al., 1998; Font et al., 1999). We present here a brief summary of the methodology and results of such simulations. We integrate the general relativistic hydrodynamic equations in the fixed background of the Kerr spacetime (including its non-rotating Schwarzschild limit) and neglect the self-gravity of the fluid as well as non-adiabatic processes such as viscosity or radiative transfer. In the black hole case the matter flows ultimately across the event horizon and becomes causally disconnected from distant observers. Near that region the problem is intrinsically relativistic and the gravitational accelerations significantly deviate from the Newtonian values.

2. Equations

The general relativistic hydrodynamic equations can be cast as a first-order flux-conservative system describing the conservation of mass, momentum and energy. Formulations of this sort are given, e.g. in (Banyuls et al., 1997; Papadopoulos and Font, 2000). In this work we follow the approach laid out in (Banyuls et al., 1997) for a perfect fluid stress-energy tensor $T^{\mu\nu}$.

The system of equation then reads:

$$\frac{1}{\sqrt{-g}} \left(\frac{\partial \sqrt{-g} \mathbf{u}}{\partial x^0} + \frac{\partial \sqrt{-g} \mathbf{f}^i}{\partial x^i} \right) = \mathbf{s} \quad (1)$$

($x^0 = t$; x^i spatial coordinates, $i = 1, 2, 3$) where $\mathbf{u} \equiv \mathbf{u}(\mathbf{w})$ are the evolved quantities, $\mathbf{u} = (D, S_j, \tau)$ and \mathbf{f}^i are the fluxes

$$\mathbf{f}^i = \left(D \left(v^i - \frac{\beta^i}{\alpha} \right), S_j \left(v^i - \frac{\beta^i}{\alpha} \right) + p \delta_j^i, \tau \left(v^i - \frac{\beta^i}{\alpha} \right) + p v^i \right), \quad (2)$$

v^i being the velocity and p the pressure. The corresponding sources \mathbf{s} are given by

$$\mathbf{s} = \left(0, T^{\mu\nu} \left(\frac{\partial g_{\nu j}}{\partial x^\mu} - \Gamma_{\nu\mu}^\delta g_{\delta j} \right), \alpha \left(T^{\mu 0} \frac{\partial \ln \alpha}{\partial x^\mu} - T^{\mu\nu} \Gamma_{\nu\mu}^0 \right) \right). \quad (3)$$

We note the presence of geometric terms in the fluxes and sources which appear as the local conservation laws of the density current and stress-energy are expressed in terms of partial derivatives. These terms are the lapse function α , the shift vector β^i and the connection coefficients $\Gamma_{\nu\mu}^\delta$ of the 3+1 spacetime metric

$$ds^2 \equiv g_{\mu\nu} dx^\mu dx^\nu = -(\alpha^2 - \beta_i \beta^i) dt^2 + 2\beta_i dx^i dt + \gamma_{ij} dx^i dx^j \quad (4)$$

Additionally $g \equiv \det(g_{\mu\nu})$ is such that $\sqrt{-g} = \alpha \sqrt{\gamma}$ and $\gamma \equiv \det(\gamma_{ij})$.

The vector \mathbf{w} , representing the primitive variables, is given by $\mathbf{w} = (\rho, v_i, \epsilon)$ where ρ is the density and ϵ the specific internal energy. The evolved quantities are defined in terms of the primitive variables as $D = \rho W$, $S_j = \rho h W^2 v_j$ and $\tau = \rho h W^2 - p - D$, W being the Lorentz factor $W = (1 - v^2)^{-1/2}$, with $v^2 = \gamma_{ij} v^i v^j$, and h the specific enthalpy, $h = 1 + \epsilon + p/\rho$. A perfect fluid equation of state $p = (\Gamma - 1)\rho\epsilon$, Γ being the constant adiabatic index, closes the system.

In our computations we specialize the above expressions to the Kerr line element which describes the exterior geometry of a rotating black hole. We use the Kerr-Schild form of the Kerr metric, which is free of coordinate singularities at the black hole horizon. Computations using the more standard Boyer-Lindquist (singular) form of the metric are presented in (Font et al., 1999). Pertinent technical details concerning the specific form of these metrics are given in (Papadopoulos and Font, 1998).

3. Numerical scheme

Our hydrodynamical code performs the numerical integration of system (1) using a Godunov-type method. The time update from t^n to t^{n+1} proceeds

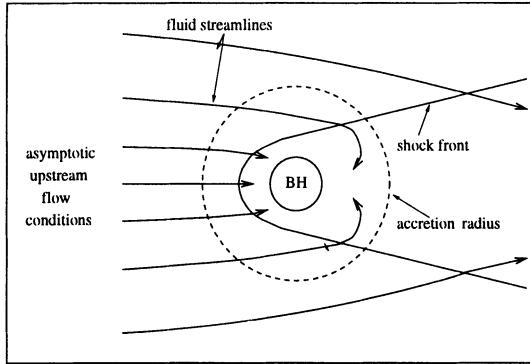


Figure 1. Schematic representation of stationary supersonic wind accretion. The shock may be detached as in the figure (bow shock) or attached to the rear part (tail shock), depending on the flow asymptotic conditions (v_∞ and $c_{s\infty}$) and thermodynamics of the gas (ρ_∞ and Γ).

according to the following algorithm in conservation form:

$$\mathbf{u}_{i,j}^{n+1} = \mathbf{u}_{i,j}^n - \frac{\Delta t}{\Delta x^k} (\hat{\mathbf{f}}_{i+1/2,j} - \hat{\mathbf{f}}_{i-1/2,j}) + \Delta t \mathbf{s}_{i,j}, \quad (5)$$

improved with the use of (second-order) conservative Runge-Kutta sub-steps to gain accuracy in time (Shu and Osher, 1988). The numerical fluxes are computed by means of Marquina's flux-formula (Donat and Marquina, 1996). After the update of the conserved quantities the primitive variables are computed via a root-finding procedure.

The flux-formula makes use of the complete characteristic information of system (1), eigenvalues (characteristic speeds) and right and left eigenvectors. Generic expressions are collected in (Ibáñez et al., 1999).

The state variables, \mathbf{u} , must be computed (reconstructed) at the left and right sides of a given interface, out of the cell-centered quantities, prior to computing the numerical fluxes. In relativistic hydrodynamics one has the freedom to reconstruct either \mathbf{w} (primitive variables) or \mathbf{u} (evolved variables). For efficiency and accuracy considerations we reconstruct the first set, from which the remaining variables are obtained algebraically. The code uses slope-limiter methods to construct second-order TVD schemes by means of monotonic piecewise linear reconstructions of the cell-centered quantities. We use the standard minmod slope which provides the desired second-order accuracy for smooth solutions, while still satisfying the TVD property.

4. Results

The classical solution for an asymptotically uniform wind of pressureless gas past a compact source (modeled analytically by a point mass) was obtained by (Bondi and Hoyle, 1944). In this solution the material is focused at the rear part of the object as a result of the gravitational pull. For a pressureless gas, the density at this symmetry line could reach an infinite value and matter would flow on to the hole along this accretion line. However, when pressure is included in the model, a cylindrical shock forms around this line and the accretion proceeds along an accretion column of high density and pressure shocked material. The predicted final accretion pattern consists of a stationary conical shock with the material inside the accretion radius being captured by the central object. An schematic representation of this solution is depicted in Fig. 1.

A numerical evolution of relativistic wind accretion past a rapidly-rotating Kerr black hole ($a = 0.999M$, a specific angular momentum, M black hole mass) is depicted in Fig. 2 (left panel). This simulation shows the steady-state pattern in the equatorial plane of the black hole. The tail shock appears stable to tangential oscillations, in contrast to Newtonian simulations with tiny accretors (Benensohn et al., 1997); see (Font and Ibáñez, 1998b) for a related discussion. The accretion rates of mass and linear and angular momentum also show a stationary behavior (Font and Ibáñez, 1998b; Font et al., 1999). As opposed to the non-rotating black hole, in the rotating case the shock becomes wrapped around the central accretor, the effect being more pronounced as the black hole angular momentum a increases. The inner boundary of the domain is located at $r = 1.0M$ (*inside* the event horizon which, for this model, is at $1.04M$) which is only possible with the adopted regular coordinate system. The flow morphology shows smooth behavior when crossing the horizon, all matter fields being regular there.

The enhancement of the pressure in the post-shock zone is responsible for the “drag” force experienced by the accretor. The rotating black hole redistributes the high pressure area, with non-trivial effects on the nature of the drag force. The pressure enhancement is predominantly on the counter-rotating side. We observe a pressure difference of almost two orders of magnitude, along the axis normal to the asymptotic flow direction (Font et al., 1999). The implication of this asymmetry is that a rotating hole moving across the interstellar medium (or accreting from a wind), will experience, on top of the drag force, a “lift” force, normal to its direction of motion (to the wind direction). Although different in origin this feature bears a superficial resemblance with the Magnus effect of classical fluid dynamics.

The right panel of Fig. 2 shows how the accretion pattern would look like

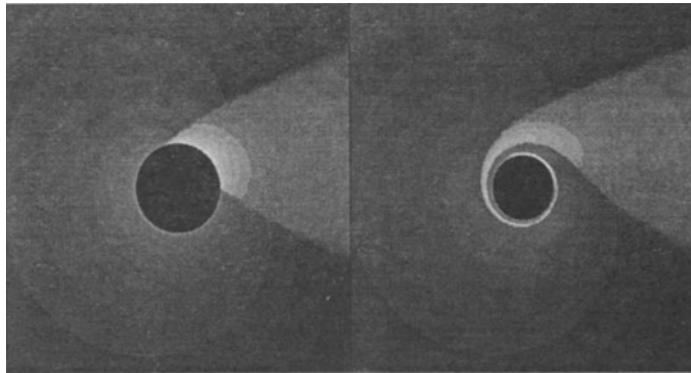


Figure 2. Relativistic wind accretion onto a rapidly rotating Kerr black hole ($a = 0.999M$, the hole spin is counter-clock wise) in Kerr-Schild coordinates (left panel). Initial model parameters: $v_\infty = 0.5$, $c_{s,\infty} = 0.1$ and $\Gamma = 5/3$. Iso-contours of the logarithm of the density are plotted at the final stationary time $t = 500M$. Brighter colours indicate high density regions while darker colours correspond to low density zones. The right panel shows how the flow solution looks like when transformed to Boyer-Lindquist coordinates. The shock appears here totally wrapped around the horizon of the black hole. The box is $12M$ units long. The simulation employed a (r, ϕ) -grid of 200×160 zones.

were the computation performed using the more common (though singular) Boyer-Lindquist coordinates. The transformation induces a noticeable wrapping of the shock around the central hole. The shock would wrap infinitely many times before reaching the horizon. As a result, the computation in these coordinates would be much more challenging than in Kerr-Schild coordinates, particularly near the horizon. Since the last stable orbit approaches closely the horizon in the case of maximal rotation, the interesting scenario of co-rotating extreme Kerr accretion would be severely affected by the strong gradients which develop in the strong-field region. This will most certainly affect the accuracy and, potentially, also the stability of numerical codes.

Acknowledgements

J.A.F. acknowledges financial support from a TMR fellowship of the European Union (contract nr. ERBFMBICT971902).

References

Banyuls F, Font JA, Ibáñez JM^a, Martí JM^a and Miralles JA (1997). Numerical 3+1

- general relativistic hydrodynamics: A local characteristic approach. *ApJ*, **476**, pp 221-231.
- Benensohn JS, Lamb DQ and Taam RE (1997). Hydrodynamical studies of wind accretion onto compact objects: Two-dimensional calculations. *ApJ*, **478**, pp 723-733.
- Bondi H and Hoyle F (1944). On the mechanism of accretion by stars. *MNRAS*, **104**, pp 273-282.
- Donat R and Marquina A (1996). Capturing shock reflections: an improved flux formula. *J. Comput. Phys.*, **125**, pp 42-58.
- Font JA and Ibáñez JM^a (1998). A numerical study of relativistic Bondi-Hoyle accretion on to a moving black hole: Axisymmetric computations in a Schwarzschild background. *ApJ*, **494**, pp 297-316.
- Font JA and Ibáñez JM^a (1998). Non-axisymmetric relativistic Bondi-Hoyle accretion on to a Schwarzschild black hole. *MNRAS*, **298**, pp 835-846.
- Font JA, Ibáñez JM^a and Papadopoulos P (1998). A horizon-adapted approach to the study of relativistic accretion flows on to rotating black holes. *ApJ*, **507**, pp L67-L70.
- Font JA, Ibáñez JM^a and Papadopoulos P (1999). Non-axisymmetric relativistic Bondi-Hoyle accretion on to a Kerr black hole. *MNRAS*, **305**, pp 920-936.
- Ibáñez JM^a, Aloy MA, Font JA, Martí JM^a, Miralles JA and Pons JA (2000). Riemann solvers in general relativistic hydrodynamics. This volume.
- Papadopoulos P and Font JA (1998). Relativistic hydrodynamics around black holes and horizon adapted coordinate systems. *Phys. Rev. D*, **58**, pp 024005,1-9.
- Papadopoulos P and Font JA (2000). Relativistic hydrodynamics on spacelike and null surfaces: Formalism and computations of spherically symmetric spacetimes. *Phys. Rev. D*, **61**, pp 024015,1-15.
- Ruffert M and Arnett D (1994). Three-dimensional hydrodynamic Bondi-Hoyle accretion. II. Homogeneous medium at Mach 3 with $\gamma = 5/3$. *ApJ*, **427**, pp 351-376.
- Shu CW and Osher S (1988). Efficient implementation of essentially non-oscillatory shock-capturing schemes. II. *J. Comput. Phys.*, **77**, pp 439-471.

A SECOND ORDER ACCURATE, SPACE-TIME LIMITED, BDF SCHEME FOR THE LINEAR ADVECTION EQUATION

S.A. FORTH

*Applied Mathematics & Operational Research Group
Cranfield University, RMCS Shrivenham
Swindon, SN6 8LA, U.K.
Email: S.A.Forth@rmcs.cranfield.ac.uk*

Abstract. Steady supersonic flow fields are frequently calculated by solution of the Euler or Parabolized Navier-Stokes (PNS) equations via a space-marching algorithm. Within space-marching the streamwise direction is treated in a time-like manner and the ensuing discretization allows solution of a 3 dimensional problem as a sequence of 2-D problems leading to high efficiency. The associated finite-volume discretizations are cell-centred in the crossflow direction and coincide with the mesh in the time-like space-marching direction. An alternative approach is the locally iterated method (Newsome et al., 1987) in which a plane-by-plane relaxation of the supersonic Euler or PNS equations is performed on a mesh centred in all 3 coordinates. Second order accuracy may be sought using an unlimited extrapolation procedure in the streamwise (time-like) direction.

In this work we consider the model problem of 1-dimensional linear advection. As noted previously (Thompson and Matus, 1989) we show that the above mentioned locally iterated methods may be regarded as implicit backward differentiation formulae of second order accuracy. If a purely upwind difference is taken in the streamwise direction then the resulting scheme is stable but dispersive. Such behaviour is explained by regarding components of the BDF time integration as a local extrapolation of the dependent variable in the time-direction and it may be eliminated by introducing limiters acting on gradients in space and time in a manner similar to that recently advocated (Sidilkover, 1998). Two schemes result from this analysis. The first is unconditionally TVD, but is second order accurate only under a CFL-like condition. The second scheme is second order accurate but subject to a CFL-like condition to maintain the TVD property. Results are presented for smooth and discontinuous solutions.

1. Space-Time Centred Schemes

We consider schemes that are centred in both space and time for the solution of the hyperbolic equation,

$$\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x} = 0, \quad (1)$$

where $f(u)$ is a flux function. In particular we shall consider the case $f(u) = au$, the linear advection equation, with $a > 0$.

1.1. DERIVATION OF SPACE-TIME CENTRED FINITE-VOLUME SCHEMES

We apply Gauss' theorem in space and time to a control volume centred about point u_j^n in space and time to obtain,

$$u_j^{n+\frac{1}{2}} = u_j^{n-\frac{1}{2}} - \frac{\Delta t}{\Delta x} \left(h_{j+\frac{1}{2}}^n - h_{j-\frac{1}{2}}^n \right). \quad (2)$$

where $u_j^{n\pm\frac{1}{2}}$, $h_{j\pm\frac{1}{2}}^n$ are numerical approximation to the space averaged value and time-averaged flux,

$$u_j^{n\pm\frac{1}{2}} \approx \bar{u}_j^{n\pm\frac{1}{2}} = \frac{1}{\Delta x} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u(x, t^{n\pm\frac{1}{2}}) dx, \quad (3)$$

$$h_{j\pm\frac{1}{2}}^n \approx \tilde{f}_{j\pm\frac{1}{2}} = \frac{1}{\Delta t} \int_{t^{n-\frac{1}{2}}}^{t^{n+\frac{1}{2}}} f(u(x_{j\pm\frac{1}{2}}, t)) dt. \quad (4)$$

1.1.1. Spatial Fluxes

For the spatial fluxes we use the spatially second-order accurate, flux-limited scheme,

$$h_{j+\frac{1}{2}}^n = a \left(u_j^n + \frac{1}{2} \phi(R_{j+\frac{1}{2}}^n) \Delta u_{j+\frac{1}{2}}^n \right), \quad (5)$$

where $\Delta u_{j+\frac{1}{2}}^n = u_{j+1}^n - u_j^n$, $\phi(R)$ is a limiter function (Sweby, 1984) and $R_{j+\frac{1}{2}}^n = \Delta u_{j-\frac{1}{2}}^n / \Delta u_{j+\frac{1}{2}}^n$. Note that since the flux is evaluated at the same point in time as the cell centred solution value then we would gain nothing from a second order in time flux formula. Thus equation (2) becomes,

$$u_j^{n+\frac{1}{2}} = u_j^{n-\frac{1}{2}} - C_{j-\frac{1}{2}}^n \Delta u_{j-\frac{1}{2}}^n \quad (6)$$

where the nonlinear coefficient $C_{j-\frac{1}{2}}^n$ is given by,

$$C_{j-\frac{1}{2}}^n = \frac{\nu}{2} \left(2 - \Phi(R_{j-\frac{1}{2}}^n) + \frac{\Phi(R_{j+\frac{1}{2}}^n)}{R_{j+\frac{1}{2}}^n} \right),$$

where $\nu = a\Delta t/\Delta x$ is the CFL number. For the resulting scheme to be TVD we require $0 \leq C_{j-\frac{1}{2}}^n \leq 1$ leading to conventional requirements of the flux limiter Φ (Sweby, 1984). The problem now remains to define the $u_j^{n \pm \frac{1}{2}}$ and this will be performed using space-time limited TVD extrapolation.

1.2. SPACE-TIME LIMITED TVD EXTRAPOLATION

We consider using an extrapolation in time,

$$u_j^{n+\frac{1}{2}} = u_j^n + \frac{1}{2}(u_j^n - u_j^{n-1})\Psi_j^n, \quad (7)$$

where Ψ_j^n is a limiter function yet to be defined. Clearly if $\Psi_j^n = 0$ then the scheme is first order accurate in time and equation (6) becomes identical to the unconditionally TVD backward Euler integration. If $\Psi_j^n = 1$ then we have a second order accurate, unlimited in time BDF scheme,

$$\frac{3}{2}u_j^n - 2u_j^{n-1} + \frac{1}{2}u_j^{n-2} = -C_{j-\frac{1}{2}}^n \Delta u_{j-\frac{1}{2}}. \quad (8)$$

Such a scheme has good linear stability properties but may yield non-monotone profiles near discontinuities. The reason for this is clearly seen in figure 1 in which extrapolation through two consecutive solutions with steep gradients leads to a non-monotone extrapolant. To prevent this we shall impose conditions on the streamwise limiter functions Ψ via a TVD analysis.

1.2.1. TVD Analysis: Approach 1

Using (7), gathering terms in $(u_j^n - u_j^{n-1})$, then following (Sidilkover, 1998), dividing by $(1 + \frac{1}{2}\Psi_j^n)$, defining,

$$r_j^n = -\frac{u_j^n - u_{j-1}^n}{u_j^n - u_j^{n-1}}, \quad (9)$$

and setting $\Psi_j^n = \Psi(r_j^n)$, we may rewrite equation (6) as

$$u_j^n + \frac{C_{j-\frac{1}{2}}^n}{1 + \frac{1}{2}\Psi(r_j^n)} \Delta u_{j-\frac{1}{2}}^n = u_j^{n-1} - \frac{\Psi(r_j^{n-1})}{2r_j^{n-1} \left(1 + \frac{1}{2}\Psi(r_j^n)\right)} \Delta u_{j-\frac{1}{2}}^{n-1}. \quad (10)$$

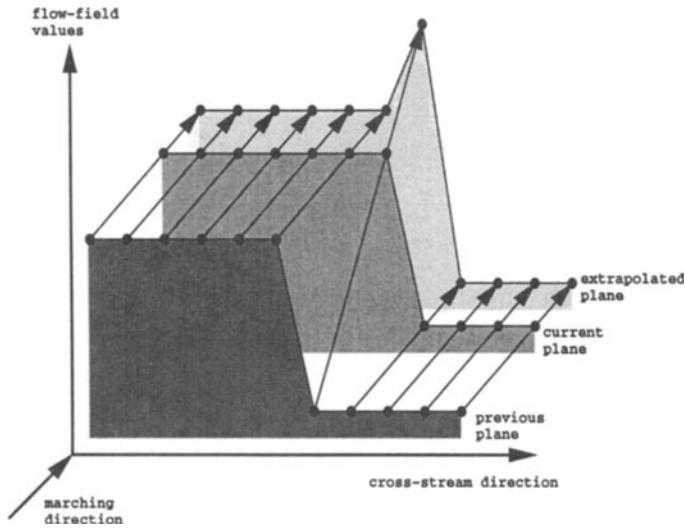


Figure 1. Extrapolation may produce non-monotone profiles

This equation is of the form $L \cdot u^n = R \cdot u^{n-1}$, where L and R are finite-difference operators and so (Harten, 1984, Lemma 3.1) may be applied giving that the scheme is TVD provided the implicit, left hand side operator L is Total Variation Increasing (TVI) and provided the explicit operator R is Total Variation Decreasing.

Now using (Harten, 1984, Lemma 3.2) we see that L is TVI provided, $0 \leq C_{j-\frac{1}{2}}^n / (1 + \frac{1}{2}\Psi(r_j^n)) \leq K$ where $K < \infty$ is some constant. This is assured by the boundedness of $C_{j-\frac{1}{2}}^n$ and provided

$$\Psi(r_j^n) \geq k > -2, \quad (11)$$

for some constant k .

The condition for the right hand side operator of equation (10) to be TVD leads to,

$$\bar{\Psi} - \Psi(s) \leq 2 \forall s, \quad (12)$$

where $\bar{\Psi} = \max_r \Psi(r)/r$.

Thus under Harten's Lemma our scheme will be TVD provided the conditions of equations (11) and (12) hold. We note that, somewhat unusually, in these conditions there is no explicit CFL number dependence.

Since our scheme evaluates spatial fluxes at the central point in time then we are assured of the second order spatial accuracy given by an appropriate spatial TVD scheme. Accuracy in time is governed by the limiter

applied to the gradients in time. In time we have the extrapolated value given by equation (7). Using a Taylor series expansion for fixed CFL number, we find that $r_j^n = -\frac{1}{\nu} + O(\Delta t)$, and so,

$$u_j^{n+\frac{1}{2}} = u_j^n + \frac{\Delta t}{2} u_t \Psi\left(\frac{1}{\nu}\right) + O(\Delta t^2).$$

Thus for the extrapolation to be first order accurate, and the overall scheme to be second order accurate, we need $\Psi(1/\nu) = 1$. This accuracy condition greatly restricts the choice of limiter function that may be used. We require the limiter to take the value 1 whenever possible. Of the commonly used limiters in TVD schemes only minmod, $\text{minmod}(r) = \max(0, \min(r, 1))$, takes the value 1 for a large range $r \geq 1$ of its argument. In this case the scheme would give second order accuracy under the CFL-like condition $\nu \leq 1$. An alternative limiter newlim could be defined by,

$$\text{newlim}(r) = \max(0, \min(2r, 1)),$$

which would be second order accurate for $\nu \leq 2$.

1.2.2. TVD Analysis: Approach 2

To eliminate the CFL number dependence on the accuracy of the scheme we now replace the limiter variable r_j^n of equation (9) by,

$$s_j^n = -\frac{\nu(u_j^n - u_{j-1}^n)}{u_j^n - u_{j-1}^n}, \quad (13)$$

and the scheme may now be written as,

$$u_j^n + \frac{C_{j-\frac{1}{2}}^n}{1 + \frac{1}{2}\Psi(s_j^n)} \Delta u_{j-\frac{1}{2}}^n = u_j^{n-1} - \frac{\nu\Psi(s_j^{n-1})}{2s_j^{n-1} \left(1 + \frac{1}{2}\Psi(s_j^n)\right)} \Delta u_j^{n-1}. \quad (14)$$

The conditions (11) for the implicit operator are unaffected. For the explicit operator we require,

$$\nu \leq \frac{2 + \min_s \Psi(s)}{\bar{\Psi}}. \quad (15)$$

The second order accuracy condition becomes the more conventional $\Psi(1) = 1$, satisfied by all conventional TVD limiter functions. Now equation (15) gives a CFL like condition for the scheme to be TVD. For example the minmod limiter would have a CFL restriction of $\nu \leq 2$.

In the next section we present some results comparing the two schemes described above.

2. Numerical Results

We present two sets of results corresponding to solution of the linear advection equation with unit advection speed, $\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = 0$, on the domain $-1 \leq x \leq 1$. The first test case corresponds to the exact solution,

$$u(x, t) = \frac{1}{2}(1 - \sin \pi(x - t)),$$

with the initial condition and left-hand side boundary set exactly. We refer to this as the *smooth test case* and we obtain solutions for $0 \leq t \leq 2$. The second *discontinuous test case* corresponds to the solution,

$$u(x, t) = \begin{cases} 1 & \text{for } x \leq -\frac{1}{2} + t, \\ 0 & \text{for } x > -\frac{1}{2} + t. \end{cases}$$

for $0 \leq t \leq 1$. In figures 2 and 3 we plot the estimated numerical order of

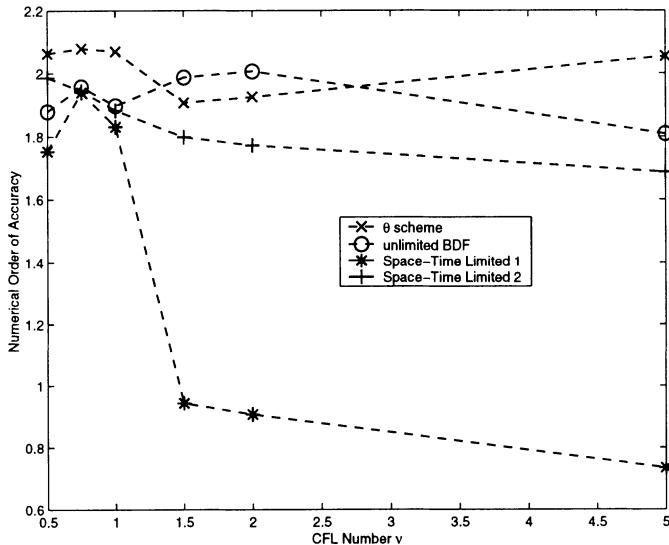


Figure 2. Numerical order of accuracy for smooth test case

accuracy $p = \log(e_{n/2}/e_n)/\log 2$ (where e_n is the solution error in the L_1 norm) for the smooth and discontinuous test cases against CFL number. These results were obtained for $n = 120$ and $n/2 = 60$ via a MATLAB code. Also included for comparison are results from a θ -type integration (with $\theta = 1/2$ for second order accuracy) and the unlimited BDF scheme of equation (8). All three schemes aspire to be second order accurate.

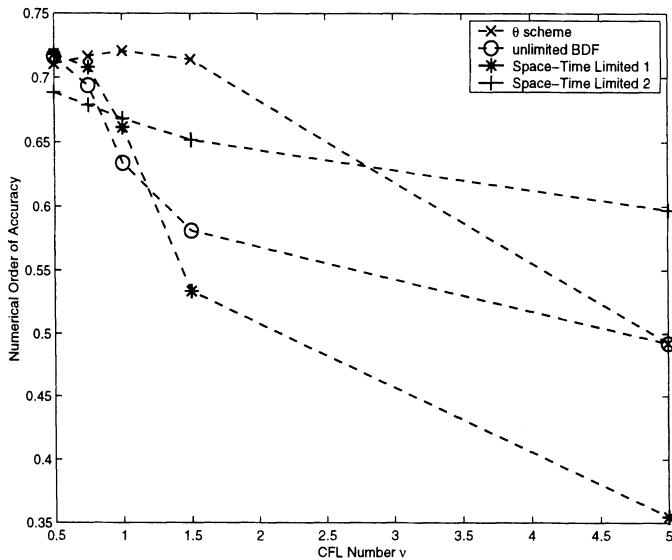


Figure 3. Numerical order of accuracy for discontinuous test case

For the smooth test case the θ -scheme, unlimited BDF scheme and space-time limited 2 scheme are all approaching second order accuracy for all CFL numbers. The space-time limited scheme of approach 1 is seen to drop to first order accuracy for CFL numbers greater than 1 as our analysis predicts.

For the discontinuous test case a second order scheme should only achieve a numerical order of accuracy of $2/3$ (LeVeque, 1992, p115). This is the case for the θ -scheme (except at CFL number $\nu = 5$ when it is no longer TVD), the unlimited BDF scheme (for $\nu \leq 1$), the space-time limited scheme 1 (within the TVD range $\nu \leq 1$) and the space-time limited scheme 2 (for all ν). These results being in accordance with either established theory or that of this paper.

3. Conclusions

We have conducted an analysis, using limiters in space and time, to ensure that a BDF scheme applied to the linear advection equation is TVD. Limiting may be applied in two ways: one giving an unconditionally TVD scheme but which drops to first order accuracy for CFL numbers larger than some limiter dependent value; the second being second order accurate but conditionally TVD. Both schemes are fully conservative.

The first scheme is a candidate for use on meshes where the grid spacing in the time-like direction is pre-determined. This is the case when space-

marching on existing Navier-Stokes meshes. The second, though mathematically elegant, requires modification of the time-step, via the CFL-like condition, for robust use.

References

- Newsome R W, Walters R W and Thomas J L (1987) An efficient iteration strategy for upwind/relaxation solutions to the thin-layer Navier-Stokes equations. *AIAA Conference Proceedings*, **87-1112-CP**
- Thompson D S and Matus R J (1989) Conservation errors and convergence characteristics of iterative space-marching algorithms. *AIAA Conference Proceedings*, **89-1935-CP**
- Sidilkover D (1998) A new time-space accurate scheme for hyperbolic problems I: quasi-explicit case. *ICASE Report* **98-25**
- Sweby P K (1984) High resolution schemes using flux limiters for hyperbolic conservation laws. *SIAM Journal on Numerical Analysis* **21**: 5
- Harten A (1984) On a class of high resolution total-variation-stable finite-difference schemes. *SIAM Journal on Numerical Analysis* **21**: 1
- LeVeque R J (1992). Numerical Methods for Conservation Laws. Birkhäuser Verlag,

MULTIDIMENSIONAL UPWIND SCHEMES: APPLICATION TO HYDRAULICS

P. GARCIA-NAVARRO

Fluid Mechanics,

University of Zaragoza, Spain.

Email: pigar@posta.unizar.es

M.E. HUBBARD

DAMTP, Cambridge, UK.

Email: M.E.Hubbard@damtp.cam.ac.uk

AND

P. BRUFAU

Fluid Mechanics,

University of Zaragoza, Spain.

Email: cucu@ideafix.cps.uniza.es

Abstract. In the field of the numerical simulation of conservation laws, upwind and TVD techniques have progressively gained acceptance. Originally, they were derived for homogeneous scalar equations or systems of equations in one spatial dimension. Their extension to more than one spatial dimension is not straightforward.

A widespread version is based on a piecewise constant representation of the solution inside cells and the application of a one dimensional Riemann solver across every cell edge. This leads to a finite volume technique based on the integral form of the equations to which the divergence theorem has been applied. In a philosophy different from concentrating on finite volumes and the changes of the variables across the cell sides, it is possible to consider solutions on grids in which the unknowns are associated with the vertices and updates to these nodal values are through the advection of linear wave solutions. This avoids the problems of taking the normal to the cell interfaces as a privileged direction. This second group of methods are based on a piecewise linear continuous representation.

Some years after their adoption for solving problems in gas dynamics, upwind schemes have been successfully used for the solution of the shallow water equations, with similar advantages. We consider the use of these tech-

niques for 2D shallow water flows and the question of whether they may be of practical use. The basis of the numerical methods is stated and their application to the shallow water system is described. Finally, some numerical results are presented.

Two dimensional wave decomposition and multi-dimensional upwinding seem a promising method of solution for the 2D shallow water equations. Two wave models have been adapted from Gas Dynamics to render the technique suited to hydraulic problems with shocks. Although the procedure is more complicated than present day generalizations of 1D upwinding techniques it is cell-based, which makes it competitive versus the edge based finite volume techniques. Both can be applied on a triangular discretization and, by taking advantage of the triangles, they can clearly be applied to arbitrary geometries, a great advantage for hydraulic engineers working on practical problems, and there is a wide variety of possibilities concerning grid movement and refinement.

1. Introduction

Upwind methods are very popular in the modelling of advection dominated flows, and in particular those which contain strong discontinuities. The essence of the upwind method in 1D depends on the reduction of the problem to a set of subproblems that are (almost) independent. The best solution techniques for these scalar subproblems can then be studied carefully and in detail(LeVeque, 1992). These methods are frequently used to solve systems of equations in higher dimensions.

Initial attempts to extend 1D upwind techniques to higher dimensions were all based on 1D upwind concepts applied within a dimensional splitting framework, and modelling the flow by solving simple Riemann problems across cell interfaces. This introduced an undesirable reliance on the computational mesh, and such techniques were not capable of adequately resolving shocks or shears which were not aligned with the grid (Deconinck et al., 1992).

Unstructured grids have many advantages for multidimensional flow analysis, particularly their flexibility when constructing boundary fitted grids for complex geometries and their general lack of preferential grid directions. Even so, locally, schemes based on structured and unstructured grids are the same when the numerical flux normal to the cell face is only evaluated in the 1D manner. The result is a solution algorithm which depends on geometrical variables which have little or no relation to the rel-

event flow directions. Specifically, cell boundaries are used to define a 1D direction along which the upwinding takes place. Consequently, the choice of the grid has a disproportionate influence on the solution.

It was soon realised that it was necessary to incorporate genuinely multidimensional physics into these algorithms. The first step was taken by Davis (Davis, 1984) who suggested that the shock capturing capabilities of upwind methods could be improved by rotating the Riemann problem to align it with the direction of physically important flow gradients. They take into account variables like flow direction or velocity gradient direction over a cell face as part of the discretization. This work was extended by Levy et al. (Levy et al., 1991) and Tamura and Fujii (Tamura and Fujii, 1991), but these methods commonly suffered problems with robustness.

An alternative method was developed independently by Rumsey et al. (Rumsey et al., 1991) and Parpia and Michalek (Parpia and Michalek, 1993). Common to these methods is the fact that the multidimensional physics is added at the cell interfaces, thus retaining some 1D aspects. Therefore new multidimensional upwind schemes for equations in more than one space dimension are still being sought which don't assume any one-dimensionality along the grid lines or normal to the cell faces.

The methods discussed in this paper use a genuinely multidimensional physical model for the upwinding which does not fit in to the standard finite volume approach where the representation of the unknowns is considered to be only piecewise continuous. In this respect the new schemes are much closer to finite element methods based on linear elements, with which they share a continuous piecewise linear representation over the cells. On the other hand, they share with upwind methods the properties of asymmetric upwinded stencils and control of monotonicity across discontinuities, and they can be considered as truly multidimensional generalizations of the 1D TVD upwind methods.

The basis of one of the original groups of these multidimensional upwinding techniques is the assumption that any observed gradients in the initial data at the start of a time step are linked to the presence of simple waves in the flow. Since an infinite number of simple wave patterns could be responsible for the same observed gradients, it is necessary to hypothesize the number and nature of the waves present: this is known as a wave model. It is important that the orientation of the waves are not constrained by the directions of the grid. The next idea is to update the solution in a way that acknowledges the direction of propagation of each wave in the model.

Initial calculations with the simple wave models confirmed an excellent capability to capture oblique discontinuities, for which the multidimensional upwind method was designed, but proved somewhat disappointing in the calculation of smooth (subcritical) flows. This is because the wave models

still employ unnecessary dissipation (Deconinck et al., 1994).

Deconinck, Hirsch and Peuteman (Deconinck et al., 1986) concurrently devised their own alternative strategy for decomposing the fluctuation. This was based on an attempt to find an approximate diagonalization of the system of equations via an appropriate similarity transformation. The original schemes lacked robustness but the idea has subsequently been improved upon by applying the diagonalization technique to a preconditioned set of equations. The correct choice of preconditioner leads to a maximally decoupled system of equations and provides a very accurate method for calculating steady state solutions to nonlinear systems of conservation laws (Deconinck et al., 1997; Mesaros and Roe, 1995; Pailiere et al., 1995).

Upwind finite volume schemes are improved by higher order interpolation, where more than the direct neighbour points are used. This is not the case for multidimensional upwind schemes, but accuracy remains a critical issue. Where steady state solutions are of interest, an essential feature of a numerical scheme is the convergence towards those steady states. Therefore, care has to be taken to provide reasonable convergence properties for multidimensional upwind scheme, in order for them to be competitive with other methods, and the issue of convergence acceleration has been studied in some detail (Deconinck et al., 1997). Unsteady problems have only really been studied in depth more recently, and then mainly in the scalar case (Hubbard and Roe, 1999; Hubbard, 1999), but evidence will be presented in this paper that even the schemes developed for steady state problems provide a significant improvement in the modelling of nonlinear problems over comparable finite volume schemes. Even so, it has become apparent that a rather fundamental distinction should be made between computational methods for steady and unsteady flows. The simple wave model approach is discussed in some detail here, not because it gives the best steady state solutions, but because it currently appears to be the most appropriate form of multidimensional upwinding for application to time-dependent problems.

The issue of approximating source terms is also discussed briefly, and suggestions made as to how they should be incorporated within the overall discretisation in order to maintain the accuracy of the homogeneous scheme, something which has so far proved to be simpler for the diagonalization approach.

2. Basic Scalar Technique

The technique for 2D problems assumes that the physical domain is discretized using triangular cells (these methods are less naturally applied to quadrilaterals (Deconinck et al., 1997)) and a set of solution values is stored at the nodes of the mesh. For each cell T , and for a linear scalar equation

of the form

$$\frac{\partial w}{\partial t} + \mathbf{a} \cdot \nabla w = 0, \quad \mathbf{a} = (a_x, a_y) \quad (1)$$

where \mathbf{a} is a constant vector, the fluctuation is defined as

$$\phi_T = \int_T \frac{\partial w}{\partial t} dS = - \int_T \mathbf{a} \cdot \nabla w dS \quad (2)$$

and the closely related quantity, the cell residual R_T , as

$$R_T = -\frac{1}{S_T} \phi_T = \frac{1}{S_T} \int_T \mathbf{a} \cdot \nabla w dS = -\frac{1}{S_T} \oint_C w \mathbf{a} \cdot \hat{\mathbf{n}} dC \quad (3)$$

where C represents the cell boundary and \mathbf{n} the inward normal to that boundary. Note that the final expression in (3) is only valid if $\nabla \cdot \mathbf{a} = 0$. The fluctuation contains information on the state of the cell and the distribution of this information to the nodes must be done in a way which ensures conservation (Deconinck et al., 1992).

From the properties of the normals in the cell and the additional assumption that the solution varies linearly within each element, it is possible to identify a discrete approximation of ∇w (Deconinck et al., 1992),

$$\nabla w_T = \frac{1}{2S_T} \sum_{i=1}^3 w_i \mathbf{n}_i \quad (4)$$

such that

$$R_T = \frac{1}{S_T} \int_T \mathbf{a} \cdot \nabla w dS = \frac{1}{S_T} \mathbf{a} \cdot \nabla w_T \int_T dS = \mathbf{a} \cdot \nabla w_T. \quad (5)$$

Equivalently,

$$\phi_T = - \sum_{i=1}^3 w_i k_i \quad (6)$$

which introduces the quantities $k_i = \frac{1}{2} \mathbf{a} \cdot \mathbf{n}_i$, containing information about the direction of the advection speed relative to the cell edges. The k_i can be used to decide whether flow enters or leaves the triangle through a particular edge and, in that sense, form a useful tool for imposing the upwind properties of the method.

Residuals and fluctuations are both cell-based quantities, and will be used to update the nodal solution values. For this purpose we need to introduce quantities known as distribution coefficients, D_T^i . By using a

simple forward Euler time differencing the following procedure can now be defined to update the variables at all the nodes in a single cell

$$\begin{aligned} S_1 w_1^{n+1} &= S_1 w_1^n - \Delta t S_T D_T^1 R_T^n \\ S_2 w_2^{n+1} &= S_2 w_2^n - \Delta t S_T D_T^2 R_T^n \\ S_3 w_3^{n+1} &= S_3 w_3^n - \Delta t S_T D_T^3 R_T^n \end{aligned} \quad (7)$$

where $S_i = \frac{1}{3} \sum_{T_i} S_T$ is the area of the median dual cell around node i , one third the total area of the triangles having i as vertex. Note that for simplicity a cell residual only contributes to its own vertices, so the condition $\sum_{i=1}^3 D_T^i = 1$ ensures conservation and consistency.

There exist many criteria for the design of advection schemes according to the choice of the distribution coefficients D_T^i . Two properties are of prime interest, positivity and linearity preservation:

- Positivity means that every solution value at the new time level can be written as a convex combination of old solution values, and is related to the 1D property of monotonicity.
- Linearity preservation has to do with higher order accuracy. It requires that the scheme preserves the exact steady state solution whenever this is a linear function in space, and for an arbitrary triangulation of the domain. This is closely related to the idea of second order accuracy in the context of finite difference schemes, although it is an accuracy requirement on the steady state space discretization only. Less accuracy is gained in the time-dependent case, but it is still significant.

It can be proved that a linear scheme cannot be both positive and linearity preserving (Deconinck et al., 1992), a result which is closely related to Godunov's theorem on the incompatibility between second order accuracy and monotonicity preservation for linear schemes in one dimension. Therefore, in order to have both of the above properties, nonlinear schemes must be considered where the update coefficients depend on the data. This leads to the generation of nonlinear schemes even for linear equations. The most commonly used of these nonlinear schemes is the PSI scheme (Deconinck et al., 1997) which is based on the linear, positive N scheme, and can be thought of in one of two ways. In its original derivation, it is based on the combination of optimal positive upwind (N) schemes based on two distinct advection velocities, the usual one \mathbf{a} , and its component in the direction of the local solution gradient, $(\mathbf{a} \cdot \nabla w)\nabla w$. More recently, it was written as the N scheme combined with a form of cross-stream limiter (Sidilkover and Roe, 1995).

The underlying advection schemes used for the shallow water equations are no different to those used for the Euler equations so we will not go into further detail about their particular construction and description. We

refer the reader to the very good reviews in (Deconinck et al., 1997) and (Deconinck et al., 1994).

2.1. NONLINEAR EQUATIONS

When the equation itself is nonlinear a suitable linearization must be performed before the technique described above is applied. In simple cases an averaged advection speed which satisfies discrete conservation can be found by assuming linear variation of the conserved quantities w over the cell, leading to a constant gradient ∇w . However, a general nonlinear 2D system of conservation laws of the form

$$\frac{\partial \mathbf{U}}{\partial t} + \nabla \cdot \mathbf{F} = 0, \quad \mathbf{F} = (\mathbf{f}, \mathbf{g}) \quad (8)$$

requires the construction of an appropriate discrete form of the system

$$\frac{\partial \mathbf{U}}{\partial t} + (\mathbf{A}, \mathbf{B}) \cdot \nabla \mathbf{U} = 0. \quad (9)$$

In particular, a consistent approximation for the cell residual is sought

$$\mathbf{R}_T = \frac{1}{S_T} \int_T (\mathbf{A}, \mathbf{B}) \cdot \nabla \mathbf{U} dS = (\tilde{\mathbf{A}}_T, \tilde{\mathbf{B}}_T) \cdot \nabla \mathbf{U}_T \quad (10)$$

where $\tilde{\mathbf{A}}_T$, $\tilde{\mathbf{B}}_T$ are discrete averages of the Jacobian matrices, constructed from the nodal values. Now, assuming linear variation of the conservative variables \mathbf{U} over each cell enables us to write

$$\mathbf{R}_T = \frac{1}{S_T} \nabla \mathbf{U}_T \cdot \int_T (\mathbf{A}, \mathbf{B}) dS \quad (11)$$

from which discrete cell gradients and cell Jacobians can be defined:

$$\tilde{\mathbf{A}} = \frac{1}{S_T} \int_T \mathbf{A} dS, \quad \tilde{\mathbf{B}} = \frac{1}{S_T} \int_T \mathbf{B} dS. \quad (12)$$

Unfortunately, the exact evaluation of the above integrals is not practical either for the Euler or for the shallow water equations. Roe et al. (Deconinck et al., 1993) suggested the introduction of “parameter vector” variables for a simpler treatment of the former system. The strategy followed for the shallow water equations is similar and makes use of the set of primitive variables, but there is no set of variables which can quite play the role of the parameter vector, as will be discussed in the following section.

3. The 2D Shallow Water System

In the conservative and homogeneous version of the shallow water system of equations with independent variables $\mathbf{U} = (h, hu, hv)^T$, where h , u and v are the depth and x - and y -velocities respectively, the fluxes are

$$\mathbf{f} = \left(hu, hu^2 + g \frac{h^2}{2}, huv \right)^T, \quad \mathbf{g} = \left(hv, huv, hv^2 + g \frac{h^2}{2} \right)^T. \quad (13)$$

and the residual is defined as (*cf.* Equation (10))

$$\mathbf{R}_T = \frac{1}{S_T} \int_T (\mathbf{f}_x + \mathbf{g}_y) dS = -\frac{1}{S_T} \oint_C (\mathbf{f}, \mathbf{g}) \cdot \hat{\mathbf{n}} dC. \quad (14)$$

We are seeking a conservative linearization of the Jacobians satisfying

$$\mathbf{R}_T = \tilde{\mathbf{f}}_x + \tilde{\mathbf{g}}_y = \tilde{\mathbf{A}} \tilde{\mathbf{U}}_x + \tilde{\mathbf{B}} \tilde{\mathbf{U}}_y. \quad (15)$$

In order to simplify the subsequent evaluation of the discrete flux Jacobians (*cf.* Equation (12)) we can use the transformation matrix $\mathbf{M} = \frac{\partial \mathbf{U}}{\partial \mathbf{V}}$ to move from the conserved variables \mathbf{U} to the primitive variables $\mathbf{V} = (h, u, v)^T$. This requires the definition of new Jacobian matrices, \mathbf{S} and \mathbf{T} , as

$$\mathbf{S} = \frac{\partial \mathbf{f}}{\partial \mathbf{V}} = \frac{\partial \mathbf{f}}{\partial \mathbf{U}} \frac{\partial \mathbf{U}}{\partial \mathbf{V}} = \mathbf{A} \mathbf{M}, \quad \mathbf{T} = \frac{\partial \mathbf{g}}{\partial \mathbf{V}} = \frac{\partial \mathbf{g}}{\partial \mathbf{U}} \frac{\partial \mathbf{U}}{\partial \mathbf{V}} = \mathbf{B} \mathbf{M} \quad (16)$$

so that

$$\mathbf{f}_x + \mathbf{g}_y = \mathbf{f}_{\mathbf{V}} \mathbf{V}_x + \mathbf{g}_{\mathbf{V}} \mathbf{V}_y = \mathbf{S} \mathbf{V}_x + \mathbf{T} \mathbf{V}_y. \quad (17)$$

Under the assumption that the variables \mathbf{V} are linear over the cells T the gradients $\nabla \mathbf{V}$ are constant, and this enables us to write the residual as

$$\begin{aligned} \mathbf{R}_T &= \frac{1}{S_T} \int_T (\mathbf{S}(\mathbf{V}) \mathbf{V}_x + \mathbf{T}(\mathbf{V}) \mathbf{V}_y) dS \\ &= \frac{1}{S_T} \left(\int_T \mathbf{S}(\mathbf{V}) dS \right) \mathbf{V}_x + \left(\int_T \mathbf{T}(\mathbf{V}) dS \right) \mathbf{V}_y \\ &= \tilde{\mathbf{S}} \mathbf{V}_x + \tilde{\mathbf{T}} \mathbf{V}_y \end{aligned} \quad (18)$$

with the definitions

$$\tilde{\mathbf{S}} = \frac{1}{S_T} \int_T \mathbf{S}(\mathbf{V}) dS, \quad \tilde{\mathbf{T}} = \frac{1}{S_T} \int_T \mathbf{T}(\mathbf{V}) dS. \quad (19)$$

We can evaluate $\tilde{\mathbf{S}}$ and $\tilde{\mathbf{T}}$ exactly, since \mathbf{S} and \mathbf{T} both vary quadratically with \mathbf{V} but, in order for the wave model to be employed, the approximate Jacobians are required to take the form $\tilde{\mathbf{S}} = \mathbf{S}(\tilde{\mathbf{V}})$ and $\tilde{\mathbf{T}} = \mathbf{T}(\tilde{\mathbf{V}})$. $\tilde{\mathbf{V}}$ can

be simply calculated by averaging over the nodal values at the vertices of the triangle T but, unlike with the Euler equations, this leaves a small correction term which must be added for conservation (Hubbard and Baines, 1997) (and depends on the choice of independent variables \mathbf{V}).

We can now reverse the transformation to define linearized derivatives of the conservative variables,

$$\tilde{\mathbf{U}}_x = \mathbf{M}(\tilde{\mathbf{V}}) \nabla \mathbf{V}_x, \quad \tilde{\mathbf{U}}_y = \mathbf{M}(\tilde{\mathbf{V}}) \nabla \mathbf{V}_y \quad (20)$$

and rewrite

$$\mathbf{R}_T = \tilde{\mathbf{R}} \mathbf{M}^{-1}(\tilde{\mathbf{V}}) \tilde{\mathbf{U}}_x + \tilde{\mathbf{S}} \mathbf{M}^{-1}(\tilde{\mathbf{V}}) \tilde{\mathbf{U}}_y = \tilde{\mathbf{A}} \tilde{\mathbf{U}}_x + \tilde{\mathbf{B}} \tilde{\mathbf{U}}_y \quad (21)$$

where $\mathbf{M} = \frac{\partial \mathbf{U}}{\partial \mathbf{V}}$, which allows the identification of $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$.

Having carried out this linearization, we have to compute the discrete residuals \mathbf{R}_T (or fluctuations $\phi_T = -S_T \mathbf{R}_T$) and distribute them to the vertices of the cells by means of an advection scheme (as discussed in Section 2). For this purpose it is necessary to decompose the residual into simple pieces, for example scalar components that can be explained as due to the passage of a simple wave. This requires a description of the wave models.

4. Simple Wave Models

Consider the linearized system of equations written in primitive variables

$$\frac{\partial \mathbf{V}}{\partial t} + \tilde{\mathbf{E}} \frac{\partial \mathbf{V}}{\partial x} + \tilde{\mathbf{H}} \frac{\partial \mathbf{V}}{\partial y} = 0. \quad (22)$$

A simple wave solution can be found according to Roe (Roe, 1986; Roe, 1986) and it is possible then to express the gradient of the independent variables as a sum

$$\nabla \mathbf{V} = \sum_{k=1}^{N_w} \alpha^k \mathbf{r}^k \mathbf{n}^k, \quad (23)$$

in which N_w is the number of waves in the decomposition and $\mathbf{n}^k = (\cos \theta^k, \sin \theta^k)$ gives the direction of propagation of the particular wave. The vectors \mathbf{r}^k are right eigenvectors of the matrix $\mathbf{M}^* = \tilde{\mathbf{E}} \cos \theta + \tilde{\mathbf{H}} \sin \theta$ and take one of three forms (representing two types of gravity wave and a shear wave):

$$\mathbf{r}_{G1} = \begin{pmatrix} 1 \\ \frac{g}{c} \cos \theta \\ \frac{g}{c} \sin \theta \end{pmatrix}, \quad \mathbf{r}_{G2} = \begin{pmatrix} 1 \\ -\frac{g}{c} \cos \theta \\ -\frac{g}{c} \sin \theta \end{pmatrix}, \quad \mathbf{r}_S = \begin{pmatrix} 0 \\ -\sin \theta \\ \cos \theta \end{pmatrix}, \quad (24)$$

where \tilde{c} is the velocity of small perturbations in still water, the equivalent of the speed of sound in gas-dynamics, and is given by $\tilde{c} = (\tilde{g}\tilde{h})^{1/2}$. The variables α^k represent weighting coefficients of the sum.

The connection between the gradient of the primitive variables and that of the averaged conservative variables (20) can be used to develop the latter as

$$\tilde{\mathbf{U}}_x = \sum_{k=1}^{N_w} \alpha^k \mathbf{r}_c^k \cos \theta^k, \quad \tilde{\mathbf{U}}_y = \sum_{k=1}^{N_w} \alpha^k \mathbf{r}_c^k \sin \theta^k \quad (25)$$

where now, \mathbf{r}_c^k represent the right eigenvectors of the matrix $\mathbf{M}^* = \tilde{\mathbf{A}} \cos \theta + \tilde{\mathbf{B}} \sin \theta$ and can be calculated using the relationship $\mathbf{r}_c^k = \mathbf{M}(\tilde{\mathbf{V}})\mathbf{r}^k$. The two matrices \mathbf{M}^* and \mathbf{M}_c^* share the same set of eigenvalues λ^k given, for the three wave types, by

$$\begin{aligned} \lambda_{G1} &= \tilde{u} \cos \theta + \tilde{v} \sin \theta + \tilde{c} \\ \lambda_{G2} &= \tilde{u} \cos \theta + \tilde{v} \sin \theta - \tilde{c} \\ \lambda_S &= \tilde{u} \cos \theta + \tilde{v} \sin \theta. \end{aligned} \quad (26)$$

Note that by implication α , λ and \mathbf{r} are all evaluated at the cell average state $\tilde{\mathbf{V}}$ (see previous section). It now follows from (25) that the residual then can be split into

$$\mathbf{R}_T = \tilde{\mathbf{A}}\tilde{\mathbf{U}}_x + \tilde{\mathbf{B}}\tilde{\mathbf{U}}_y = \sum_{k=1}^{N_w} \alpha^k \lambda^k \mathbf{r}_c^k \quad (27)$$

so that it can be written in the form

$$\mathbf{R}_T = \sum_{k=1}^{N_w} (\lambda^k \cdot \nabla w^k) \mathbf{r}_c^k \quad (28)$$

for appropriate choices of the wave velocities λ^k and “characteristic” gradient ∇w . This is simply a sum of components of precisely the form seen in (5).

We next describe two of the most successful simple wave models proposed in the literature as suitable to accomplish the above decomposition.

4.1. ROE'S WAVE MODELS

The wave decomposition of the gradient of the primitive variables,

$$\nabla \mathbf{V} = \sum_{k=1}^{N_w} \alpha^k \mathbf{r}^k \mathbf{n}^k, \quad (29)$$

represents a system of six equations in the 2D shallow water case, where we have two spatial derivatives for each of the three variables. Therefore, it allows for six unknowns. These must correspond to coefficients or angles of propagation of suitable choices of waves whose advection will represent the total fluctuation.

Following Roe's suggested Model D for the treatment of the Euler equations (Roe, 1986), a splitting can be made into four orthogonal acoustic waves, labelled by their strengths (coefficients) and one angle θ which determines the four directions (α_1, θ) , $(\alpha_2, \theta + \pi)$, $(\alpha_3, \theta + \frac{\pi}{2})$, $(\alpha_4, \theta + \frac{3\pi}{2})$, along with one shear wave (β, ϕ) of strength β at an angle ϕ . The six unknowns are taken to be α_1 , α_2 , α_3 , α_4 , β and θ . The value of the angle ϕ is determined in terms of the solution as

$$\phi = \theta - \frac{\pi}{4} \operatorname{sign}(\beta). \quad (30)$$

Making use of the equivalents of the basic trigonometric functions to those of the first quadrant of the unit radius circle, and after some algebraic manipulations (Garcia-Navarro et al., 1995),

$$\beta = v_x - u_y, \quad \tan 2\theta = \frac{u_y + v_x}{u_x - v_y} \quad (31)$$

and

$$\begin{aligned} \alpha_1 &= \frac{1}{2} \left(h_x \cos \theta + h_y \sin \theta + \frac{c}{g} \left(\frac{u_x \cos^2 \theta - v_y \sin^2 \theta}{\cos 2\theta} - \frac{1}{2} |\beta| \right) \right) \\ \alpha_2 &= \frac{1}{2} \left(-(h_x \cos \theta + h_y \sin \theta) + \frac{c}{g} \left(\frac{u_x \cos^2 \theta - v_y \sin^2 \theta}{\cos 2\theta} - \frac{1}{2} |\beta| \right) \right) \\ \alpha_3 &= \frac{1}{2} \left(h_y \cos \theta - h_x \sin \theta + \frac{c}{g} \left(\frac{v_y \cos^2 \theta - u_x \sin^2 \theta}{\cos 2\theta} + \frac{1}{2} |\beta| \right) \right) \\ \alpha_4 &= \frac{1}{2} \left(-(h_y \cos \theta - h_x \sin \theta) + \frac{c}{g} \left(\frac{v_y \cos^2 \theta - u_x \sin^2 \theta}{\cos 2\theta} + \frac{1}{2} |\beta| \right) \right). \end{aligned} \quad (32)$$

λ and ∇w are then constructed so that the advection velocities each take one of the appropriate forms represented by

$$\lambda_{G1} = \begin{pmatrix} u + c \cos \theta \\ v + c \sin \theta \end{pmatrix}, \quad \lambda_{G2} = \begin{pmatrix} u - c \cos \theta \\ v - c \sin \theta \end{pmatrix}, \quad \lambda_S = \begin{pmatrix} u \\ v \end{pmatrix}. \quad (33)$$

The main problems with this model are that it has more than three waves and so introduces unnecessary numerical dissipation, and the dependence

of the propagation directions on solution gradients hinders convergence to the steady state.

4.2. RUDGYARD'S WAVE MODELS

Rudgyard (Rudgyard, 1993) based his wave models on the idea of obtaining the six waves by choosing two, in principle, arbitrary propagation angles, θ_1 and θ_2 , and performing a decomposition of the gradient of the form

$$\nabla \mathbf{V} = \sum_{k=1}^3 \alpha_{\theta_1}^k \mathbf{r}_{\theta_1}^k \mathbf{n}_{\theta_1} + \sum_{k=1}^3 \alpha_{\theta_2}^k \mathbf{r}_{\theta_2}^k \mathbf{n}_{\theta_2} \quad (34)$$

which contains six free parameters, the six coefficients α_θ^k . The vectors $\mathbf{n}_\theta = (\cos \theta, \sin \theta)$ are again unit vectors in the direction θ , and \mathbf{r}_θ^k are the right eigenvectors of the matrix \mathbf{M}^* , as defined in (24) (a full set of eigenvectors is used in the decomposition for each of θ_1 and θ_2). In order to solve for the unknowns, use is also made of the left eigenvectors of that matrix

$$\mathbf{l}_\theta^{G1} = \begin{pmatrix} \frac{1}{2} \\ \frac{c}{2g} \cos \theta \\ \frac{c}{2g} \sin \theta \end{pmatrix}, \quad \mathbf{l}_\theta^{G2} = \begin{pmatrix} \frac{1}{2} \\ -\frac{c}{2g} \cos \theta \\ -\frac{c}{2g} \sin \theta \end{pmatrix}, \quad \mathbf{l}_\theta^S = \begin{pmatrix} 0 \\ -\sin \theta \\ \cos \theta \end{pmatrix} \quad (35)$$

and of the unit vector normal to \mathbf{n}_θ , $\mathbf{s}_\theta = (-\sin \theta, \cos \theta)$, leading to

$$\alpha_{\theta_1}^k = -\frac{\mathbf{s}_{\theta_2} \cdot (\mathbf{l}_{\theta_1}^k \cdot \nabla \mathbf{V})}{\sin(\theta_2 - \theta_1)}, \quad \alpha_{\theta_2}^k = \frac{\mathbf{s}_{\theta_1} \cdot (\mathbf{l}_{\theta_2}^k \cdot \nabla \mathbf{V})}{\sin(\theta_2 - \theta_1)}. \quad (36)$$

In this case the associated advection velocities in (28) are chosen so that, from (36), $\nabla w^k = \mathbf{l}_\theta^k \cdot \nabla \mathbf{V}$.

The best of the options proposed for Rudgyard's wave models is the choice of angles which satisfy the equation $u \cdot \mathbf{n}_\theta - c = 0$, that is, those angles that make the velocity of one of the gravity waves vanish. They do not depend on solution gradients and can be expressed as

$$\begin{aligned} \theta_1 &= \arctan \left(\frac{v + u\sqrt{F^2 - 1}}{u - v\sqrt{F^2 - 1}} \right) \\ \theta_2 &= \arctan \left(\frac{v - u\sqrt{F^2 - 1}}{u + v\sqrt{F^2 - 1}} \right), \end{aligned} \quad (37)$$

with $F^2 = \frac{u^2 + v^2}{c^2}$ representing the Froude number (or the Mach number in the gas dynamics problem). This technique gives very good results for

supercritical flows but is not directly applicable to the subcritical case. It can nevertheless be adapted for subcritical flows by replacing $F^2 - 1$ with $\max(F^2 - 1, \epsilon)$. The tolerance ϵ typically takes a value of 0.01. Results become increasingly poor as the Froude number decreases and the effect of having more than three waves becomes more significant.

5. Approximate Diagonalizations

In 1D shallow water flows it is possible to diagonalize the flux Jacobian, thus splitting the problem into independent scalar subproblems. Unfortunately, in 2D the matrices \mathbf{A} and \mathbf{B} cannot generally be diagonalized simultaneously (hence the difficulty in constructing suitable wave models). Instead, an approximate diagonalization can be constructed via a 3-parameter similarity transformation (Deconinck et al., 1986), giving a system in ‘characteristic’ variables \mathbf{W} ,

$$\mathbf{W}_t + \mathbf{A}_W \mathbf{W}_x + \mathbf{B}_W \mathbf{W}_y = \mathbf{0}, \quad (38)$$

in which the 3 free parameters are chosen so that the new Jacobians \mathbf{A}_W and \mathbf{B}_W are, in some sense, close to being diagonal. This is treated as a decoupled set of inhomogeneous equations, each with a residual of the form

$$R_T = \boldsymbol{\lambda} \cdot \nabla w + q, \quad (39)$$

(the distribution coefficients can be calculated as they would for the homogeneous fluctuation but then used to distribute the complete R_T) and a conservative flux balance,

$$\mathbf{R}_T = \sum_{k=1}^{N_{eq}} (\boldsymbol{\lambda}^k \cdot \nabla w^k + q^k) \mathbf{r}_c^k, \quad (40)$$

in which \mathbf{r}_c^k are the columns of the similarity transformation matrix $\frac{\partial \mathbf{U}}{\partial \mathbf{W}}$, and $N_{eq} = 3$ is the number of equations in the system.

These methods have an advantage over the existing simple wave models in having the correct number of components for linearity preservation ($N_w = N_{eq}$) but the propagation directions, which depend on the parameters which define the similarity transformation, are usually chosen to depend on solution gradients, creating problems with convergence to a steady state. However, their main disadvantage is the presence of the source terms q^k which destroy positivity and hence robustness.

The effect of the source terms created by the characteristic decomposition can be minimised by attempting to diagonalize a preconditioned form of the shallow water equations (Pailiere et al., 1995; Mesaros and Roe,

1995). The decomposed flux balance once more takes the form (40), but now \mathbf{r}_c^k are the columns of the matrix $\frac{\partial \mathbf{U}}{\partial \mathbf{Q}} \mathbf{P}^{-1} \frac{\partial \mathbf{Q}}{\partial \mathbf{W}}$, \mathbf{Q} being an intermediate set of (symmetrizing) variables, introduced to simplify the algebra, and \mathbf{P} a preconditioning matrix. Careful choice of the preconditioner gives an optimal decoupling of the system, complete in supercritical flow but unavoidably including a coupled 2×2 elliptic subsystem for subcritical flow.

Briefly, these most recent decompositions are constructed by first transforming the shallow water equations into symmetrizing variables,

$$\partial \mathbf{Q} = \begin{pmatrix} \sqrt{\frac{q}{d}} \partial d \\ \partial q \\ q \partial \theta \end{pmatrix}, \quad (41)$$

where $q = \sqrt{u^2 + v^2}$ is the flow speed and $\theta = \tan^{-1}(\frac{v}{u})$ is the direction of the flow. The system therefore becomes

$$\mathbf{Q}_t + \mathbf{A}_Q \mathbf{Q}_x + \mathbf{B}_Q \mathbf{Q}_y = \mathbf{0}, \quad (42)$$

in which the flux Jacobians are symmetric matrices given by

$$\mathbf{A}_Q = \frac{\partial \mathbf{Q}}{\partial \mathbf{U}} \mathbf{A} \frac{\partial \mathbf{U}}{\partial \mathbf{Q}}, \quad \mathbf{B}_Q = \frac{\partial \mathbf{Q}}{\partial \mathbf{U}} \mathbf{B} \frac{\partial \mathbf{U}}{\partial \mathbf{Q}}. \quad (43)$$

The equations (42) are simplified even further when they are written in terms of the streamwise coordinates, ξ and η , which leads to

$$\mathbf{Q}_t + \mathbf{A}_Q^S \mathbf{Q}_\xi + \mathbf{B}_Q^S \mathbf{Q}_\eta = \mathbf{0}, \quad (44)$$

where

$$\mathbf{A}_Q^S = \frac{u \mathbf{A}_Q + v \mathbf{B}_Q}{q}, \quad \mathbf{B}_Q^S = \frac{-v \mathbf{A}_Q + u \mathbf{B}_Q}{q}. \quad (45)$$

These equations are now preconditioned by an appropriate matrix \mathbf{P} , giving

$$\mathbf{Q}_t + \mathbf{P} (\mathbf{A}_Q^S \mathbf{Q}_\xi + \mathbf{B}_Q^S \mathbf{Q}_\eta) = \mathbf{0}, \quad (46)$$

and this system is transformed into “characteristic” variables,

$$\mathbf{W}_t + \mathbf{A}_W^S \mathbf{W}_\xi + \mathbf{B}_W^S \mathbf{W}_\eta = \mathbf{0}, \quad (47)$$

which can be treated as (38), but where

$$\mathbf{A}_W^S = \frac{\partial \mathbf{W}}{\partial \mathbf{Q}} \mathbf{P} \mathbf{A}_Q^S \frac{\partial \mathbf{Q}}{\partial \mathbf{W}}, \quad \mathbf{B}_W^S = \frac{\partial \mathbf{W}}{\partial \mathbf{Q}} \mathbf{P} \mathbf{B}_Q^S \frac{\partial \mathbf{Q}}{\partial \mathbf{W}}. \quad (48)$$

For an appropriate choice of \mathbf{P} the system (47) is either fully or partially diagonalized depending on whether the flow is supercritical or subcritical. The hyperbolic components are treated using the standard scalar schemes, but the subcritical elliptic subsystem can usefully be distributed in a different manner: a system Lax-Wendroff scheme has been shown to work well with this component (Mesaros and Roe, 1995).

An example of a suitable preconditioner, based on that of Mesaros and Roe (Mesaros and Roe, 1995), is given by (Hubbard and Baines, 1997)

$$\mathbf{P} = \frac{1}{q} \begin{pmatrix} \frac{\varepsilon F^2}{\beta \kappa} & -\frac{\varepsilon F}{\beta \kappa} & 0 \\ -\frac{\varepsilon F}{\beta \kappa} & \frac{\varepsilon}{\beta \kappa} + \varepsilon & 0 \\ 0 & 0 & \frac{\beta}{\kappa} \end{pmatrix}, \quad (49)$$

where

$$\beta = \sqrt{|F^2 - 1|}, \quad \kappa = \max(F, 1), \quad (50)$$

and $\varepsilon = \varepsilon(F)$ is a function which satisfies $\varepsilon(0) = \frac{1}{2}$ and $\varepsilon(F) = 1$ for $F \geq 1$ (giving the correct behaviour in the preconditioned system at stagnation and continuity of the optimal decomposition through the critical point).

These preconditioned wave models have proved to be the best of the current decompositions for the modelling of steady state problems but, in contrast to the simple wave models, seem unlikely to provide a simple extension to unsteady problems (not least because they have a singularity at stagnation points which can only be treated in the steady case).

6. Source terms

The modelling of shallow water flows commonly requires the inclusion of source terms, and hence the approximation of inhomogeneous equations of the form

$$\frac{\partial \mathbf{U}}{\partial t} + \nabla \cdot \mathbf{F} = \mathbf{S}. \quad (51)$$

For example, taking

$$\mathbf{S} = (0, -ghz_x, -ghz_y)^T \quad (52)$$

allows the modelling of flow over a varying bed topography, defined by the gradient of the bed height, ∇z .

In principle it is simple to incorporate source terms within the structure of multidimensional upwind schemes by decomposing them in the same manner as the flux terms (Hubbard and Baines, 1997). For the approximate diagonalization approach this results in a slightly different residual to (40), namely

$$\mathbf{R}_T = \sum_{k=1}^{N_{eq}} (\lambda^k \cdot \nabla w^k + q^k - s_W^k) \mathbf{r}_c^k, \quad (53)$$

where s_W^k are the components of $\mathbf{S}_W = \frac{\partial \mathbf{W}}{\partial \mathbf{U}} \mathbf{S}$ (or, when a preconditioner is included, $\mathbf{S}_W = \frac{\partial \mathbf{W}}{\partial \mathbf{Q}} \mathbf{P} \frac{\partial \mathbf{Q}}{\partial \mathbf{U}} \mathbf{S}$). Each component of (53) is distributed using the coefficients calculated for the homogeneous equations. When simple wave models are used the correct treatment is not so obvious, since the components of the decomposition are not independent, and this represents a subject for further research.

7. Numerical results

7.1 Test1: Steady flow

The first test case presented here is that of flow through a symmetric constricted open channel of length 4, whose breadth is given by

$$B(x) = \begin{cases} 1 - (1 - B_{\min}) \cos^2(\pi(x - 2)) & |x - 2| \leq .5 \\ 1 & \text{otherwise,} \end{cases} \quad (54)$$

where $B_{\min} = 0.92$ is the minimum channel breadth and x is the distance into the channel (so the throat is positioned at the midpoint of the constriction, $x = 2.0$). The 2114 node, 4054 cell grid on which the numerical results have been obtained is shown in Fig. 1 along with three steady state solutions distinguished by their freestream Froude numbers: (i) $F_\infty = 0.5$, completely subcritical and hence symmetric about the throat of the channel, (ii) $F_\infty = 0.71$, transcritical with a stationary hydraulic jump in the constriction downstream of the throat, and (iii) $F_\infty = 2.0$, completely supercritical, with a criss-cross pattern of undular jumps downstream of the throat. Simple characteristic boundary conditions are applied in each case. The solutions have been obtained using the hyperbolic/elliptic decomposition described in Section 5, applying the PSI scheme to each of the decoupled scalar components and a system Lax-Wendroff scheme to the subcritical elliptic subsystem. The results illustrate that the scheme can accurately model each of these different types of steady state flow.

7.2 Test2: L-shaped channel

Results from experimental test cases are presented, proposed by Prof. Zech (Civil Engineering Dept., UCL Belgium) from the Working Group on Dam Break Flow Modelling in which the authors are involved. Experimental and numerical results will be compared in these two time-dependent test cases. The treatment of the solution at the boundaries has been kept as close as possible to the theory of the characteristics in 2D. In all cases, the number of physical conditions to be imposed has been determined by this theory.

The flow domain, depicted in Fig. 2, consists of a square reservoir that initially contains a wall to separate it from the L-shaped channel. The initial conditions are zero flow with 0.2m depth to the left and 0.01m depth to the

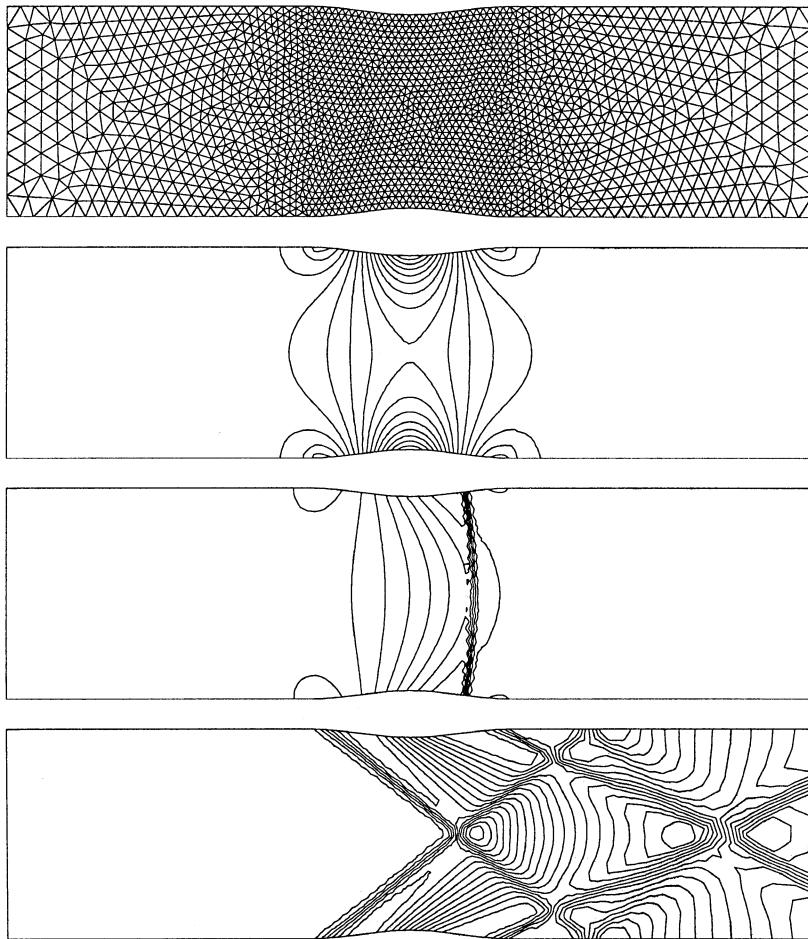


Figure 1. The grid and contours of depth for solutions of the subcritical (top), transcritical (middle) and supercritical (bottom) symmetric constricted open channel test cases.

right of the wall. All boundaries are solid non-slip walls except the outlet which is considered free. The Manning coefficient is 0.0095 and the bed slope is zero. The number of elements used in the mesh is 2954. Comparisons of the time evolution of the water depth predicted by experiment and numerics once the wall is removed using Rudgyard's wave model and the PSI scheme are made at the points P1, P2, P3, P4, P5 and P6, and are shown in Fig.3. They indicate that the multidimensional upwind schemes do provide an accurate numerical model of the flow.

7.3 Test3: Comparison with upwind finite volume scheme results

Results will now be presented which are obtained with first order upwind

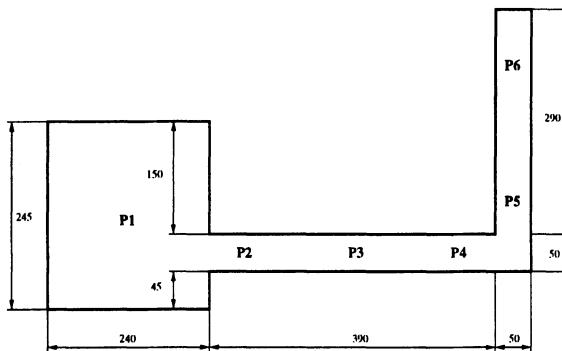


Figure 2. Geometry for the L-shaped channel test.

finite volume and the multidimensional upwind (same as before) approximations on the same unstructured Delaunay triangular mesh for a second experimental test case, also proposed by Prof. Zech.

The test to be studied combines a square shaped upstream reservoir and a channel with a 45° bend (see Fig. 4). The channel is made of 4.25m and 4.15m long and 0.495m wide rectilinear reaches connected at a 45° angle by an element. There is no slope in the channel. A gate (which is opened at $t = 0$) connects the channel to a reservoir. The initial conditions are water at rest with the free surface 25cm above the bed level in the upstream reservoir and 1cm water depth in the channel. All boundaries are solid walls except the outlet which is considered free. The Manning coefficient is $n_b = 0.0095$ for the bed and $n_w = 0.0195$ for the walls. The number of elements used in the mesh is 15397.

The flow will be essentially two-dimensional in the reservoir and at the angle between the two straight reaches of the channel. Two features of the resulting dam break flow are of special interest: the damping effect of the corner, and the upstream moving hydraulic jump which is formed by reflection at the corner.

Nine gauging points were used in the laboratory to measure water level in time. Their locations are shown in Fig. 4. The measurements at these stations are compared with the numerical results and displayed in Figs. 5 and 6. Fig. 7 shows snapshots of the free surface at time 18s.

In general, the figures indicate good performance of both numerical schemes. The arrival time of the main shock fronts is better captured by the standard upwind method. Some differences are noticeable in P2, P3 and P4 in terms of the reflected shock front celerity, which may be attributed to the treatment of the boundary conditions. However, the great improvement shown by the multidimensional upwinding is only really visible in the free

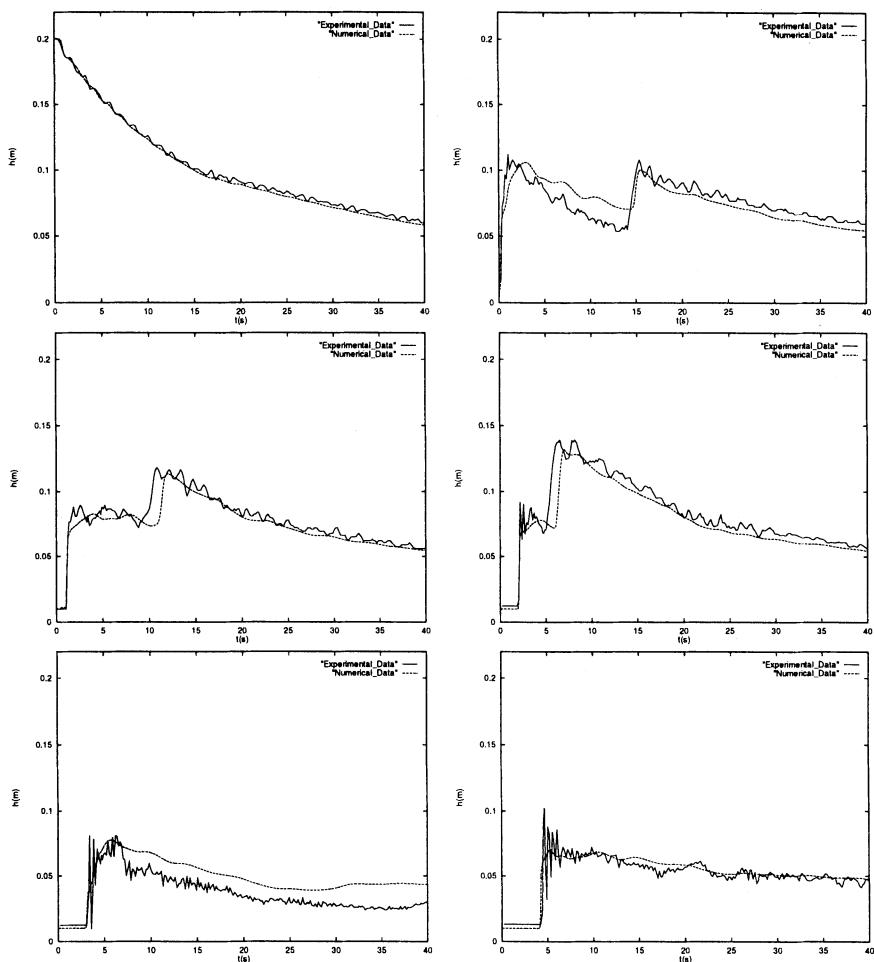


Figure 3. Time evolution of the depth of water at points P1 to P6 for Test1.

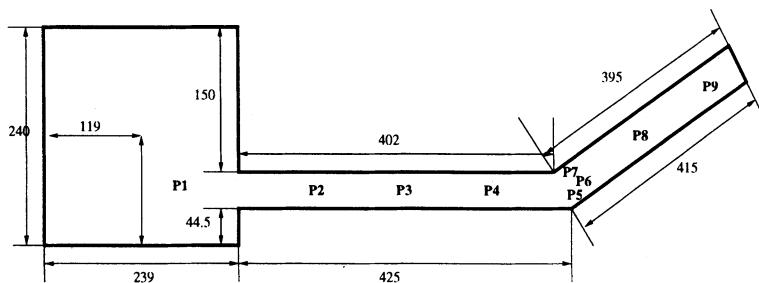


Figure 4. Plane view of the Test2 channel with gauging points.

surface plots, in which it is clear that it models the shock structure far better. This indicates that perhaps measurements along the walls of the channel should be taken into account to demonstrate which approximation is better. Up to now, only data in the central axis have been measured.

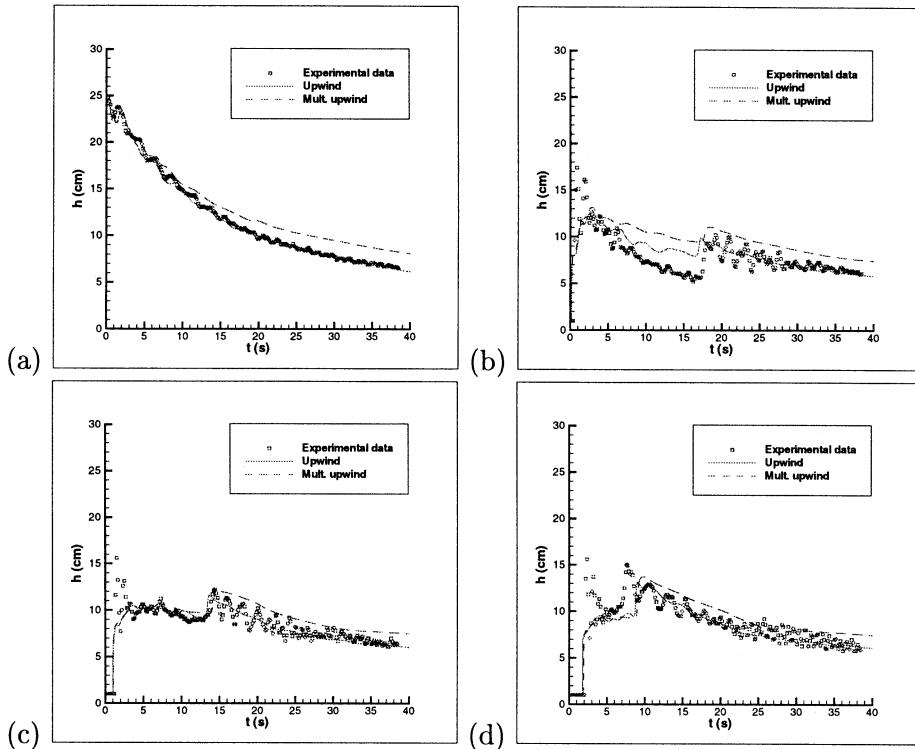


Figure 5. Water depth history at points a) P1, b) P2, c) P3 and d) P4 for Test2.

8. Conclusions

Two-dimensional wave decomposition and multidimensional upwinding seem a promising method of solution for the 2D shallow water equations. A number of schemes have been adapted to render the technique suited to hydraulic problems with shocks. As with the 1D TVD schemes, our experience with using the multidimensional upwind approach for the shallow water equations has closely followed that of the researchers solving the Euler equations (with both the advection schemes and the wave models), showing the same properties as for that system of equations.

The procedure is more complicated and costly than the most efficient present day generalizations of 1D upwind finite volume techniques. How-

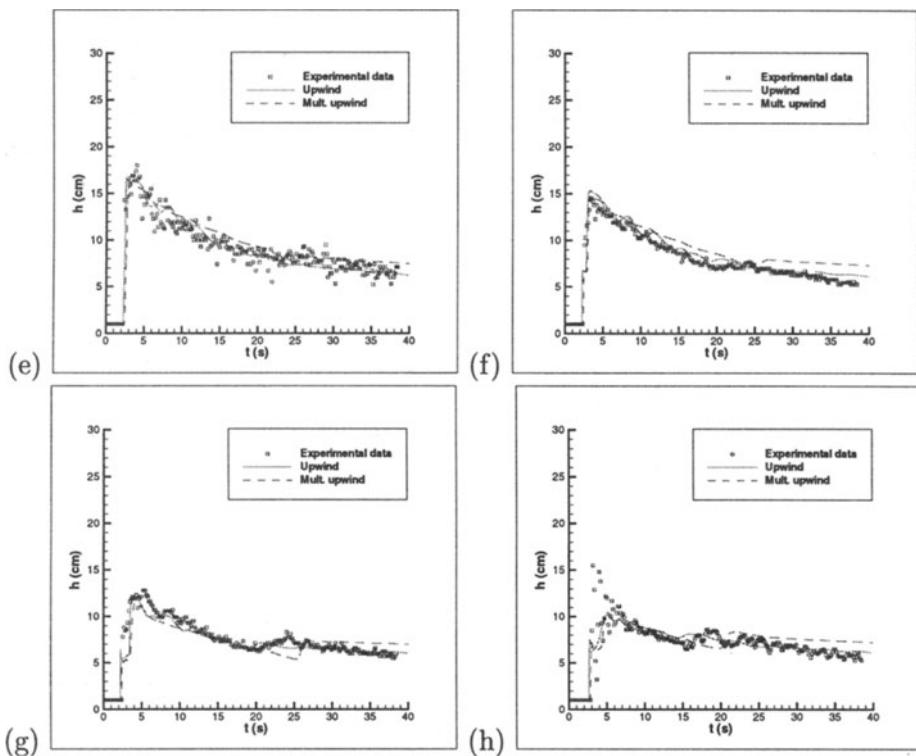


Figure 6. Water depth history at points e) P5, f) P6, g) P7 and h) P8 for Test2.

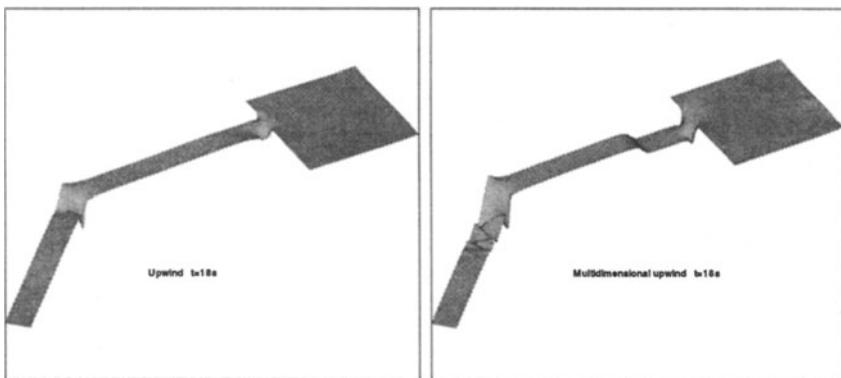


Figure 7. Free surface at time $t=18s$ for Test2 with upwind finite volume (left) and multidimensional upwind schemes (right).

ever, it is based on a triangular discretization and, by taking advantage of the triangles, the disadvantages can be overcome, making the schemes very competitive. The future for these schemes then looks much more promising, since they can clearly be applied to arbitrary geometries, a great advantage for hydraulic engineers working on practical problems, particularly as there is a wide variety of possibilities concerning grid movement and adaptation (Deconinck et al., 1997; ?).

Future work is envisaged to find better ways to deal with the source terms present in the shallow water equations when applied to realistic problems, while there is still much work to be done in the development of more efficient and accurate schemes in unsteady cases, possibly following recent work on the scalar schemes (Hubbard and Roe, 1999; März, 1996) but definitely requiring a more detailed study of possible wave models for time-dependent problems.

References

- M.J. Baines and M.E. Hubbard. Multidimensional upwinding with grid adaptation, *Numerical Methods for Wave Propagation*, E.F.Toro and J.F.Clarke (Eds.), pp. 33–54, Kluwer Academic Publishers (1998).
- S.F. Davis. A rotationally biased upwind difference scheme for the Euler equations., *J. Comput. Phys.*, **56** (1984).
- H. Deconinck, C. Hirsch, J. Peuteman. Characteristics decomposition methods for the multidimensional Euler equations, 10th Int. Conf. in Num. Met. in Fluid Dyn. 216–221 (1986).
- H. Deconinck, P.L. Roe and R. Struijs, A multi-dimensional generalization of Roe's flux difference splitter for the Euler equations, *Computers and Fluids*, **22**, 215–222 (1993).
- H. Deconinck, R. Struijs, G. Bourgois, H. Paillere, P.L. Roe, Multidimensional upwind methods for unstructured grids, Unstructured grid methods for advection dominated flows, AGARD787 (1992).
- H. Deconinck, R. Struijs, G. Bourgois and P.L. Roe, High resolution shock capturing cell vertex advection schemes for unstructured grids, in VKI LS 1994-05, Computational Fluid Dynamics (1994).
- H. Deconinck, Analysis of wave propagation properties for the Euler equations in two-space dimensions, in VKI LS 1994-05, Computational Fluid Dynamics (1994).
- H. Deconinck, B. Koren (editors), Euler and Navier-Stokes solvers using multidimensional upwind schemes and multigrid acceleration, *Notes on Numerical Fluid Mechanics*, **57**, Vieweg (1997).
- P. Garcia-Navarro, M.E. Hubbard and A. Priestley, Genuinely Multidimensional Upwinding for the 2D Shallow Water Equations, *Journal of Computational Physics*, **121**, 79-93 (1995).
- M.E. Hubbard, Aspects of multidimensional upwinding: time-dependent nonlinear systems, source terms, spherical geometries, and three-dimensional grid adaptation, Report NA-4/99, Dept. of Math., Univ. of Reading, UK (1999).
- M.E. Hubbard and M.J. Baines, Conservative multidimensional upwinding for the steady two dimensional shallow water equations, *J. Comput. Phys.*, **138** 419–448 (1997).
- M.E. Hubbard and P.L. Roe, Compact high-resolution algorithms for time-dependent advection on unstructured grids, to appear, *Int. J. for Num. Methods in Fluids* (1999).
- R.J. LeVeque, *Numerical methods for conservation laws*, Birkhauser, Basel, 2nd edition (1992).

- D. Levy, K.G. Powell, B. Van Leer, Implementation of a grid-independent upwind scheme for the Euler equations, 91-0635, *AIAA* (1991).
- J. März, Improving time accuracy for residual distribution schemes, VKI PR 1996-17, von Karman Institute for Fluid Dynamics (1996).
- L.M. Mesaros and P.L. Roe, Multidimensional fluctuation splitting schemes based on decomposition methods, *AIAA* 95-1699 (1995).
- H. Paillere, E. van der Weide and H. Deconinck, Multidimensional upwind methods for inviscid and viscous compressible flows, in VKI LS 1995-02, Computational Fluid Dynamics (1995).
- I.M. Parpia and D.J. Michalek, Grid-independent upwind scheme for multidimensional flow, *AIAA*, 31(4): 646-651 (1993).
- P.L. Roe, Discrete models for the numerical analysis of time-dependent multidimensional gas dynamics, *J. Comput. Phys.*, **63**, pp 458-476 (1986).
- P.L. Roe, A basis for upwind differencing of the two-dimensional unsteady Euler equations, *Num. Met. for Fluid Dyn. II*, pp 55-80 (1986).
- M.A. Rudgyard, Multidimensional wave decompositions for the Euler equations, VKI Lecture notes (1993).
- C.L. Rumsey, B. Van Leer, P.L. Roe, A grid-independent approximate Riemann solver with applications to the Euler and Navier-Stokes equations, 91-1530, *AIAA* (1991).
- D. Sidilkover and P.L. Roe, Unification of some advection schemes in two dimensions, ICASE Report 95-10 (1995).
- Y. Tamura and K. Fujii, A multidimensional upwind scheme for the Euler equations on unstructured grids, *4th ISCFD Conference* (1991).

HELMIT - A NEW INTERFACE RECONSTRUCTION ALGORITHM

R.D. GIDDINGS

*Atomic Weapons Establishment,
Aldermaston, Reading, RG7 4PR, U.K.
Email: Rob.Giddings@awe.co.uk*

Abstract. A new interface reconstruction algorithm is presented which represents the interface within each cell as two connected line segments. This enables interfaces to be continuous across cell boundaries and allows T-junctions to be represented.

1. Introduction

Practical applications of computational fluid dynamics can often contain a number of qualitatively different regions. These can represent different materials in multiphase flow or shocked/unshocked regions in gas dynamics problems, for example. Many strategies have evolved over the years to keep track of the interfaces between these regions, falling into two main categories -‘capturing’ and ‘tracking’. Interface capturing schemes, see (Woodward and Colella, 1984) for a survey, including Godunov’s, do not explicitly track interfaces, but rely instead on implicit indicators eg sharp gradients in density across shocks. This has the disadvantage that the discontinuity can be smeared across a few cells. Much of the machinery of modern schemes eg flux limiting is aimed at minimising this effect.

In contrast, interface tracking schemes maintain a record of exactly where the interfaces are, see (Hyman, 1984), (Oran and Boris, 1987), (Benson, 1992), (Puckett et al., 1997) for surveys. This introduces extra computational effort but is necessary where interface physics is important eg surface tension, radiation transport, reaction kinetics etc. For some problems tracking is used where capturing introduces unacceptable diffusion eg Rayleigh-Taylor instability, free surfaces. In (Puckett et al., 1990) a tracking scheme is coupled to Godunov’s scheme for the calculation of shock refraction between two compressible fluids of differing equation of state.

Tracking can be accomplished in two ways:

- Directly, using marker particles ie a set of points initially on the interfaces which move with the flow (Glimm, 1985). This is accurate but slow because it is necessary to check if interfaces have intersected and markers may need to be added or removed in places to maintain detail.
- Indirectly eg
 - Level-set methods - store ‘distance to the interface’ field ϕ .
 - Volume-of-fluid methods - store ‘volume fraction’ for each cell.
 - But with these the interface needs to be reconstructed each timestep.

The level-set method (Sethian, 1999) has many advantages (simplicity, ease of extension to higher dimensions etc) but it has the disadvantage that the numerical implementation is not inherently conservative (p 142).

The volume-of-fluid (VOF) method (Noh and Woodward, 1976) (Hirt and Nichols, 1981) is inherently conservative but has the disadvantage that current algorithms represent the interface within each cell as a single straight line which results in jagged edges when these do not meet across cell boundaries. If many materials are present in a cell then they are grouped together in a certain order and the algorithm repeated each time. T-junctions (where three materials meet at a point) cannot be represented and can result in successive interfaces crossing.

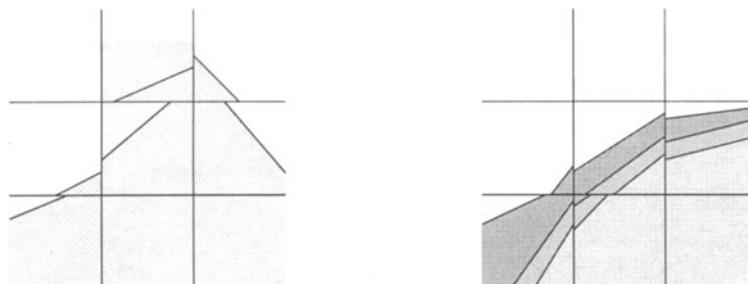


Figure 1. Volume-of-fluid reconstruction of one (left) and many (right) materials

2. HELMIT

HELMIT is an attempt to remedy these problems within the VOF framework. The algorithm consists of three basic steps:

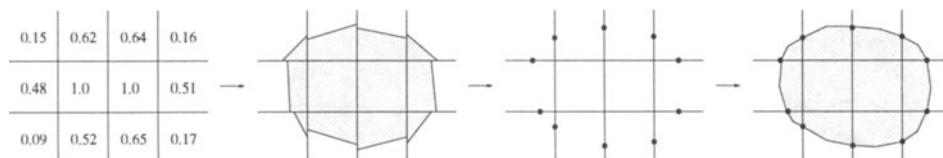


Figure 2. Main steps of the HELMIT algorithm

Step one: Use an existing technique eg (Youngs, 1982)

Step two: Where interfaces do not meet across a cell boundary, they will both be forced to go through a common point, called a Mesh Crossing Point or MCP, placed halfway between them. For the top centre cells in figure 2, this is shown in figure 3. This is repeated for all interfaces.

Step Three: Now the MCPs have been placed all that is left is to join these up enclosing the correct volume fractions. If this is naively done with straight lines it is unlikely the areas will be correct so they are adjusted with isosceles triangles (figure 4).

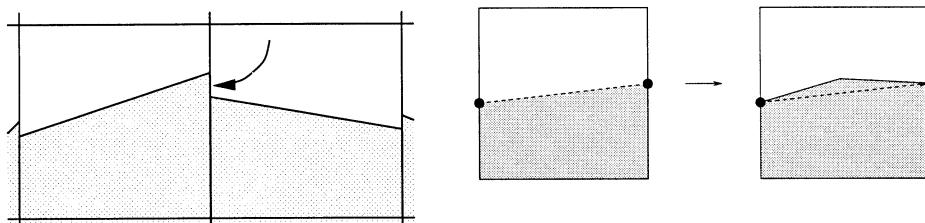


Figure 4. Hinged Line interface

Figure 3. Location of the MCP

Each interface now consists of two line segments joined or ‘hinged’ at the apex of the triangle. Thus the method is called the Hinged Line Method of Interface Tracking or HELMIT for short.

Exceptional cases

1. It is possible for the hinge to be placed outside the cell. In this case it is retracted along the perpendicular bisector of the MCPs onto the cell boundary. Then the MCPs are moved to correct the area (figure 5).

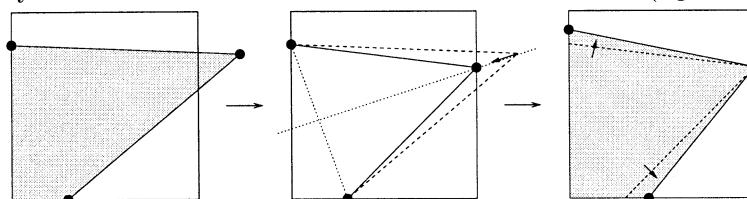


Figure 5. Exceptional case 1 - Hinge placed outside cell

Step two assumes that it is always possible to find pairs of interfaces by which to locate the MCP. Often this is not true. An interface might have no partner (figure 6 left) or many from which a choice needs to be made (figure 6 right).

A new strategy is required. The example on the left suggests that the interfaces themselves are not as important as the body of material they enclose. For the example on the right, the body of material relevant to the top centre cell is magnified in figure 7 left. The gap in the body of material just outside the cell boundary is due to a poorly placed interface. Removing

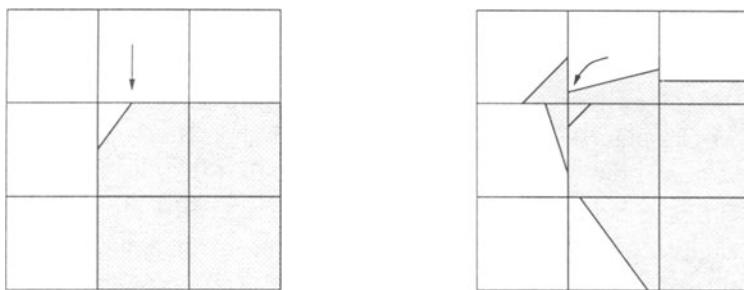


Figure 6. Exceptional case 2 - no suitable partner

this gap (figure 7 right) leaves two suitable pairs of interfaces by which to place the MCPs. In practice an array is used to record which parts of the exterior of the cell boundary are in contact with material (call these 'black') and which parts are not (call these 'white'). Then the colour of the shortest segment (black or white) is inverted, and this is repeated until just one black segment remains.

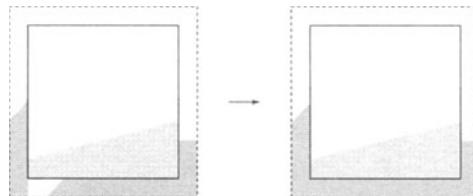


Figure 7. Removing the gaps

If originally there were two suitable interfaces this procedure would give the same result as before.

3. Results (Two Materials)

In the following examples the results of Youngs scheme and HELMIT can be compared with the original polygon.

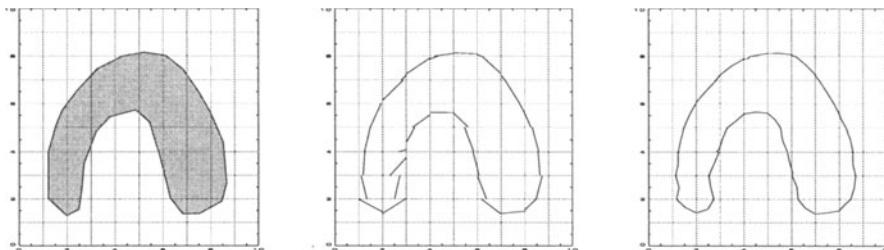


Figure 8. Original polygons

Youngs' interfaces

HELMIT

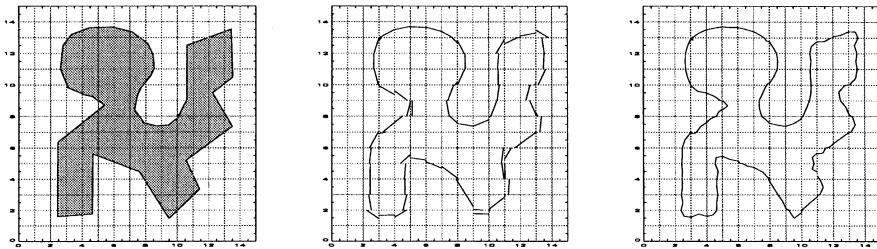


Figure 9. Original polygons Youngs' interfaces HELMIT

The HELMIT interfaces are continuous almost everywhere. Notice that where the original polygon is smooth, the Youngs interfaces are very good and HELMIT can do little to improve them, but at sharp corners, where the Youngs interfaces become jagged, HELMIT has corrected this, albeit sometimes at the expense of making the interface slightly 'wiggly'. This problem will be solved later.

4. T-junctions

The benefit of changing the interface from a single line to a hinged line is that now T-junctions can be represented. Suppose a cell contains three materials and three MCPs have been found on the boundary. Between each pair an isosceles triangle can be constructed to enclose the correct volume fraction. However the apex of the triangle, the hinge, can lie anywhere on a line parallel to the base and still satisfy the area requirement. Where these lines meet is the T-junction :

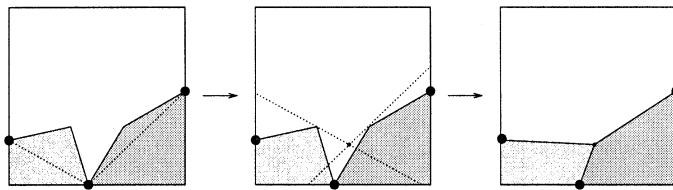


Figure 10. Construction of the T-junction

In practice a T-junction is fitted whenever two successive HELMIT interfaces intersect, two of the three points coming from the most recent interface and one from the previous.

5. Results (Many Materials)

In figure 11, middle row, cells (8,11) and (8,13) contain a normal interface in addition to a T-junction. This example also contains a thin shell around the dumb-bell shape in the middle which also causes no problems. In the

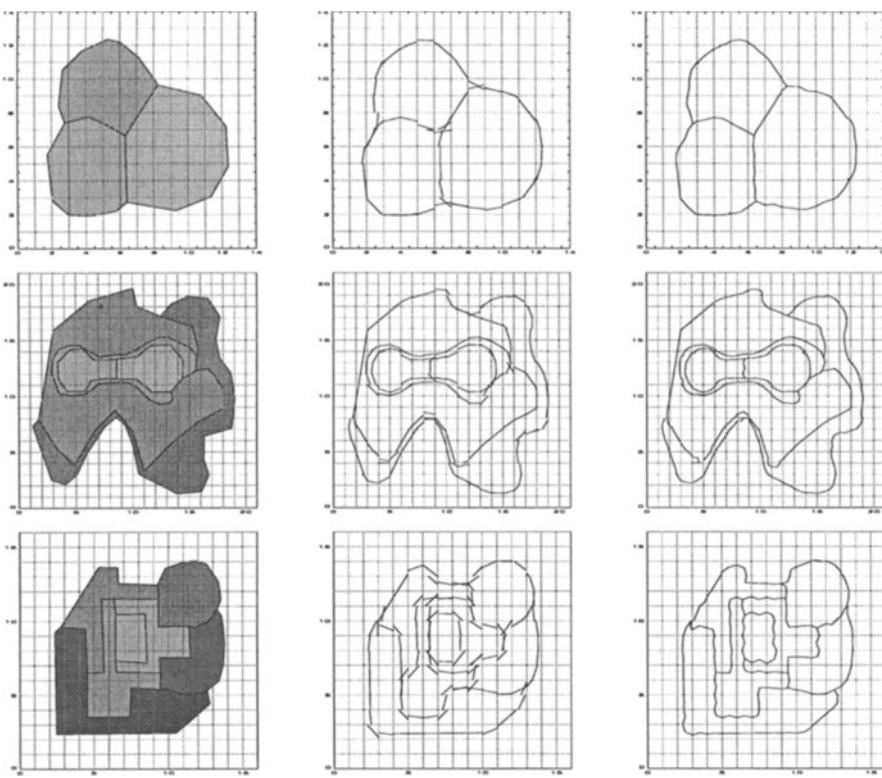


Figure 11. Original polygons Youngs' interfaces HELMIT
bottom row, HELMIT has successfully maintained a continuous interface
through two neighbouring T-junctions in cells (6,10) and (6,11).

6. Half-HELMIT

HELMIT has a drawback when it comes to be implemented in software. It needs to have available all the Youngs interfaces. So these either need to be stored, or recalculated each time they need to be used (which can be up to five times). Half-HELMIT avoids this by reusing information calculated in the original Youngs step. The first part of Youngs algorithm takes each bordering cell in turn and fits a single straight line through that cell and the centre cell, satisfying both volume fractions (figure 12). It then extracts the length of common border containing material and throws everything else away. Half-HELMIT uses these interfaces instead of the full Youngs interfaces in the surrounding cells (figure 13).

Previously the MCP was placed halfway between pairs of interfaces to ensure the HELMIT interfaces joined up. But now there is only one proper interface - in the centre cell - so this is no longer possible. With no extra

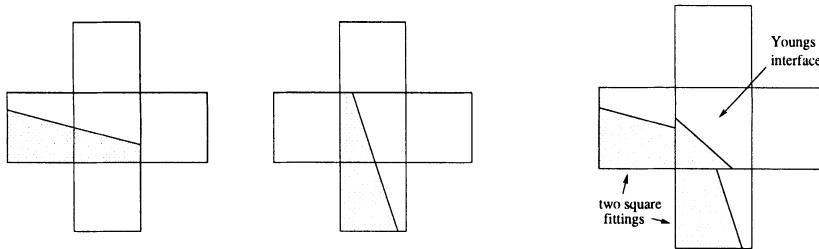
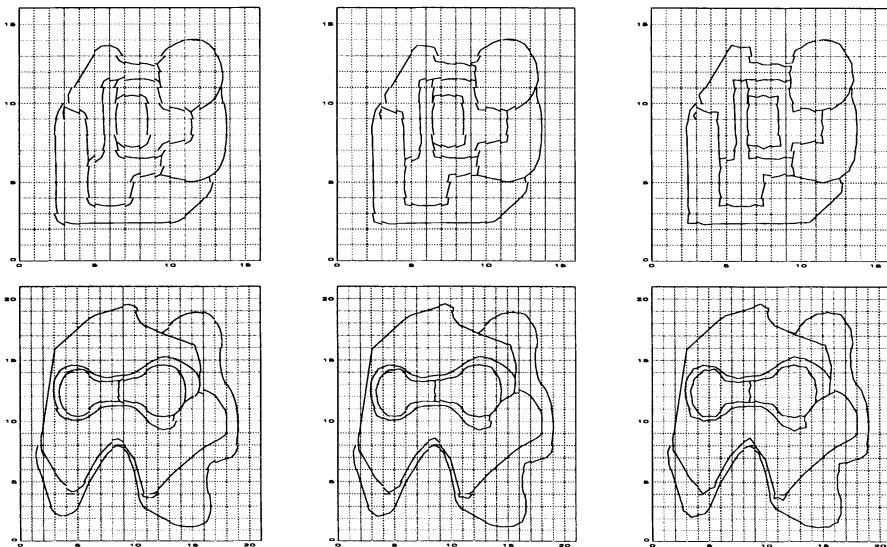


Figure 12. Two square fittings

Figure 13. Half-HELMIT

information to decide where, the MCP could be placed anywhere between the interfaces. Suppose it is placed a fixed fraction λ , $0 \leq \lambda \leq 1$ of the distance from the Youngs interface to the two square fitting. Figure 14 shows the results for three values of λ .

 $\lambda = 0.3$ $\lambda = 0.5$ $\lambda = 0.7$

In both these examples, the higher λ , the sharper the reconstructed interface. In the top row, where there are many sharp corners, $\lambda = 0.5$ looks best, but in the bottom row, where there are few sharp corners $\lambda = 0.3$ seems best. This is especially obvious around the dumb-bell shape in the middle. After experimentation the author found 0.5 to be reasonable in most cases.

7. Double HELMIT

Half-HELMIT improves Youngs interfaces but does not make them join up. So it is possible to apply the original HELMIT algorithm again, putting the MCPs halfway between the Half-HELMIT interfaces (figure 15).

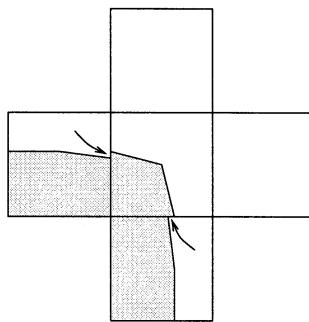


Figure 15. Double-HELMIT

This combination is called Double-HELMIT. The results are shown in figure 16.

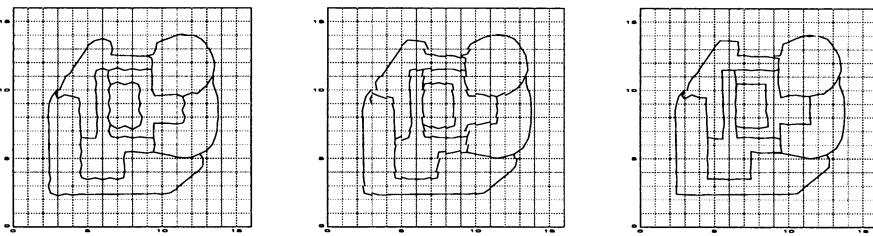


Figure 16. HELMIT

Half-HELMIT

Double-HELMIT

The 'wiggles', present in the HELMIT interfaces, have been removed.

8. Advection Test

Here Half-HELMIT is compared to Youngs scheme for the advection of a 'jack'. This is a cross shape, with each arm being four cells wide and six cells long. Figure 17 shows the jack after being advected diagonally up and to the right by about sixty cells.

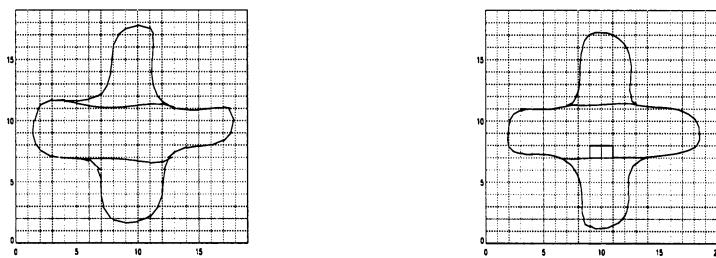


Figure 17. Advection test

Both jacks have lost their sharp corners. However, the internal interfaces clearly show that Youngs scheme has caused material to get slightly left

behind, gathering in the bottom and left arms. With the improved interfaces of the Half-HELMIT scheme, this has not happened.

9. Quantitative Tests

In this section a quantitative comparison of the four schemes (Youngs, Half-HELMIT, HELMIT, Double-HELMIT) will be made using two test shapes - a circle and a pentagon. If T is the test shape and P the polygon(s) reconstructed by a scheme then two error estimators are used:

1. Area discrepancy - defined to be $1 - \text{Area}(T \cap P)/\text{Area}(T)$
2. Length discrepancy - defined to be $\text{Length}(P)/\text{Length}(T) - 1$

To try to remove mesh effects, for each mesh size h the test was repeated 50 times with differing positions for the circle centre and rotations of the pentagon, and the average errors taken (figures 18 and 19). The noisy curves indicate this has not entirely succeeded. As expected, the more sophisticated schemes perform better, with Double-HELMIT the best for both errors. For the circle the area errors are all scaling better than second order, but fall to second order for the pentagon.

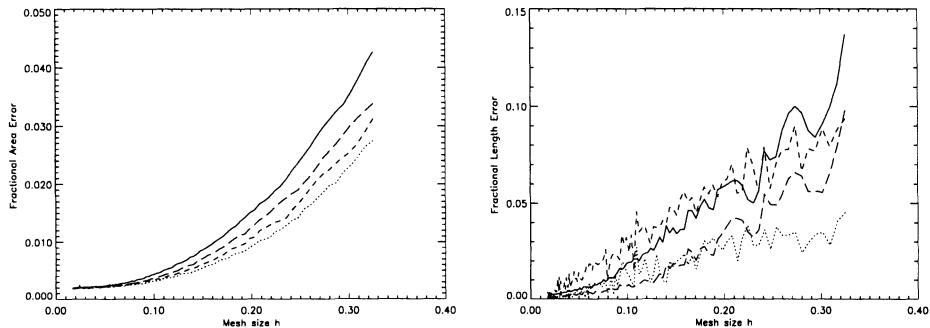


Figure 18. Circle Test. Left - Area discrepancy, Right - Length Discrepancy

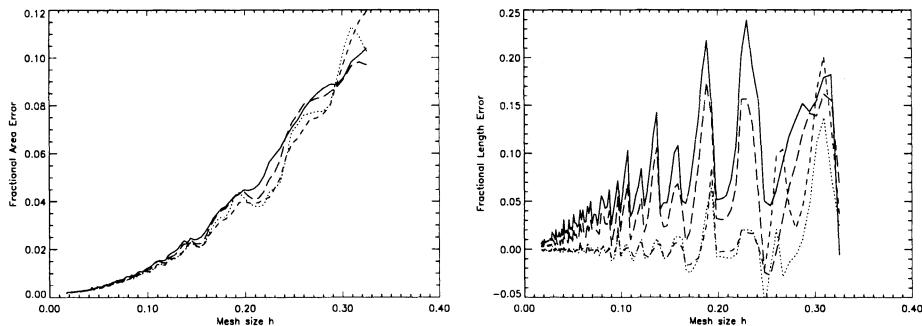


Figure 19. Pentagon Test. Left - Area discrepancy, Right - Length Discrepancy

Key: Solid line = Youngs scheme, Long Dashed Line = Half-HELMIT, Short Dashed Line = HELMIT, Dotted Line = Double-HELMIT

For the pentagon, the length error accentuates the differences between the schemes. For Youngs and Half-HELMIT the length error falls approximately linearly with mesh size. But for those schemes where the interfaces join up, HELMIT and Double-HELMIT, the length error is roughly zero for $h < 0.18$. In places it is negative which means the reconstructed polygon has shorter length than the target pentagon - indicating the large amount of information thrown away in the conversion from the polygon to the volume fractions.

10. Summary

A new interface reconstruction scheme, HELMIT, and two extensions of it, Half-HELMIT and Double-HELMIT, have been presented. By representing the interface within each cell as a ‘hinged line’, it is possible to achieve continuity of the interface across cell boundaries as well as representing T-junctions. The schemes are shown to perform better than Youngs in a number of examples presented here. Considered as a post-processor for other schemes, it is thus extendable to non-orthogonal and non-cartesian meshes.

References

- Bell J, Marcus D L, Puckett E G, Almgren A S and Ryder W J (1997). A high order projection method for tracking fluid interfaces in variable density incompressible flows. *J. Comput. Phys.* **130**, pp 269-282.
- Benson D J (1992). Computational methods in lagrangian and eulerian hydrocodes. *Computer Methods in Applied Mechanics and Engineering*, **99**, pp 235-394.
- Glimm J and McBryan O (1985). A computational model for interfaces. *Adv. App. Math.*, **6**:422.
- Henderson L F, Puckett E G and Colella P (1990). On the Anomalous Refraction of Shock Waves. Proc., Second Japan-Soviet Union Symposium on Computational Fluid Dynamics, Tsukuba, Japan, p 144.
- Hirt C W and Nichols B D (1981). Volume of fluid (VOF) method for the dynamics of free boundaries. *J. Comput. Phys.* **39**, pp 201-225.
- Hyman J M (1984). Numerical methods for tracking interfaces. *Physica 12D*, pp 396-407.
- Noh W F and Woodward P (1976). SLIC (simple line interface calculation). Lecture Notes in Physics 59. Springer, Berlin.
- Oran E S and Boris J P (1987). Numerical Simulation of Reactive Flow. Elsevier, New York.
- Sethian J A (1999). Level Set Methods and Fast Marching Methods. Cambridge University Press.
- Woodward P and Colella P (1984). The numerical simulation of two-dimensional flows with strong shocks. *J. Comput. Phys.* **54**(1), pp 115-173.
- Youngs D L (1982). Time dependent multi-material flow with large fluid distortion. Numerical Methods for Fluid Dynamics, pp 273-285. Morten K W and Baines M J (Editors). Academic Press.

A GODUNOV-TYPE METHOD FOR STUDYING THE LINEARISED STABILITY OF A FLOW. APPLICATION TO THE RICHTMYER-MESHKOV INSTABILITY

E. GODELEWSKI

LAN

Université Paris VI

75252 Paris Cedex 05, France.

Email: godlewski@ann.jussieu.fr

M. OLAZABAL

DCSA

CEA/DAM Ile-de-France

BP12

91680 Bruyères-Le-Châtel, France.

Email: olazabal@bruyeres.cea.fr

AND

P.-A. RAVIART

CMAP

Ecole Polytechnique

91128 Palaiseau Cedex, France.

Email: raviart@cmapx.polytechnique.fr

Abstract. We introduce a direct formulation of the linearisation of a non-linear hyperbolic system of conservation laws at a discontinuous solution. In this formulation, the solution of the linearised Cauchy problem is to be found in a space of measures. We give a numerical scheme of solution of this Cauchy problem based on a Roe linearisation and we apply it to study the linear phase of the Richtmyer-Meshkov instability in Lagrangian coordinates.

1. Introduction

In many fluid flow problems, we are led to study the linearised stability of a solution of a nonlinear hyperbolic system of conservation laws with respect to perturbations of the initial data. If this solution, called *basic solution*, is smooth there is no difficulty in linearising the nonlinear system around it. This is no longer the case if the basic solution presents discontinuities such as shock waves, contact discontinuities... . Indeed the linearised hyperbolic system has discontinuous coefficients and the Cauchy problem is in general ill-posed in any class of functions. However, if we look for measure solutions in the form of a sum of a function and a Dirac measure carried by the discontinuities of the basic solution, we are able to solve the linearised Cauchy problem provided we give a suitable natural sense to the nonconservative product of a Dirac measure and a discontinuous function. In fact, this so-called direct formulation of the linearised problem is formally equivalent to the classical one used for instance in Fluid Mechanics. Moreover, the measure solution appears to depend continuously on the data and can be accurately approximated by means of Godunov-type numerical schemes although these properties have been proved only in the scalar case. In this paper, we show how the direct formulation of the linearised problem can be used in order to study numerically the linearised stability of fluid flows in Lagrangian coordinates and, as an illustration of the effectiveness of the method, we apply it to the Richtmyer-Meshkov instability.

The plan of the paper is as follows. In Section 2, we present the direct formulation of the linearised problem and we state the existence result of the measure solution. We also show the equivalence of the direct approach with the classical one. In Section 3, we restrict ourselves to nonlinear hyperbolic systems for which there exists a formulation in Lagrangian coordinates. We exhibit the linearised system in Lagrangian coordinates and relate the linearised solutions in Eulerian and Lagrangian coordinates. Section 4 is devoted to the numerical solution of the linearised problem in Lagrangian coordinates. Starting from a Godunov-type method solution of the basic problem using a Roe linearisation, we show how to construct a fairly unexpensive numerical scheme for the linearised problem. In Section 5, we consider the Richtmyer-Meshkov instability which describes the instability of an interface initiated by the interaction of a shock wave. We show that, using the above method, one can easily compute the rate of growth of the linear phase of the instability. We compare the computed rate of growth with the exact one and also with that predicted by the impulsive model.

2. The linearised problem

We consider a nonlinear hyperbolic system of conservation laws in d space-variables

$$\frac{\partial \mathbf{U}}{\partial t} + \sum_{j=1}^d \frac{\partial}{\partial x_j} \mathbf{F}_j(\mathbf{U}) = \mathbf{0}. \quad (1)$$

In (1), each \mathbf{F}_j , $1 \leq j \leq d$, is a smooth function from the state space $\Omega \subset \mathbb{R}^n$ into \mathbb{R}^n . Let \mathbf{U}^0 be a solution of (1) called hereafter the *basic solution*: we assume that $\mathbf{U}^0 = \mathbf{U}^0(x_1, t)$ does not depend on the variables x_2, \dots, x_d . In other words, \mathbf{U}^0 satisfies

$$\frac{\partial \mathbf{U}^0}{\partial t} + \frac{\partial}{\partial x_1} \mathbf{F}_1(\mathbf{U}^0) = \mathbf{0}. \quad (2)$$

We denote by $\mathbf{U}_0^0 = \mathbf{U}_0^0(x_1) = \mathbf{U}^0(x_1, 0)$ the associated initial condition.

Now, we want to study the stability of the basic solution \mathbf{U}^0 with respect to multidimensional perturbations of the initial condition \mathbf{U}_0^0 . For any "small" perturbation

$$\mathbf{U}_0^\epsilon = \mathbf{U}_0^\epsilon(\mathbf{x}) = \mathbf{U}_0^0(\mathbf{x}) + \epsilon \mathbf{U}_0^1(\mathbf{x}) + \dots \quad (3)$$

of \mathbf{U}_0^0 , this amounts to analyse the behaviour in time of the solution \mathbf{U}^ϵ of (1) associated with the initial condition

$$\mathbf{U}^\epsilon(\mathbf{x}, 0) = \mathbf{U}_0^\epsilon. \quad (4)$$

Writing

$$\mathbf{U}^\epsilon(\mathbf{x}, 0) = \mathbf{U}^0(x_1, 0) + \epsilon \mathbf{U}^1(\mathbf{x}, 0), \quad (5)$$

we obtain that the first order perturbation \mathbf{U}^1 of the basic solution \mathbf{U}^0 is solution of the *linearised Cauchy problem*

$$\begin{cases} \frac{\partial \mathbf{U}^1}{\partial t} + \sum_{j=1}^d \frac{\partial}{\partial x_j} (\mathbf{A}_j(\mathbf{U}^0) \mathbf{U}^1) = \mathbf{0}, \mathbf{x} \in \mathbb{R}^d, t > 0, \\ \mathbf{U}^1(\mathbf{x}, 0) = \mathbf{U}_0^1(\mathbf{x}). \end{cases} \quad (6)$$

In (6), $\mathbf{A}_j(\mathbf{U})$ denotes the Jacobian matrix of $\mathbf{F}_j(\mathbf{U})$, $1 \leq j \leq d$. Then, studying the *linearised stability* of the basic solution \mathbf{U}^0 consists in analysing the behaviour in time of the first order perturbation \mathbf{U}^1 .

Since by hypothesis the linearised system is hyperbolic, the Cauchy problem (6) is well posed as soon as the basic solution \mathbf{U}^0 and therefore

$\mathbf{A}_j(\mathbf{U}^0)$, $1 \leq j \leq d - 1$, are smooth enough. However, in many cases of interest, the basic solution presents discontinuities (contact discontinuities, shock waves...) and moreover the initial perturbation \mathbf{U}_0^1 is not a function but a *measure*. Let us give an example of such a situation. Assume that \mathbf{U}^0 presents a jump discontinuity along the curve $x_1 = \phi^0(t)$, *i.e.*,

$$\mathbf{U}^0(x_1, t) = \begin{cases} \mathbf{U}_L(x_1, t), & x_1 < \phi^0(t), \\ \mathbf{U}_R(x_1, t), & x_1 > \phi^0(t), \end{cases} \quad (7)$$

where \mathbf{U}_L and \mathbf{U}_R are smooth solutions of (2) in the domains $\{(x_1, t); x_1 < \phi^0(t), t > 0\}$ and $\{(x_1, t); x_1 > \phi^0(t), t > 0\}$ respectively. We may suppose $\phi^0(0) = 0$ so that the initial discontinuity is located at $x_1 = 0$. Now, if we introduce a perturbation of the initial position of the discontinuity $x_1 = \epsilon\phi_0^1(\mathbf{y})$, $\mathbf{y} = (x_2, \dots, x_d)$, we are led to the perturbed initial condition

$$\mathbf{U}_0^\epsilon(\mathbf{x}) = \mathbf{U}_0^0(x_1 - \epsilon\phi_0^1(\mathbf{y}), 0) = \begin{cases} \mathbf{U}_L(x_1 - \epsilon\phi_0^1(\mathbf{y}), 0), & x_1 < \epsilon\phi_0^1(\mathbf{y}), \\ \mathbf{U}_R(x_1 - \epsilon\phi_0^1(\mathbf{y}), 0), & x_1 > \epsilon\phi_0^1(\mathbf{y}). \end{cases}$$

Then (3) holds with

$$\mathbf{U}_0^1(\mathbf{x}) = -\phi_0^1(\mathbf{y}) \frac{d\mathbf{U}_0^0}{dx_1}(x_1).$$

Since

$$\frac{d\mathbf{U}_0^0}{dx_1} = \left\{ \frac{d\mathbf{U}_0^0}{dx_1} \right\} + [\mathbf{U}_0^0] \delta(x_1),$$

where

$$\left\{ \frac{d\mathbf{U}_0^0}{dx_1} \right\}(x_1) = \begin{cases} \frac{\partial \mathbf{U}_L}{\partial x_1}(x_1, 0), & x_1 < 0, \\ \frac{\partial \mathbf{U}_R}{\partial x_1}(x_1, 0), & x_1 > 0, \end{cases}$$

is the "function part" of the distributional derivative $\frac{d\mathbf{U}_0^0}{dx_1}$ and $[\mathbf{U}_0^0] = (\mathbf{U}_R - \mathbf{U}_L)(0, 0)$ is the jump of \mathbf{U}_0^0 at the discontinuity, we obtain that \mathbf{U}_0^1 is indeed a measure.

An analogous but somewhat more complex situation will occur when studying first-order perturbations of the solution of a Riemann problem associated with perturbations of the position of the initial discontinuity. Such a problem is indeed related to the study of the linearised Richtmyer-Meshkov instability as it will become clear in Section 5.

The above example suggests to look for a *measure solution* of the linearised problem (6) when the basic solution \mathbf{U}^0 presents jump discontinuities. In fact, assuming that the basic solution \mathbf{U}^0 is of the form (7), we look for a solution to (6) of the form

$$\mathbf{U}^1(\mathbf{x}, t) = \left\{ \mathbf{U}^1 \right\}(\mathbf{x}, t) - \phi^1(\mathbf{y}, t) [\mathbf{U}^0] \delta_\Sigma, \quad (8)$$

where $\left\{ \mathbf{U}^1 \right\}$ is a function and δ_Σ is the Dirac measure carried by the curve $\Sigma = \{(x_1, t); x_1 = \phi^0(t), t \geq 0\}$. However a difficulty arises immediately when trying to give a sense to the expressions

$$\mathbf{A}_j(\mathbf{U}^0) \mathbf{U}^1, \quad 1 \leq j \leq d,$$

which involve the *nonconservative products*

$$\mathbf{A}_j(\mathbf{U}^0) [\mathbf{U}^0] \delta_\Sigma, \quad 1 \leq j \leq d,$$

of a discontinuous function and a Dirac measure whose support is the discontinuity of the function. All these products are of the form

$$\mathbf{A}(\mathbf{U}^0) [\mathbf{U}^0] \delta_\Sigma,$$

where \mathbf{A} is the Jacobian matrix of a mapping $\mathcal{F} : \Omega \rightarrow I\!\!R^n$ and it appears natural to define the above expression by using Volpert's product (Dal Maso, Le Floch and Murat, 1983)

$$\mathbf{A}(\mathbf{U}^0) [\mathbf{U}^0] \delta_\Sigma = [\mathcal{F}(\mathbf{U}^0)] \delta_\Sigma. \quad (9)$$

Then, we can state

Theorem 1

Assume that the nonconservative products are defined as in (9). Then, given an initial condition of the form

$$\mathbf{U}_0^1(\mathbf{x}) = \left\{ \mathbf{U}_0^1 \right\}(\mathbf{x}) - \phi_0^1(\mathbf{y}) [\mathbf{U}_0^0] \delta(\mathbf{x}), \quad (10)$$

where $\left\{ \mathbf{U}_0^1 \right\}$ and ϕ_0^1 are smooth functions, there exists a unique pair of smooth functions ($\left\{ \mathbf{U}^1 \right\}, \phi^1$), $\left\{ \mathbf{U}^1 \right\} = \left\{ \mathbf{U}^1 \right\}(\mathbf{x}, t)$, $\phi^1 = \phi^1(\mathbf{y}, t)$ such that

$$\mathbf{U}^1 = \left\{ \mathbf{U}^1 \right\} - \phi^1 [\mathbf{U}^0] \delta_\Sigma, \quad (11)$$

is a measure solution of the linearised problem (6).

For the proof, we refer to (Godlewski and Raviart, 2000).

Several remarks are now in order. First we note that the function $\{\mathbf{U}^1\}$ is indeed the solution of the linearised problem (6) in the domains $\{(\mathbf{x}, t); x_1 < \phi^0(t), t \geq 0\}$ and $\{(\mathbf{x}, t); x_1 > \phi^0(t), t \geq 0\}$ separately. Moreover, we have

$$\{\mathbf{U}^1\}(\mathbf{x}, 0) = \{\mathbf{U}_0^1\}(\mathbf{x}), \quad \phi^1(\mathbf{y}, 0) = \phi_0^1(\mathbf{y}),$$

and thus $\phi^1(0, \mathbf{y}) = 0$ if \mathbf{U}_0^1 is a function. However, we have in general $\phi^0(\mathbf{y}, t) \neq 0$ for all $t > 0$ so that the solution of (6) is a measure whatever the smoothness of \mathbf{U}_0^1 can be.

On the other hand, the question arises immediately to compare this so-called *direct approach* of the linearisation of a nonlinear system of conservation laws at a discontinuous basic solution with the classical one. In this later approach (Majda, 1983), it is assumed that the solution \mathbf{U}^ϵ of the perturbed problem presents a jump discontinuity along some surface Σ^ϵ of equation

$$x_1 = \phi^\epsilon(\mathbf{y}, t).$$

Then, we perform a change of variables and function

$$\begin{cases} \hat{x}_1 = x_1 - \phi^\epsilon(\mathbf{y}, t), \quad \hat{x}_j = x_j, 1 \leq j \leq d, \\ \hat{\mathbf{U}}^\epsilon(\hat{\mathbf{x}}, t) = \mathbf{U}^\epsilon(\hat{x}_1 + \phi^\epsilon(\mathbf{y}, t), \mathbf{y}, t), \end{cases} \quad (12)$$

so that $\hat{\mathbf{U}}^\epsilon$ is now discontinuous along the hyperplane $\hat{x}_1 = 0$. By writing

$$\begin{cases} \hat{\mathbf{U}}^\epsilon(\mathbf{x}, t) = \hat{\mathbf{U}}^0(\hat{x}_1, t) + \epsilon \hat{\mathbf{U}}^1(\hat{\mathbf{x}}, t) + \dots \\ \phi^\epsilon(\mathbf{y}, t) = \phi^0(t) + \epsilon \phi^1(\mathbf{y}, t) + \dots \end{cases} \quad (13)$$

we obtain on one hand that $\hat{\mathbf{U}}^0$ is given by

$$\hat{\mathbf{U}}^0(\hat{x}_1, t) = \mathbf{U}^0(\hat{x}_1 + \phi^0(t), t), \quad (14)$$

and on the other hand that the pair $(\hat{\mathbf{U}}^1, \phi^1)$ satisfies the linearised system

$$\begin{cases} \frac{\partial \hat{\mathbf{U}}^1}{\partial t} + \frac{\partial}{\partial \hat{x}_1} \left(\left[\mathbf{A}_1(\hat{\mathbf{U}}^0) - \frac{d\phi^0}{dt} \right] \hat{\mathbf{U}}^1 - \frac{\partial \phi^1}{\partial t} \hat{\mathbf{U}}^0 - \sum_{j=2}^d \frac{\partial \phi^1}{\partial x_j} \mathbf{F}_j(\hat{\mathbf{U}}^0) \right) + \\ + \sum_{j=2}^d \frac{\partial}{\partial x_j} (\mathbf{B}_j(\hat{\mathbf{U}}^0) \hat{\mathbf{U}}^1) = \mathbf{0} \end{cases} \quad (15)$$

in each domain $\{(\hat{\mathbf{x}}, t); \hat{x}_1 < 0, t > 0\}$ and $\{(\hat{\mathbf{x}}, t); \hat{x}_1 > 0, t > 0\}$ separately together with the initial condition

$$\hat{\mathbf{U}}^1(\hat{\mathbf{x}}, 0) = \hat{\mathbf{U}}_0^1(\hat{\mathbf{x}}) + \phi_0^1(\mathbf{y}) \frac{d\mathbf{U}^0}{dx_1}(\hat{\mathbf{x}}_1) \quad (16)$$

for $\hat{x}_1 < 0$ and $\hat{x}_1 > 0$ and the linearised Rankine-Hugoniot jump conditions at $\hat{x}_1 = 0$ which read

$$\left[\left(\mathbf{A}_1(\hat{\mathbf{U}}^0) - \frac{d\phi^0}{dt} \right) \hat{\mathbf{U}}^1 \right] = \frac{\partial \phi^1}{\partial t} [\hat{\mathbf{U}}^0] + \sum_{j=2}^d \frac{\partial \phi^1}{\partial x_j} [\mathbf{F}_j(\hat{\mathbf{U}}^0)]. \quad (17)$$

Such a problem (15)-(17) has a unique solution $(\hat{\mathbf{U}}^1, \phi^1)$ (Majda, 1983) and moreover one can prove (Godlewski and Raviart, 2000) that the pair $(\{\mathbf{U}^1\}, \phi^1)$ where

$$\{\mathbf{U}^1\}(\mathbf{x}, t) = \hat{\mathbf{U}}^1(x_1 - \phi^0(t), \mathbf{y}, t) - \phi^1(\mathbf{y}, t) \left\{ \frac{\partial \mathbf{U}^0}{\partial x_1} \right\} (x_1, t), \quad (18)$$

provides the measure solution (11) of the linearised problem (6).

Hence the direct formulation of the linearised problem appears to be equivalent to the classical one but is by far more suited to numerical computations as we will illustrate it in Section 4. However Theorem 1 relies on Volpert's definition of the non conservative product (9) which may be considered as artificial. Indeed, for fully justify the direct approach, it remains to show that the measure solution (11) to (6) depends continuously on the data. In fact, this is proved in the case of a scalar conservation law (Bouchut and James, 1997), (Godlewski, Olazabal and Raviart, 1998) but remains an open difficult problem in the case of systems although this continuity property is clearly demonstrated by numerical experiments.

3. The linearised problem in Lagrangian coordinates

From now on, we restrict ourselves to nonlinear hyperbolic systems which arise in Fluid Mechanics and enjoy a formulation in Lagrangian coordinates. Such systems are of the form

$$\begin{cases} \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0, \\ \frac{\partial}{\partial t}(\rho \psi) + \nabla \cdot ((\rho \psi \otimes \mathbf{u}) + \mathbf{f}(\rho, \psi)) = \mathbf{0}. \end{cases} \quad (19)$$

In (19), ρ denotes the mass density of the fluid, $\rho\psi = (\rho\psi_1, \dots, \rho\psi_{n-1})$ the vector of the other nonconservative variables, $\mathbf{u} = (u_1, u_2, u_3) = \mathbf{u}(\rho, \psi)$ the velocity of the fluid and

$$\mathbf{f} = (f_{ij})_{1 \leq i \leq n-1, 1 \leq j \leq 3},$$

is a $(n-1) \times 3$ tensor. Setting $\mathbf{f}_j = (f_{1j}, \dots, f_{n-1,j})^T$, $1 \leq j \leq 3$, the system (19) is indeed of the form (1) with

$$\mathbf{U} = \begin{pmatrix} \rho \\ \rho\psi \end{pmatrix}, \quad \mathbf{F}_j(\mathbf{U}) = \begin{pmatrix} \rho u_j \\ \rho u_j \psi + \mathbf{f}_j \end{pmatrix}, \quad 1 \leq j \leq 3.$$

Let us next introduce the Lagrangian coordinates. For any $\xi = (\xi_1, \xi_2, \xi_3) \in \mathbb{R}^3$, we define $t \rightarrow \mathbf{x}(\xi, t)$ to be the solution of the differential problem

$$\begin{cases} \frac{\partial \mathbf{x}}{\partial t} = \mathbf{u}(\mathbf{x}, t), \\ \mathbf{x}(0) = \xi \end{cases} \quad (20)$$

and we set

$$J(\xi, t) = \det \left(\frac{\partial x_i}{\partial \xi_j}(\xi, t) \right). \quad (21)$$

Now, with any function $\varphi = \varphi(\mathbf{x}, t)$, we associate the function $\bar{\varphi} = \bar{\varphi}(\xi, t)$ defined by

$$\bar{\varphi}(\xi, t) = \varphi(\mathbf{x}(\xi, t), t). \quad (22)$$

In other words, we express the function φ in the Lagrangian coordinates (ξ, t) . Then, introducing the specific volume $\tau = \rho^{-1}$ of the fluid and setting

$$\mathbf{V} = \begin{pmatrix} \tau \\ \psi \end{pmatrix}, \quad \mathbf{G}_j(\mathbf{V}) = \begin{pmatrix} -u_j \\ \mathbf{f}_j \left(\frac{1}{\tau}, \psi \right) \end{pmatrix}, \quad 1 \leq j \leq 3, \quad (23)$$

and passing in Lagrangian coordinates, it is a standard matter to check that the pair $(\bar{\mathbf{V}}, \mathbf{x})$ is solution of the system

$$\begin{cases} \rho_0 \frac{\partial \bar{\mathbf{V}}}{\partial t} + J \sum_{j=1}^3 \frac{\partial}{\partial x_j} \mathbf{G}_j(\mathbf{V}) = \mathbf{0}, \\ \frac{\partial \mathbf{x}}{\partial t} = \bar{\mathbf{u}}, \end{cases} \quad (24)$$

where

$$\rho_0(\xi) = \rho(\xi, 0), \quad J = \rho_0 \bar{\tau}. \quad (25)$$

Note that the basic solution in Lagrangian coordinates $(\bar{\mathbf{V}}^0, x_1^0) = (\mathbf{V}^0(\xi_1, t), x_1^0(\xi_1, t))$ is solution of the system in one-space dimension

$$\begin{cases} \rho_0^0 \frac{\partial \bar{\mathbf{V}}^0}{\partial t} + \frac{\partial}{\partial \xi_1} \mathbf{G}_1(\bar{\mathbf{V}}^0) = \mathbf{0}, \\ \frac{\partial x_1^0}{\partial t} = \bar{u}_1^0, \end{cases} \quad (26)$$

while for $j = 2, 3$, the function x_j^0 defined by

$$\frac{\partial x_j^0}{\partial t} = \bar{u}_j^0, \quad x_j^0(\xi, 0) = \xi_j, \quad (27)$$

depends on ξ_1, ξ_j and t . Note also that the basic solutions in Eulerian and Lagrangian coordinates are related by

$$\bar{\mathbf{V}}^0(\xi_1, t) = \mathbf{V}^0(x_1^0(\xi_1, t), t), \quad \mathbf{V}^0 = (\mathbf{V}(U^0)). \quad (28)$$

Now, with any function $\varphi = \varphi(\mathbf{x}, t)$ it is convenient to associate another function $\tilde{\varphi} = \tilde{\varphi}(\mathbf{x}^0, t)$ defined by

$$\tilde{\varphi}(\mathbf{x}^0(\xi, t), t) = \bar{\varphi}(\xi, t) = \varphi(\mathbf{x}(\xi, t), t). \quad (29)$$

This amounts to express the function φ in the frame of reference of the basic flow. Then (28) reads

$$\tilde{\mathbf{V}}^0 = \mathbf{V}^0. \quad (30)$$

Let us next relate the first order perturbations in Eulerian and Lagrangian coordinates. On one hand, we write

$$\mathbf{V}^\epsilon = \mathbf{V}(U^\epsilon) = \mathbf{V}(U^0 + \epsilon U^1 + \dots) = \mathbf{V}^0 + \epsilon \mathbf{V}^1 + \dots$$

which gives

$$\mathbf{V}^1 = \mathbf{V}'(U^0) \mathbf{U}^1. \quad (31)$$

On the other hand, starting from an asymptotic expansion of the perturbed solution $(\bar{\mathbf{V}}^\epsilon, \mathbf{x}^\epsilon)$ in Lagrangian coordinates

$$\begin{cases} \bar{\mathbf{V}}^\epsilon = \bar{\mathbf{V}}^0 + \epsilon \bar{\mathbf{V}}^1 + \dots, \\ \mathbf{x}^\epsilon = \mathbf{x}^0 + \epsilon \mathbf{x}^1 + \dots, \end{cases} \quad (32)$$

it is a simple matter to check that

$$\bar{\mathbf{V}}^1(\xi, t) = \mathbf{V}^1(\mathbf{x}^0(\xi, t), t) + x_1^1(\xi, t) \frac{\partial \mathbf{V}^0}{\partial x_1}(x_1^0(\xi, t), t), \quad (33)$$

or equivalently

$$\tilde{\mathbf{V}}^1 = \mathbf{V}^1 + \tilde{x}_1^1 \frac{\partial \mathbf{V}^0}{\partial x_1}. \quad (34)$$

Remark 1. Assume that the basic solution \mathbf{U}^0 is of the form (7), i.e., presents a jump discontinuity along the curve Σ . Then, using (9), (11) and (31), we obtain *formally*

$$\mathbf{V}^1 = \{\mathbf{V}^1\} - \phi^1 [\mathbf{V}^0] \delta_\Sigma, \quad \{\mathbf{V}^1\} = \mathbf{V}'(\mathbf{U}^0) \{\mathbf{U}^1\}$$

and by (34)

$$\tilde{\mathbf{V}}^1 = \{\tilde{\mathbf{V}}^1\} + (\tilde{x}_1^1 - \phi^1) [\mathbf{V}^0] \delta_\Sigma, \quad \{\tilde{\mathbf{V}}\}^1 = \{\mathbf{V}^1\} + \tilde{x}_1^1 \left\{ \frac{\partial \mathbf{V}^0}{\partial x_1^0} \right\}.$$

This gives

$$\bar{\mathbf{V}}^1 = \{\bar{\mathbf{V}}^1\}(\boldsymbol{\xi}, t) + (x_1^1(\boldsymbol{\xi}, t) - \phi^1(\mathbf{y}^0(\boldsymbol{\xi}, t), t)) [\bar{\mathbf{V}}^0] \delta_{\bar{\Sigma}} \quad (35)$$

where $\bar{\Sigma}$ is the curve of discontinuity of $\bar{\mathbf{V}}^0$. Hence $\bar{\mathbf{V}}^1$ is also a measure at least in general. Consider the case where Σ is a *material* contact discontinuity so that

$$\bar{\Sigma} = \{(\xi_1, t); \xi_1 = 0, t \geq 0\}.$$

If we assume $u_j^0, j = 2, 3$, we have

$$\mathbf{y}^0(\boldsymbol{\xi}, t) = \boldsymbol{\eta}, \quad \boldsymbol{\eta} = (\xi_2, \xi_3)$$

and (35) reads

$$\bar{\mathbf{V}}^1(\boldsymbol{\xi}, t) = \{\bar{\mathbf{V}}^1\}(\boldsymbol{\xi}, t) - (x_1^1(0, \boldsymbol{\eta}, t) - \phi^1(\boldsymbol{\eta}, t)) [\bar{\mathbf{V}}^0] \delta(\xi_1).$$

Next, we obtain from the linearised Rankine-Hugoniot jump conditions that

$$\bar{u}_1^1 = \frac{\partial \phi^1}{\partial t}$$

is continuous across $\bar{\Sigma}$. Together with

$$\frac{\partial x_1^1}{\partial t} = \bar{u}_1^1,$$

this yields

$$\frac{\partial}{\partial t} (x_1^1(0, \boldsymbol{\eta}, t) - \phi^1(\boldsymbol{\eta}, t)) = 0.$$

Note that $\mathbf{x}^1(\boldsymbol{\xi}, 0) = \mathbf{0}$, and therefore $x_1^1(0, \boldsymbol{\eta}, 0) - \phi^1(\boldsymbol{\eta}, 0) = -\phi^1(\boldsymbol{\eta}, 0)$. Hence if $\phi^1(\boldsymbol{\eta}, 0) = 0$, *i.e.*, if we perturb the initial position of the material interface, we have that $\bar{\mathbf{V}}^1 = \{\bar{\mathbf{V}}^1\}$ is no longer a measure but a function. For a precise mathematical proof of this result, we refer to (Godlewski, Olazabal and Raviart, 1999). However if we perturb this initial position, $\bar{\mathbf{V}}^1$ is a measure but the i -th component $\bar{\mathbf{V}}_i^1$ of $\bar{\mathbf{V}}$ still remains a function as soon as $[\bar{\mathbf{V}}_i^1] = 0$. This will be indeed the situation encountered in the Richtmyer-Meshkov instability considered in Section 5.

Since the coefficients of the linearised system (6) do not depend on the transverse variables $\mathbf{y} = (x_2, x_3)$, it is natural to perform a Fourier transform of \mathbf{U}^1 with respect to \mathbf{y} which amounts to consider the Fourier modes, *i.e.*, the solutions of the linearised system of the form

$$\mathbf{U}^1(\mathbf{x}, t) = \mathbf{U}^{1,k}(x_1, t) \exp(i\mathbf{k} \cdot \mathbf{y}), \quad \mathbf{k} = (k_1, k_2) \in \text{IR}^2,$$

or equivalently

$$\mathbf{V}^1(\mathbf{x}, t) = \mathbf{V}^{1,k}(x_1, t) \exp(i\mathbf{k} \cdot \mathbf{y}), \quad \mathbf{V}^{1,k} = \mathbf{V}'(\mathbf{U}^0) \mathbf{U}^{1,k}.$$

In terms of functions $\tilde{\mathbf{V}}^1$ and $\tilde{\mathbf{x}}^1$, we are thus led to consider solutions of the form

$$\begin{cases} \tilde{\mathbf{V}}^1(\mathbf{x}^0, t) = \tilde{\mathbf{V}}^{1,k}(x_1^0, t) \exp(i\mathbf{k} \cdot \mathbf{y}^0) \\ \tilde{\mathbf{x}}^1(\mathbf{x}^0, t) = \tilde{\mathbf{x}}^{1,k}(x_1^0, t) \exp(i\mathbf{k} \cdot \mathbf{y}^0) \end{cases}, \quad \mathbf{y}^0 = (x_2^0, x_3^0), \quad (36)$$

where by (34)

$$\tilde{\mathbf{V}}^{1,k} = \mathbf{V}^{1,k} + \tilde{x}_1^{1,k} \frac{\partial \mathbf{V}^0}{\partial x_1}.$$

Now, one can check that the pair $(\bar{\mathbf{V}}^{1,k}, \mathbf{x}^{1,k})$ defined by

$$\begin{cases} \bar{\mathbf{V}}^{1,k}(\xi_1, t) = \tilde{\mathbf{V}}^1(x_1^0(\xi_1, t), t) \\ \mathbf{x}^{1,k}(\xi_1, t) = \tilde{\mathbf{x}}^1(x_1^0(\xi_1, t), t) \end{cases} \quad (37)$$

is solution of the linearised system in one-space dimension and in Lagrangian coordinates

$$\left\{ \begin{array}{l} \rho_0^0 \frac{\partial \bar{\mathbf{V}}^{1,k}}{\partial t} + \frac{\partial}{\partial \xi_1} \left(\mathbf{B}_1(\bar{\mathbf{V}}^0) \bar{\mathbf{V}}^{1,k} \right) + \\ + i \sum_{j=2}^3 k_j \left(\rho_0^0 \mathbf{C}_j(\bar{\mathbf{V}}^0) \bar{\mathbf{V}}^{1,k} + \mathbf{D}_j(\bar{\mathbf{V}}^0) \mathbf{x}^{1,k} \right) = -\rho_0^{1,k} \frac{\partial \bar{\mathbf{V}}^0}{\partial t}, \\ \frac{\partial \mathbf{x}^{1,k}}{\partial t} + i \left(\sum_{j=2}^3 k_j \mathbf{u}_j^0 \right) \mathbf{x}^{1,k} = \bar{\mathbf{u}}^{1,k}, \end{array} \right. \quad (38)$$

where $\mathbf{B}_j(\mathbf{V})$ is the Jacobian matrix of $\mathbf{G}_j(\mathbf{V})$, $1 \leq j \leq 3$, and

$$\left\{ \begin{array}{l} \mathbf{C}_j(\bar{\mathbf{V}}^0) = \bar{u}_j^0 \mathbf{I} + \bar{\tau}^0 \mathbf{B}_j(\bar{\mathbf{V}}^0), \\ \mathbf{D}_j(\bar{\mathbf{V}}^0) \mathbf{x} = x_j \frac{\partial}{\partial \xi_1} \mathbf{G}_1(\bar{\mathbf{V}}^0) - x_1 \frac{\partial}{\partial \xi_1} \mathbf{G}_j(\bar{\mathbf{V}}^0). \end{array} \right. \quad (39)$$

In Lagrangian coordinates the Fourier modes thus read

$$\left\{ \begin{array}{l} \bar{\mathbf{V}}^1(\xi, t) = \bar{\mathbf{V}}^{1,k} \exp(i\mathbf{k} \cdot \mathbf{y}^0(\xi, t)), \\ \mathbf{x}^1(\xi, t) = \mathbf{x}^{1,k} \exp(i\mathbf{k} \cdot \mathbf{y}^0(\xi, t)). \end{array} \right. \quad (40)$$

4. A numerical method of Godunov-type

As it is classical in the Lagrangian formalism, we introduce a mass variable m defined by

$$dm = \rho_0^0(\xi_1) d\xi_1. \quad (41)$$

Then, setting for simplicity

$$\mathbf{U} = \bar{\mathbf{V}}^0, \quad \mathbf{V} = \bar{\mathbf{V}}^{1,k}, \quad \mathbf{x} = \bar{\mathbf{x}}^{1,k},$$

the basic problem is of the form

$$\left\{ \begin{array}{l} \frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{G}(\mathbf{U})}{\partial m} = 0, \\ \mathbf{U}(m, 0) = \mathbf{U}_0(m), \end{array} \right. \quad (42)$$

while, for each wave vector \mathbf{k} , the linearised problem reads

$$\left\{ \begin{array}{l} \frac{\partial \mathbf{V}}{\partial t} + \frac{\partial}{\partial m} (\mathbf{B}(\mathbf{U}) \mathbf{V}) = S(\mathbf{k}, \mathbf{U}; \mathbf{V}, \mathbf{x}), \\ \frac{\partial \mathbf{x}}{\partial t} + i \sum_{j=2}^3 k_j u_j^0 \mathbf{x} = \mathbf{u}, \\ \mathbf{V}(m, 0) = \mathbf{V}_0(m), \quad \mathbf{x}(m, 0) = \mathbf{x}_0(m). \end{array} \right. \quad (43)$$

In (43), $\mathbf{B}(\mathbf{U})$ is the Jacobian matrix of $\mathbf{G}(\mathbf{U})$.

Except for a few simple model problems, the solution of both the basic problem and the linearised one can be obtained numerically only. On one hand, finding a numerical solution of (42) is a standard problem in Computational Fluid Dynamics. Indeed, we have used a Godunov-type method based on a Roe linearisation $\mathbf{B}(\mathbf{U}_L, \mathbf{U}_R)$ of the flux function: for each pair of states $(\mathbf{U}_L, \mathbf{U}_R)$, $\mathbf{B}(\mathbf{U}_L, \mathbf{U}_R)$ is a $n \times n$ matrix with the standard properties

- (i) $\mathbf{B}(\mathbf{U}_L, \mathbf{U}_R)(\mathbf{U}_R - \mathbf{U}_L) = (\mathbf{G}(\mathbf{U}_R) - \mathbf{G}(\mathbf{U}_L))$
- (ii) $\mathbf{B}(\mathbf{U}_L, \mathbf{U}_R)$ is a diagonalisable matrix with real eigenvalues,
- (iii) $\mathbf{B}(\mathbf{U}, \mathbf{U}) = \mathbf{B}(\mathbf{U})$.

Indeed, a Roe linearisation of the gas dynamics equations has been constructed (Munz, 1994) and generalised to the compressible M.H.D. equations (Cargo, Gallice and Raviart, 1996). Now, given a mass increment Δm and a time increment Δt , we introduce the one-step explicit scheme

$$\left\{ \begin{array}{l} \mathbf{U}_j^{n+1} = \mathbf{U}_j^n - \frac{\Delta t}{\Delta m} \left[\left(\mathbf{B}_{j+1/2}^n \right)^- \left(\mathbf{U}_{j+1}^n - \mathbf{U}_j^n \right) + \right. \\ \left. + \left(\mathbf{B}_{j-1/2}^n \right)^+ \left(\mathbf{U}_j^n - \mathbf{U}_{j-1}^n \right) \right], \end{array} \right. \quad (44)$$

where \mathbf{U}_j^n stands for an approximation of $\mathbf{U}(j\Delta m, n\Delta t)$ and

$$\mathbf{B}_{j+1/2} = \mathbf{B}(\mathbf{U}_j, \mathbf{U}_{j+1}), \quad \mathbf{B}^\pm = \frac{1}{2} (\mathbf{B}_\pm |\mathbf{B}|).$$

On the other hand, it is a simple matter to construct a numerical scheme for the linearised problem (43) using the above Roe linearisation. We associate with (44) its “linearised” counterpart

$$\left\{ \begin{array}{l} \mathbf{V}_j^{n+1} = \mathbf{V}_j^n - \frac{\Delta t}{\Delta m} \left[(\mathbf{B}_{j+1/2}^n)^- \mathbf{V}_{j+1}^n + (\mathbf{B}_{j+1/2}^n)^+ \mathbf{V}_j^n - \right. \\ \left. - (\mathbf{B}_{j-1/2}^n)^- \mathbf{V}_j^n - (\mathbf{B}_{j-1/2}^n)^+ \mathbf{V}_{j-1}^n \right] + \Delta t \mathbf{S}(\mathbf{k}, \mathbf{U}_j^n, \mathbf{V}_j^n, \mathbf{x}_j^n), \\ \mathbf{x}_j^{n+1} = \left(1 - i\Delta t (k_2 u_2^0 + k_3 u_3^0)_j^n \right) \mathbf{x}_j^n + \Delta t \mathbf{u}_j^n. \end{array} \right. \quad (45)$$

Such a scheme has been proved to produce convergent approximations to measure solutions of scalar linear hyperbolic equations with discontinuous coefficients (Gosse and James, 1999).

It is worthwhile to notice that both numerical solutions of the basic problem (42) and the linearised one (43) are computed simultaneously. Moreover, most of the computational effort is done when computing \mathbf{U}^{n+1} , *i.e.*, when determining the Roe matrices $\mathbf{B}_{j+1/2}^n$ so that the computation of \mathbf{V}^{n+1} appears to be fairly cheap. Indeed linearising exactly the basic scheme (44) would have led to a complicated and inefficient “linearised” scheme.

The above schemes are only first order accurate and, as usual, one needs to use second-order schemes in order to obtain accurate solutions to the linearised problem and more specifically to predict reliable rates of growth of linear instabilities. This is achieved by using the MUSCL technique of Van Leer. For details, we refer to (Olazabal, 1998) and (Olazabal, Raviart and Cahen, 2000).

5. Application to the Richtmyer-Meshkov instability

The numerical method of Section 4 has been successfully applied to the study of the linearised stability of continuous and discontinuous hydrodynamic and magnetohydrodynamic flows. In particular, rates of growth of Rayleigh-Taylor instabilities or rippled shocks simulations agree fairly accurately with analytic computations (Olazabal, 1998), (Olazabal, Cahen and Raviart, 1999). Here we apply the method to the Richtmyer-Meshkov instability. We consider two fluids at rest separated by a plane interface. An incident plane shock parallel to the interface is propagating in the fluid 1 towards the fluid 2. At time $t=0$, the shock hits the interface. The initial condition of the basic problem thus consists of two states: the shocked fluid 1 and the fluid 2 at rest. The basic flow is then modelled as a solution of a Riemann problem for the system of gas dynamics: it consists of four constant states separated by a reflected wave (rarefaction or shock), a contact discontinuity (the interface) and a transmitted shock. The nature

of the reflected wave depends of the relative characteristics of the fluids and the incident shock strength. For a detailed description of the basic solution, we refer to (Yang, Zang and Sharp, 1994).

Suppose that we perturb the interface before the incident shock hits it. Then, after the shock has passed the interface, instabilities of the interface and the transmitted and reflected waves develop. This phenomenon is known as the Richtmyer-Meshkov instability. For studying the linear phase of this instability, we solve both the basic problem (26) and, for each “relevant” wave vector \mathbf{k} , the linearised system (38),(39), associated with the equations of gas dynamics in Lagrangian coordinates by means of the numerical method of Section 4. As an initial time $t=0$, we choose an instant before the incident shock hits the interface. At $t=0$, we thus introduce a perturbation of the position of the interface, which amounts to take

$$\bar{\mathbf{V}}^{1,k}(\xi_1, 0) = -\phi_0^{1,k} \left[\mathbf{V}_0^0 \right] \delta(\xi_1), \quad (46)$$

where here

$$\mathbf{V} = (\tau, \mathbf{u}^T, e)^T$$

and $e = \epsilon + \frac{1}{2}\mathbf{u}^2$ is the specific total energy. In (46),

$$\left[\mathbf{V}_0^0 \right] = (\left[\tau_0^0 \right], \mathbf{0}^T, \left[\epsilon_0^0 \right])^T,$$

denotes the initial jump of \mathbf{V}^0 at the interface $\xi_1 = 0$. Since the basic solution \mathbf{V}^0 presents jump discontinuities and using Remark 1, we expect that for each \mathbf{k} , the linearised solution $\bar{\mathbf{V}}^{1,k}$ is of the form

$$\bar{\mathbf{V}}^{1,k}(\xi_1, t) = \left\{ \bar{\mathbf{V}}^{1,k} \right\}(\xi_1, t) - \sum_{j=1}^3 \psi_j^{1,k} \left[\bar{\mathbf{V}}^0 \right] \delta_{\bar{\Sigma}_j}, \quad (47)$$

where $\bar{\Sigma}_j$ denotes the j th wave of the basic solution. In (47), $j=1$ and 3 correspond to the reflected wave and the transmitted shock wave while 2 refers to the contact discontinuity which is stationary in Lagrangian coordinates, *i.e.*,

$$\delta_{\bar{\Sigma}_2} = \delta(\xi_1). \quad (48)$$

If the reflected wave is a rarefaction, the contribution of this wave to the singular part of (47) does not exist which amounts to set $\psi_1^{1,k} = 0$. Since $[\bar{\mathbf{u}}^0](0, t) = \mathbf{0}$, it follows from (47) that $\bar{\mathbf{u}}^{1,k}$ does not present any Dirac measure at $\xi_1 = 0$. In fact, the function $\xi_1 \rightarrow \bar{\mathbf{u}}^{1,k}(\xi_1, t)$ is continuous at $\xi_1 = 0$.

Now, theoretical and experimental studies (Richtmyer, 1960), (Meshkov, 1969) of the linear phase of the Richtmyer-Meshkov instability show that the perturbed position amplitude $a(t)$ at the interface grows asymptotically linearly in time. Since

$$\dot{a}^k(t) = \frac{\partial x_1^{1,k}}{\partial t}(0, t) = \bar{u}_1^{1,k}(0, t),$$

the rate of growth of the instability can be measured by the perturbed velocity amplitude at the interface $\bar{u}_1^{1,k}(0, t)$.

We present below numerical simulations of the linear phase of the instability when the two fluids are ideal gases of different densities with the same adiabatic coefficient. We distinguish two cases: the *light/heavy case* where the incident shock propagates from the light gas to the heavy one and the *heavy/light case* where the incident shock propagates from the heavy gas to the light one. In the *light/heavy case*, the reflected wave is a shock while in the *light/heavy case*, the reflected wave is a rarefaction. We compare the computed asymptotic growth rates with those provided by impulsive models as the one introduced for the *heavy/light case* (Richtmyer, 1960) and extended to the *heavy/light case* as well to the case of two gases with different adiabatic coefficients (Meyer and Blewett, 1972), (Vandenboomgaerde, Mugler and Gauthier, 1998).

Comparison with the impulsive model

We first present two examples of simulations compared with the impulsive model formulation (Vandenboomgaerde, Mugler and Gauthier, 1998). This formulation is available in both *heavy/light* and *light/heavy* cases, when the shock strength $S = 1 - p_1/p_0$ is smaller than 0.4. Here, p_0 and p_1 denotes respectively the pressure of the shocked and unshocked gas. The two gases have the same adiabatic coefficient $\gamma = 5/3$.

In the case of a reflected shock, Figure 1 presents the densities of the unperturbed flow before and after the incident shock crosses the interface. The calculation is performed with the following initial density, pressure and velocity of the shocked gas: $\rho_0 = 0.5$, $p_0 = 5$, $u_0 = 1$, and the velocity of the unshocked gas: $u_1 = 0$. The incident shock strength is equal to $S = 0.347$. We take the density of the second gas as $\rho_2 = 0.6\rho_1$. The perturbation wave vector is $\mathbf{k} = (2\pi, 0)^T$.

In Figure 2, we plot the perturbed velocity amplitude at the interface, which gives the growth rate of the linear instability. The oscillations observed on the perturbed velocity are due the presence of acoustic modes

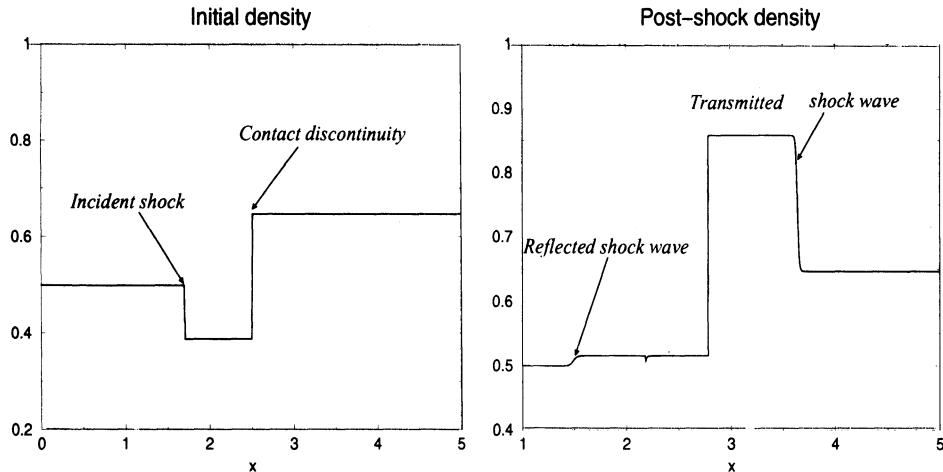


Figure 1. Unperturbed density before and after the shock/interface interaction: case of a reflected shock wave

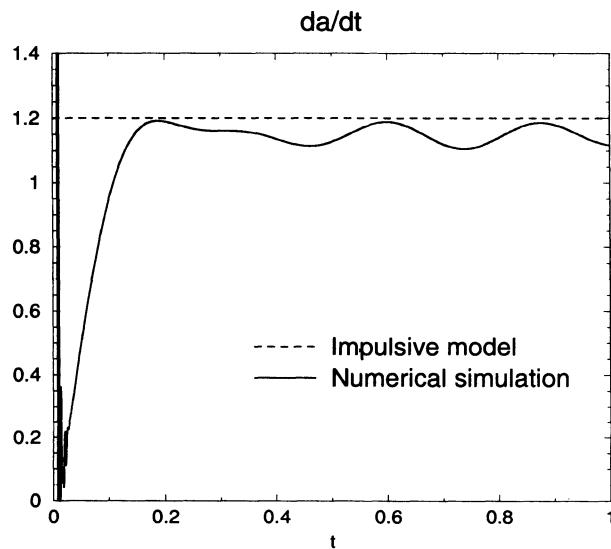


Figure 2. Perturbed velocity amplitude $u^k(t)$ at the interface, *i.e.* the growth rate da/dt of the instability, after the shock/interface interaction: case of a reflected shock wave.

and vortex modes which propagate between the transmitted and reflected shocks. We compare this growth rate to the asymptotic rate predicted by the impulsive model. In Figure 3, we have plotted the perturbed velocity profile at two different times. Calculations have been performed for several values of k and $S \leq 0.4$, taking 100 cells per wavelength, with the same accuracy, *i.e.* a relative error smaller than 5%. We can observe on the first figure that a Dirac mass appears at each shock locations. These Dirac mass

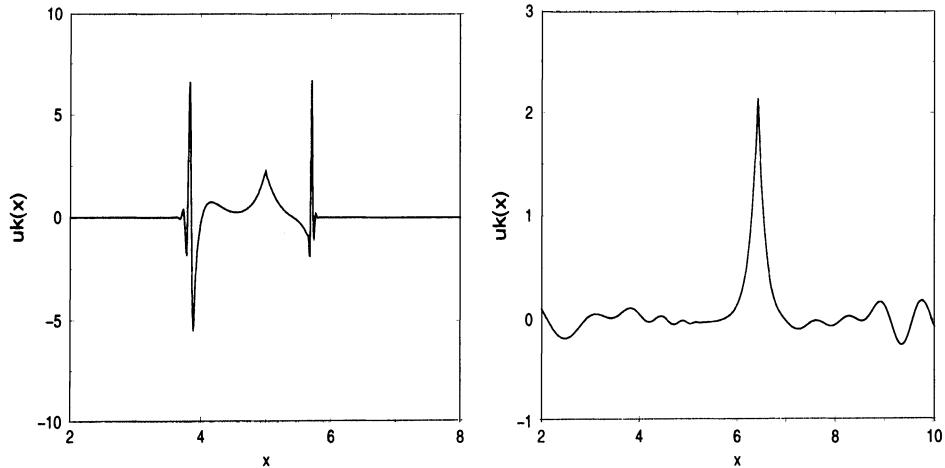


Figure 3. Perturbed velocity profiles $u^k(x)$ at two different times in the case of a reflected shock wave: at $t=0.4$ (on the left), we observe a dirac mass at each shock location and its interaction with the modes (acoustics, vortex) which propagate between the two shocks. At a later time $t=2.0$ (on the right), a typical interface mode profile appears.

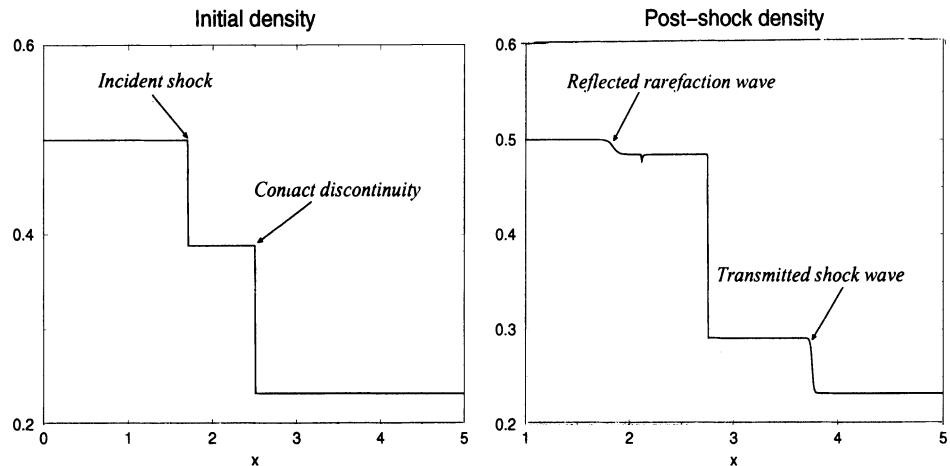


Figure 4. Unperturbed density before and after the shock/interface interaction: case of a reflected rarefaction wave.

interact with the acoustic waves which propagate between the two shock waves. At the interface, we observe the development of a perturbation. At a later time, when the transmitted and reflected waves are far enough from the interface, we obtain a typical interface mode profile.

Figures 4 and 5 illustrate the case of a reflected rarefaction wave. The initial data are the same as in the previous calculations except that we have

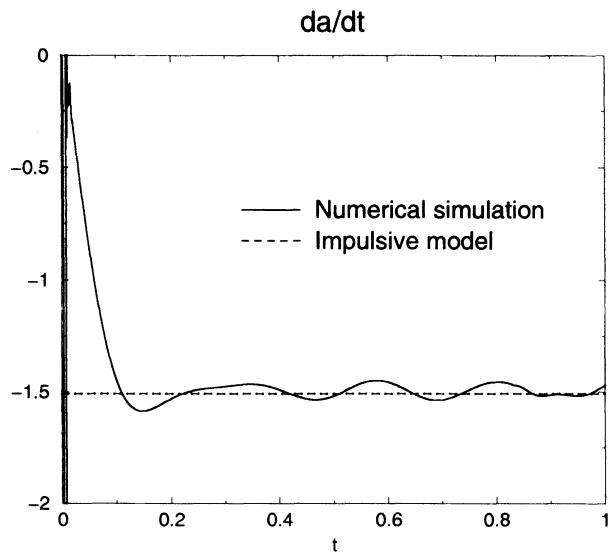


Figure 5. Perturbed velocity amplitude at the interface $u^k(t) = da/dt$ after the shock/interface interaction: case of a reflected rarefaction wave.

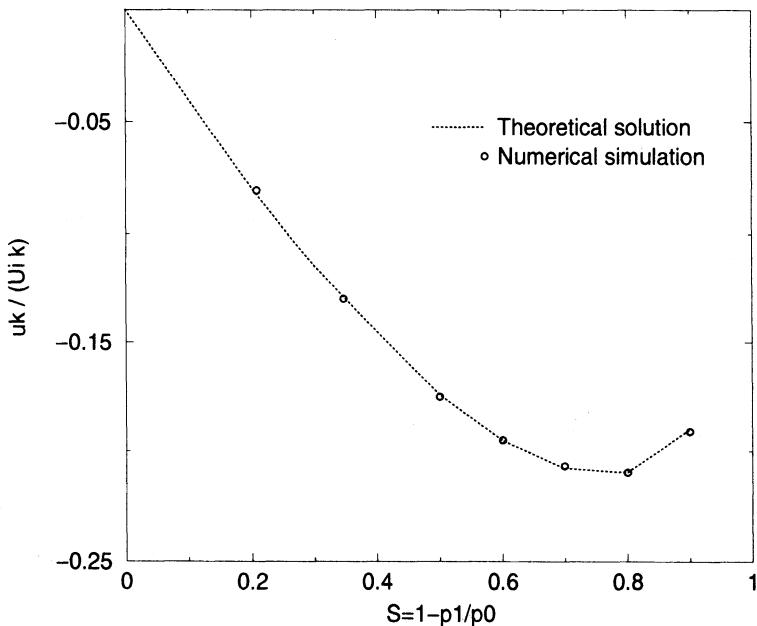


Figure 6. Perturbed velocity amplitude at the interface $u^k(t) = da/dt$ in terms of the incident shock strength S : case of a reflected rarefaction wave (U_i is the velocity of the incident shock).

$\rho_1 = 0.6\rho_2$. Again we obtain a good agreement with the impulsive model.

Comparison with an exact solution

In order to validate the numerical calculations for larger values of the incident shock strength, we have compared them with analytic solutions (Wouchuk and Nishihara, 1996) obtained as sums of series expansions in Bessel functions.

Now, the densities of the two gases verify $\rho_1/\rho_2 = 3.29$. Fig. 6 shows a good agreement between the growth rates of the numerical and analytic approaches for several values of S , in the case of a reflected rarefaction wave.

In the reflected shock case, a good agreement have been also obtained for small values of S . More detailed results will be published in a forthcoming paper.

6. Conclusion

We have presented here a direct approach to the linearisation of a non-linear hyperbolic system at a discontinuous solution. The solution of the corresponding perturbed problem has to be found in the class of measures, which implies that we have to give a sense to the associated non-conservative product. This direct formulation, established in Eulerian and Lagrangian coordinates, had led to the development of a Godunov-type scheme for the linearised problem. In previous works, the method has been successfully applied to the calculation of different instabilities occuring in Fluids Mechanics problems. We have presented here an application to the Richthmyer-Meshkov instability. A good agreement has been observed between numerical results and analytic models.

References

- Bouchut F and James F (1997) One-dimensional transport equations with discontinuous coefficients, *Nonlinear Analysis TMA* **32**, 891-933.
- Cargo P, Gallice G and Raviart P A (1996). Construction d'une linéarisée de Roe pour les équations de la MHD idéale, *Num. Anal., C.R. Acad. Sci. Paris* **323** (Série I) , 951-955.
- Dal Maso G, Le Floch P and Murat F (1983). Definition and weak stability of nonconservative products, *J. Math. Pures Appl.* **74** , 483-548.
- Gosse L and James F. Numerical approximations of one-dimensional conservation equations with discontinuous coefficients (to appear).
- Godlewski E and Raviart P A (1999). *Numerical Approximation of hyperbolic Systems of conservation laws*, Appl. Math. Science **118**, Springer, New York.
- Godlewski E and Raviart P A (1999). The linearized stability of solutions of nonlinear hyperbolic systems of conservation laws. A general numerical approach, *Mathematics and Computers in Simulation* **50**, 77-95.

- Godlewski E, Olazabal M and Raviart P A (1998). On the linearization of hyperbolic systems of conservation laws. Application to stability, *Equations aux dérivées partielles et Applications (Articles dédiés à J.-L. LIONS)*, Gauthier Villars, Paris , 549-570.
- Godlewski E, Olazabal M, Raviart P A (1999). On the linearization of systems of conservation laws for fluids at a material contact discontinuity, *J. Math. Pures Appl.*, **78**, 1013-1042.
- Majda A (1983) *The stability of multidimensional shock fronts*, Memoirs of the A.M.S. **275**, Amer. Math. Soc., Providence.
- Meshkov E E (1969). Izv. Akad. Nauk SSSR, Mekh.Zhidk.Gaz **5**
- Meyer K A and Blewet P J(1972). Numerical investigation of the stability of a shock-accelerated interface between two fluids, *Phys. Fluids* **15** (5), 735-759.
- Munz C D (1994). On Godunov-type schemes for lagrangian gaz dynamics, *SIAM J.Numer.Anal.* **31** (1), 17-42.
- Olazabal M (1998). 'Modélisation numérique de problèmes de stabilité linéaire. Application aux équations de la dynamique des gaz et de la MHD idéale', Doctoral dissertation, Université P. et M. Curie, Paris.
- Olazabal M, Cahen J and Raviart P A (1999). *Numerical modeling of linear stability problems applied to gas dynamics and MHD*, in Proceedings of the First international conference on Inertial Fusion Science and Applications (IFSA'99), Bordeaux, Elsevier.
- Olazabal M, Raviart P A and Cahen J (2000). Numerical modeling using Lagrangian formalism for studying the linear stability of compressible flows (to appear)
- Richtmyer R D (1960). Taylor instability in shock acceleration of compressible fluids, *Comm. Pure Appl. Math.* **13**, 297-319.
- Vandenboomgaerde M, Mugler C and Gauthier S (1998). Impulsive model for the Richtmyer-Meshkov instability, *Phys. Rev. E* **58** (2), 1874-1882.
- Wouchuk J G and Nishihara K (1996). Linear perturbation growth at a shocked interface, *Phys. Plasmas* **3** (10), 3761-3776.
- Yang Y, Zhang Q and Sharp D H (1994). Small amplitude theory of Richtmyer-Meshkov instability, *Phys. Fluids* **6**, 1856-1873.

THERMODYNAMICS, CONSERVATION LAWS AND THEIR ROTATION INVARIANCE

S. K. GODUNOV

Sobolev Institute of Mathematics

Novosibirsk, 630090, RUSSIA

Email:godunov@math.nsc.ru

Late in the 1950s, when I developed methods of gasdynamics computations, I, being under the influence of variational principles, arrive at a conclusion that the equations of gasdynamics belong to the class of equations of the following form:

$$\frac{\partial L_{q_i}}{\partial t} + \frac{\partial M_{q_i}^j}{\partial x_j} = 0. \quad (1)$$

On smooth solutions to these equations, one more additional relation holds:

$$\frac{\partial(q_i L_{q_i} - L)}{\partial t} + \frac{\partial(q_i M_{q_i}^j - M^j)}{\partial x_j} = 0. \quad (2)$$

In fact, this relation is a thermodynamic identity.

The trouble with this relation is that we have to use four thermodynamic potentials: $L = L(q_1, q_2, \dots)$, $M^1 = M^1(q_1, q_2, \dots)$, M^2 , and M^3 , whereas only one potential usually used in thermodynamics.

To overcome this obstacle, it is convenient to distinguish variables u_1 , u_2 , and u_3 (the components of the velocity), consider only one generating potential $L = L(u_1, u_2, u_3; q_1, q_2, \dots, q_n)$, set $M^j = u_j L$, and write out the equations and the additional conservation law as follows:

$$\begin{aligned} \frac{\partial L_{q_j}}{\partial t} + \frac{\partial(u_j L_{q_j})}{\partial x_j} &= 0, \\ \frac{\partial L_{u_k}}{\partial t} + \frac{\partial[(u_j L)_{u_k}]}{\partial x_j} &= 0 \\ \frac{\partial}{\partial t}(q_k L_{q_k} + u_k L_{u_k} - L) + \frac{\partial}{\partial x_j}[u_j(q_k L_{q_k} + u_k L_{u_k})] &= 0. \end{aligned} \quad (3)$$

The number of equations is greater than the number of unknown functions by 1. The system is overdetermined but compatible.

The following example is the system of equation of nonlinear theory of elasticity. We write it in the Euler coordinates. To this end, we use the generating thermodynamic potential

$$L = L(\{u_i\}, T, q, \{r_{ij}\})$$

depending on the velocity vector with components u_i , the temperature T , some nonsymmetric tensor r_{ij} describing deformation, and a parameter q which could be referred to as “free enthalpy”. The systems consists of the basis equations

$$\begin{array}{c|l} T & \frac{\partial L_T}{\partial t} + \frac{\partial(u_k L_T)}{\partial x_k} = 0, \\ q & \frac{\partial L_q}{\partial t} + \frac{\partial(u_k L_q)}{\partial x_k} = 0, \\ u_i & \frac{\partial L_{u_i}}{\partial t} + \frac{\partial[(u_k L)_{u_i} - r_{i\alpha} L_{r_{k\alpha}}]}{\partial x_k} = 0, \\ r_{ij} & \frac{\partial L_{r_{ij}}}{\partial t} + \frac{\partial[u_k L_{r_{ij}} - u_i L_{r_{kj}}]}{\partial x_k} = 0. \end{array} \quad (4)$$

The number of equations of this system coincides with the number of unknown functions. On the left-hand side of the last formula, we indicate factors of the corresponding equations. Multiplying equations by these factors, we can derive, as a consequence of these equations, the conservation law of energy. Unlike gasdynamics, we need the following additional equations:

$$u_1 r_{1j} + u_2 r_{2j} + u_3 r_{3j} \quad \left| \quad \frac{\partial L_{r_{ij}}}{\partial x_i} = 0 \quad (j = 1, 2, 3) \right. \quad (5)$$

These equations are not a consequence of the above equations but are compatible with them. From the basis equations we can obtain only the equalities

$$\frac{\partial}{\partial t} \left(\frac{\partial L_{ij}}{\partial x_i} \right) = 0.$$

The numbers to the left of the additional equations play role of factors when we derive the conservation law of energy.

Consider a linear combination of all the equations (conservation laws) multiplied by the indicated factors (playing the role of the coefficients of the linear combination). The linear combination can be represented as the conservation law:

$$\begin{aligned} & \frac{\partial}{\partial t} (T L_T + q L_q + u_i L_{u_i} + r_{ij} L_{r_{ij}} - L) + \\ & + \frac{\partial}{\partial x_k} [u_k (T L_T + q L_q + u_i L_{u_i} + r_{ij} L_{r_{ij}}) - u_\alpha r_{\alpha\beta} L_{r_{k\beta}}] = 0. \end{aligned} \quad (6)$$

Thus, in the elasticity theory, we have 14 basis equations and 4=3+1 additional equations, which are compatible with the basis ones, for 14 unknown functions (T , q , three components u_i , and nine components r_{ij}).

Similar systematization was done for some other equations of modern classical physics. These equations were represented in the form of compatible overdetermined system of conservation laws, called *thermodynamically compatible equations*. Description and motivation of our investigation concerning systematization of such equations can be found in two last chapters of the recently published (in Russian) book S.K. Godunov and E.I. Romen-sky “Elements of Mechanics of Continuous Media, and Conservation Laws” [1]

However, the obtained systematization is not final because the derivation rules turn out to be different for different concrete problems. It should be note that we studied various examples from magnetohydrodynamics, electrically conducting media, superfluidity and superconductivity.

Such a situation called our attention to the invariance of physical laws relative to the choice of coordinates and to conditions on thermodynamically compatible equations which must be required in view of this invariance. First of all, it is necessary to consider equations invariant relative to rotations, orthogonal transformations of coordinates.

We present a rather large but not yet complete class of conservation laws that are invariant relative to rotations. On the whole, we follow the work [2] devoted to equations in the Lagrange coordinates. However, the mentioned work uses some cumbersome tool of generating functions, which hampers the understanding of a rather simple idea. We try to find a more available way of presentation. To this end, we introduce a special symbolics that generalizes the usual notation of the inner product and the vector product. Using this symbolics, we can formally write the equations in a simpler form. I would like to thank V. M. Gordienko for his help in preparing this material.

In the case of the Galilei invariance, a rather large class of thermodynamically compatible conservation laws was described with the help of generating functions in [3]. Perhaps, the ideas of [3] become clear if the result will be obtained by using the formalism presented below. It is hoped that the further development of the group approach to the classification of conservation laws and thermodynamical identities will lead to a better understanding of the structure of equations of mathematical physics and equations describing evolution of continuous media, which provides, in turn, the further development of the mathematical theory and the computational algorithms as well.

Any vector $u^{(1)}$ of the three-dimensional space is given by three numbers (coordinates) $u_{-1}^{(1)}$, $u_0^{(1)}$, and $u_1^{(1)}$ in an orthogonal basis. The superscript (1)

indicates the dimension 3. In the sequel, we will deal with vectors of some odd dimension. Such a vector is given by $2N + 1$ coordinates

$$u_{-N}^{(N)}, u_{-N+1}^{(N)}, \dots, u_{-1}^{(N)}, u_0^{(N)}, u_1^{(N)}, \dots, u_{N-1}^{(N)}, u_N^{(N)}.$$

The integer N is called a *spin*. In the quantum mechanics, vectors given by some even number of components corresponding to semi-integer spins are widely used. However, we restrict ourselves to the case where all unknown vector-valued functions are of odd dimension.

With every rotation of the coordinates of the initial three-dimensional space, i.e., with every transformation of $\text{SO}(3)$, we associate an orthogonal transformation of the coordinates $u_n^{(N)}$ of the vector $u^{(N)}$ with the help of an irreducible representation of the group $\text{SO}(3)$ of weight N . The matrices of the irreducible representation are defined if a certain canonical basis for the representation space is chosen.

The table formed by the products $u_k^{(K)} v_l^{(L)}$ of the coordinates of the vectors $u^{(K)}$ and $v^{(L)}$, i.e., the Kronecker products $[u^{(K)} \times v^{(L)}]$, can be regarded as a vector of the space of dimension $(2K + 1) \cdot (2L + 1)$. A naturally defined representation of the rotation group acts in this space. It is induced by the representations by which the vectors $u^{(K)}$ and $v^{(L)}$ are transformed. However, the obtained representation is not irreducible in the $(2K + 1)(2L + 1)$ -dimensional space. It is convenient to write the table

$$[u^{(K)} \times v^{(L)}] = \begin{bmatrix} u_{-K} v_{-L} & u_{-K} v_{-L+1} & \dots & u_{-K} v_L \\ u_{-K+1} v_{-L} & u_{-K+1} v_{-L+1} & \dots & u_{-K+1} v_L \\ \vdots & \vdots & & \vdots \\ u_K v_{-L} & u_K v_{-L+1} & \dots & u_K v_L \end{bmatrix} \quad (7)$$

as a linear combination of some special tables that form a basis for the space of tables. This basis is orthonormal with respect to some inner product introduced below.

For elements of the basis we take $(2K + 1) \times (2L + 1)$ -matrices $C_{M[KL]}^m$ formed by elements $C_{M[KL]}^{m[kl]}$ ($-K \leq k \leq K, -L \leq l \leq L$). These matrices are gathered in $K + L - |K - L| + 1$ series with number $M = K + L, K + L - 1, \dots, |K - L| + 1, |K - L|$. In every series, the matrices are enumerated by index m ($-M \leq m \leq M$). Elements $C_{M[KL]}^{m[kl]}$ of such matrices are called the *Clebsch-Gordan coefficients*. The matrices will be called the *basis Clebsch-Gordan matrices*. Sometimes, we will not indicate the dimensions K, L ($C_M^m \equiv C_{M[KL]}^m$) in the notation. These matrices are real and orthonormal with respect to the inner product $(A, B) = \text{tr}[A^T B]$, i.e.,

$$\text{tr}\{[C_{M[KL]}]^T C_{N[KL]}^n\} = \text{tr}\{[C_M^m]^T C_N^n\} = \delta_{MN} \cdot \delta_{mn}. \quad (8)$$

Representing the Kronecker product $[u^{(K)} \times v^{(L)}]$ as a linear combination of the basis matrices C_M^m ,

$$[u^{(K)} \times v^{(L)}] = \sum_{M=|K-L|}^{M=K+L} \sum_{m=-M}^{m=M} w_m^{(M)} C_M^m, \quad (9)$$

we find that

$$w_m^{(M)} = \text{tr} \{ [C_M^m]^T [u^{(K)} \times v^{(L)}] \}. \quad (10)$$

The Clebsch–Gordan matrices C_M^m possess the following remarkable property. Under linear transformations of the vectors $u^{(K)}$ and $v^{(L)}$ that realize irreducible representations (in the canonical basis) of weights K and L of the rotation group, each vector $w^{(M)}$ with coordinates $w_m^{(M)}$ ($-M \leq m \leq M$) is also transformed according to an irreducible representation of the corresponding weight M . This allows us to define the “vector product of weight M ” in a natural way:

$$[u^{(K)} \times v^{(L)}]^{(M)} = w^{(M)}. \quad (11)$$

Moreover, for a fixed M the matrices $C_M^m \equiv C_{M[KL]}^m$ can be regarded as a canonical basis for some uniquely defined $2M + 1$ -dimensional subspace of the linear space of $(2K + 1) \times (2L + 1)$ -matrices. The irreducible representation of weight M acts in this subspace.

To clarify the idea, we consider an example. We represent the Kronecker product of two three-dimensional vectors

$$[u^{(1)} \times v^{(1)}] = \begin{bmatrix} u_{-1}^{(1)} v_{-1}^{(1)} & u_{-1}^{(1)} v_0^{(1)} & u_{-1}^{(1)} v_1^{(1)} \\ u_0^{(1)} v_{-1}^{(1)} & u_0^{(1)} v_0^{(1)} & u_0^{(1)} v_1^{(1)} \\ u_1^{(1)} v_{-1}^{(1)} & u_1^{(1)} v_0^{(1)} & u_1^{(1)} v_1^{(1)} \end{bmatrix}$$

in the form of a linear combination as follows:

$$\begin{aligned} [u^{(1)} \times v^{(1)}] &= \frac{u_{-1}^{(1)} v_{-1}^{(1)} + u_0^{(1)} v_0^{(1)} + u_1^{(1)} v_1^{(1)}}{\sqrt{3}} \begin{bmatrix} 1/\sqrt{3} & 0 & 0 \\ 0 & 1/\sqrt{3} & 0 \\ 0 & 0 & 1/\sqrt{3} \end{bmatrix} + \\ &+ \frac{u_{-1}^{(1)} v_0^{(1)} - u_0^{(1)} v_{-1}^{(1)}}{\sqrt{2}} \begin{bmatrix} 0 & 1/\sqrt{2} & 0 \\ -1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \\ &+ \frac{u_0^{(1)} v_1^{(1)} - u_1^{(1)} v_0^{(1)}}{\sqrt{2}} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 0 \end{bmatrix} + \end{aligned}$$

$$\begin{aligned}
& + \frac{u_1^{(1)} v_{-1}^{(1)} - u_{-1}^{(1)} v_1^{(1)}}{\sqrt{2}} \begin{bmatrix} 0 & 0 & 1/\sqrt{2} \\ 0 & 0 & 0 \\ -1/\sqrt{2} & 0 & 0 \end{bmatrix} + \\
& + \frac{u_{-1}^{(1)} v_{-1}^{(1)} - 2u_0^{(1)} v_0^{(1)} + u_1^{(1)} v_1^{(1)}}{\sqrt{6}} \begin{bmatrix} 1/\sqrt{6} & 0 & 0 \\ 0 & -2/\sqrt{6} & 0 \\ 0 & 0 & 1/\sqrt{6} \end{bmatrix} + \\
& + \frac{u_{-1}^{(1)} v_{-1}^{(1)} - u_1^{(1)} v_1^{(1)}}{\sqrt{2}} \begin{bmatrix} 1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1/\sqrt{2} \end{bmatrix} + \\
& + \frac{u_{-1}^{(1)} v_0^{(1)} + u_0^{(1)} v_{-1}^{(1)}}{\sqrt{2}} \begin{bmatrix} 0 & 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \\
& + \frac{u_0^{(1)} v_1^{(1)} + u_1^{(1)} v_0^{(1)}}{\sqrt{2}} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 0 \end{bmatrix} + \\
& + \frac{u_1^{(1)} v_{-1}^{(1)} + u_{-1}^{(1)} v_1^{(1)}}{\sqrt{2}} \begin{bmatrix} 0 & 0 & 1/\sqrt{2} \\ 0 & 0 & 0 \\ 1/\sqrt{2} & 0 & 0 \end{bmatrix}.
\end{aligned}$$

On the right-hand side of this linear combination there is the sum of the Clebsch–Gordan matrices multiplied by some factors:

$$\begin{aligned}
C_0^0 &= \begin{bmatrix} 1/\sqrt{3} & 0 & 0 \\ 0 & 1/\sqrt{3} & 0 \\ 0 & 0 & 1/\sqrt{3} \end{bmatrix}, \\
C_1^{-1} &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 0 \end{bmatrix}, \quad C_1^0 = \begin{bmatrix} 0 & 0 & 1/\sqrt{2} \\ 0 & 0 & 0 \\ -1/\sqrt{2} & 0 & 0 \end{bmatrix}, \\
C_1^1 &= \begin{bmatrix} 0 & 1/\sqrt{2} & 0 \\ -1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \\
C_2^{-2} &= \begin{bmatrix} 0 & 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad C_2^{-1} = \begin{bmatrix} 0 & 0 & 1/\sqrt{2} \\ 0 & 0 & 0 \\ 1/\sqrt{2} & 0 & 0 \end{bmatrix},
\end{aligned}$$

$$C_2^0 = \begin{bmatrix} 1/\sqrt{6} & 0 & 0 \\ 0 & -2/\sqrt{6} & 0 \\ 0 & 0 & 1/\sqrt{6} \end{bmatrix}, \quad C_2^1 = \begin{bmatrix} 1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1/\sqrt{2} \end{bmatrix},$$

$$C_2^2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 0 \end{bmatrix},$$

whereas the coefficients of this linear combination are the coordinates of the vectors

$$w^{(0)} \equiv w_0^{(0)} = \frac{u_{-1}^{(1)} v_{-1}^{(1)} + u_0^{(1)} v_0^{(1)} + u_1^{(1)} v_1^{(1)}}{\sqrt{3}}$$

(a vector of dimension 0 or weight 1, i.e., a scalar)

$$w^{(1)} = \begin{bmatrix} w_{-1}^{(1)} \\ w_0^{(1)} \\ w_1^{(1)} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} u_0^{(1)} v_1^{(1)} - u_1^{(1)} v_0^{(1)} \\ u_1^{(1)} v_{-1}^{(1)} - u_{-1}^{(1)} v_1^{(1)} \\ u_{-1}^{(1)} v_0^{(1)} - u_0^{(1)} v_{-1}^{(1)} \end{bmatrix}$$

(a vector of dimension 3 or weight 1)

$$w^{(2)} = \begin{bmatrix} w_{-2}^{(2)} \\ w_{-1}^{(2)} \\ w_0^{(2)} \\ w_1^{(2)} \\ w_2^{(2)} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} u_{-1}^{(1)} v_0^{(1)} + u_0^{(1)} v_{-1}^{(1)} \\ \frac{u_1^{(1)} v_{-1}^{(1)} + u_{-1}^{(1)} v_1^{(1)}}{\sqrt{3}} \\ u_1^{(1)} v_1^{(1)} - u_{-1}^{(1)} v_{-1}^{(1)} \\ u_0^{(1)} v_1^{(1)} + u_1^{(1)} v_0^{(1)} \end{bmatrix}$$

(a vector of dimension 5 or weight 2)

It is obvious that $w^{(0)}$ and $w^{(1)}$ differ from the usual inner product and the vector product of the vectors $u^{(1)}$ and $v^{(1)}$ only by normalized factors $1/\sqrt{3}$ and $1/\sqrt{2}$ respectively.

In the sequel, we need invariant differential operators transforming vector valued functions $v^{(N)}(x_{-1}, x_0, x_1)$ that can be defined with the help of the gradient operator

$$\nabla^{(1)} = \begin{bmatrix} \frac{\partial}{\partial x_{-1}} \\ \frac{\partial}{\partial x_0} \\ \frac{\partial}{\partial x_1} \end{bmatrix} \quad (12)$$

by the following equalities:

$$\begin{aligned}\Delta_+^{(N)} v^{(N)} &\equiv [\nabla^{(1)} \times v^{(N)}]^{(N+1)} = w^{(N+1)} \\ \Delta_0^{(N)} v^{(N)} &\equiv [\nabla^{(1)} \times v^{(N)}]^{(N)} = w^{(N)} \\ \Delta_-^{(N)} v^{(N)} &\equiv [\nabla^{(1)} \times v^{(N)}]^{(N-1)} = w^{(N-1)}.\end{aligned}\tag{13}$$

These operators transform $v^{(N)}$ into vector-valued functions that are transformed by irreducible representations of weights $N + 1$, N , and $N - 1$ respectively.

The following operator identities hold:

$$\begin{aligned}\sqrt{L+2} \Delta_0^{(L+1)} \Delta_+^{(L)} - \sqrt{L} \Delta_+^{(L)} \Delta_0^{(L)} &= 0, \\ \sqrt{L-1} \Delta_0^{(L-1)} \Delta_+^{(L)} - \sqrt{L+1} \Delta_-^{(L)} \Delta_0^{(L)} &= 0, \\ (L+1) \sqrt{2L-1} \Delta_+^{(L-1)} \Delta_-^{(L)} - 2\sqrt{2L+1} \Delta_0^{(L)} \Delta_0^{(L)} - L\sqrt{2L+3} \Delta_-^{(L+1)} \Delta_+^{(L)} &= 0.\end{aligned}\tag{14}$$

Using the above notation, we can describe a rather large class of thermodynamically compatible overdetermined system of equations in which the unknowns are transformed, under rotations of the coordinates, according to representations of the rotation group of some, generally speaking, arbitrary weights. In particular, this class contains the hydrodynamics equations and the equations of the elasticity theory written in the Lagrange coordinates.

Suppose that the unknown functions form vectors $q^{(M)}$, $u^{(L-1)}$, $v^{(L)}$, $w^{(L+1)}$, $p^{(L)}$ which, under rotations of coordinates, are transformed by irreducible representations of the corresponding weights indicated in the notation. Let $\mathcal{L} = \mathcal{L}(q^{(M)}, u^{(L-1)}, v^{(L)}, w^{(L+1)}, p^{(L)})$ be a convex function of components of vectors such that this function remains constant when vectors are transformed under rotations of the coordinates. This function \mathcal{L} plays role of the generating thermodynamic potential for the above equations. The system consists of the basis equations (a_L , b_L , and c_L are some constant coefficients)

$$\begin{aligned}\frac{\partial \mathcal{L}_{q^{(M)}}}{\partial t} &= 0, \\ \frac{\partial}{\partial t} \mathcal{L}_{u^{(L-1)}} + a_L [\nabla^{(1)} \times p^{(L)}]^{(L-1)} &= 0, \\ \frac{\partial}{\partial t} \mathcal{L}_{v^{(L)}} + b_L [\nabla^{(1)} \times p^{(L)}]^{(L)} &= 0, \\ \frac{\partial}{\partial t} \mathcal{L}_{w^{(L+1)}} + c_L [\nabla^{(1)} \times p^{(L)}]^{(L+1)} &= 0,\end{aligned}\tag{15}$$

$$\begin{aligned} \frac{\partial}{\partial t} \mathcal{L}_{p^{(L)}} + \sqrt{\frac{2L-1}{2L+1}} a_L [\nabla^{(1)} \times u^{(L-1)}]^{(L)} - b_L [\nabla^{(1)} \times v^{(L)}]^{(L)} \\ + \sqrt{\frac{2L+3}{2L+1}} c_L [\nabla^{(1)} \times w^{(L)}]^{(L)} = 0, \end{aligned}$$

the following compatible additional relations, which follow from (14):

$$\begin{aligned} \sqrt{L-1} b_L [\nabla^1 \times \mathcal{L}_{u^{(L-1)}}]^{(L-1)} - \sqrt{L+1} a_L [\nabla^1 \times \mathcal{L}_{v^{(L)}}]^{(L-1)} = 0, \\ \sqrt{L+2} b_L [\nabla^1 \times \mathcal{L}_{w^{(L+1)}}]^{(L+1)} - \sqrt{L} c_L [\nabla^1 \times \mathcal{L}_{v^{(L)}}]^{(L+1)} = 0, \quad (16) \\ (L+1)\sqrt{2L-1} \ b_L c_L [\nabla^{(1)} \times \mathcal{L}_{u^{(L-1)}}]^{(L)} - 2\sqrt{2L+1} \ a_L c_L [\nabla^{(1)} \times \mathcal{L}_{v^{(L)}}]^{(L)} \\ - L\sqrt{2L+3} \ a_L b_L [\nabla^{(1)} \times \mathcal{L}_{w^{(L+1)}}]^{(L)} = 0. \end{aligned}$$

and the conservation law (an analog of the thermodynamical identity)

$$\begin{aligned} \frac{\partial}{\partial t} [q^{(M)} \mathcal{L}_{q^{(M)}} + p^{(L)} \mathcal{L}_{p^{(L)}} + u^{(L-1)} \mathcal{L}_{u^{(L-1)}} + v^{(L)} \mathcal{L}_{v^{(L)}} + w^{(L+1)} \mathcal{L}_{w^{(L+1)}} - \mathcal{L}] \\ + \sqrt{2L-1} \ a_L [\nabla^{(1)} \times [p^{(L)} \times u^{(L-1)}]^{(1)}]^{(0)} + \sqrt{2L+1} \ b_L [\nabla^{(1)} \times [p^{(L)} \times v^{(L)}]^{(1)}]^{(0)} \\ + \sqrt{2L+3} \ c_L [\nabla^{(1)} \times [p^{(L)} \times w^{(L+1)}]^{(1)}]^{(0)} = 0. \quad (17) \end{aligned}$$

We use the following notation:

$$q^{(M)} \mathcal{L}_{q^{(M)}} \equiv \sum_m q_m^{(M)} \mathcal{L}_{q_m^{(M)}} \equiv \sqrt{2M+1} [q^{(M)} \times \mathcal{L}_{q^{(M)}}]^{(0)}.$$

Irreducible orthogonal representations of the rotation group are uniquely (up to an equivalence) defined by the dimension of the representation space, i.e. they can be written by the same matrices in the corresponding bases. Therefore, to justify the rules (14) and compute the Clebsh–Gordan coefficients, we can use some concrete realizations of the representation. In particular, it is possible to use the classic representations of rotations in the space of homogeneous harmonic polynomials in variables x_{-1}, x_0, x_1 . An orthonormal basis is usually formed by such polynomials $g_m^M(x_{-1}, x_0, x_1)$ of the form

$$\begin{aligned} g_{-|m|}^M &= \gamma_{-|m|}^M (x_{-1}^2 + x_0^2 + x_1^2)^M \sin m\varphi P_M^{|m|}(\cos \theta) \\ g_0^M &= \gamma_0^M (x_{-1}^2 + x_0^2 + x_1^2)^M P_M(\cos \theta) \quad (18) \\ g_{|m|}^M &= \gamma_m^M (x_{-1}^2 + x_0^2 + x_1^2)^M \cos m\varphi P_M^{|m|}(\cos \theta). \end{aligned}$$

Here $P_M(\mu)$, $P_M^{[m]}(\mu)$ are the Legendre polynomials, in particular, associated ones, $\cos \theta = x_0/(x_{-1}^2 + x_0^2 + x_1^2)^{1/2}$, $e^{i\varphi} = (x_{-1} + ix_1)/(x_{-1}^2 + x_1^2)^{1/2}$. The Clebsh–Gordan coefficients are uniquely (up to the sign) defined from the equalities

$$\begin{aligned} & q_k^K(x_{-1}, x_0, x_1) \cdot q_l^L(x_{-1}, x_0, x_1) \\ &= \sum_{M=|K-L|}^{K+L} \rho_{M[K,L]}(x_{-1}^2 + x_0^2 + x_1^2)^{K+L-M} C_{M[K,L]}^{m[k,l]} q_m^M(x_{-1}, x_0, x_1) \quad (19) \end{aligned}$$

and the normalization condition

$$\sum_{m=-M}^M \text{tr}[(C_M^m)^T C_M^m] = 1. \quad (20)$$

To justify the identities (14) and their consequence (16), it is convenient to use the unitary realization of representations of the group $SU(2)$ which is the universal covering of $SO(3)$. This realization has the form of transformations of spinor polynomials

$$P_N(\xi, \eta) = \sum_{n=-N}^{n=N} \alpha_n \frac{\xi^{N-n} \eta^{N+n}}{\sqrt{(N-n)!(N+n)!}} \quad (21)$$

induced by transformations from $SU(2)$ in two-dimensional complex space of vector with coordinates ξ, η . For representations of integer weight N in the space of spinor polynomials, we can choose a basis such that the matrices corresponding to representations do not differ from the matrices obtained by the above realization by harmonic polynomials.

In the spinor realization, the action of the operators $\Delta_{\pm}^{(L)}$ and $\Delta_0^{(L)}$ is equivalent to the action of the differentiation operators

$$\begin{aligned} \Delta_+^{(L)} f^{(L)}(\xi, \eta) &= \sqrt{\frac{2(2L+3)}{2L+1}} \left[-\frac{1}{2} \left(\frac{\partial}{\partial x_{-1}} + i \frac{\partial}{\partial x_1} \right) \eta^2 \right. \\ &\quad \left. - \frac{\partial}{\partial x_0} \xi \eta + \frac{1}{2} \left(\frac{\partial}{\partial x_{-1}} - i \frac{\partial}{\partial x_1} \right) \xi^2 \right] f^{(1)}(\xi, \eta), \\ \Delta_0^{(L)} f^{(L)}(\xi, \eta) &= -\frac{1}{\sqrt{L(L+1)}} \left[\frac{1}{2} \left(\frac{\partial}{\partial x_{-1}} + i \frac{\partial}{\partial x_1} \right) \eta \frac{\partial}{\partial \xi} \right. \\ &\quad \left. + \frac{1}{2} \frac{\partial}{\partial x_0} \left(\xi \left(\frac{\partial}{\partial \xi} - \eta \frac{\partial}{\partial \eta} \right) + \frac{1}{2} \left(\frac{\partial}{\partial x_{-1}} - i \frac{\partial}{\partial x_1} \right) \xi \frac{\partial}{\partial \eta} \right) \right] f^{(1)}(\xi, \eta), \quad (22) \end{aligned}$$

$$\Delta_-^{(L)} f^{(L)}(\xi, \eta) = \frac{\sqrt{2}}{L(2L+1)} = \left[\frac{1}{2} \left(\frac{\partial}{\partial x_{-1}} - i \frac{\partial}{\partial x_1} \right) \frac{\partial^2}{\partial \xi^2} \right]$$

$$+ \frac{\partial}{\partial x_0} \frac{\partial^2}{\partial \xi \partial \xi} - \frac{1}{2} \left(\frac{\partial}{\partial x_{-1}} + i \frac{\partial}{\partial x_1} \right) \frac{\partial^2}{\partial \xi^2}] f^{(1)}(\xi, \eta).$$

The proof of the identities (14) and (16) in the spinor realization is reduced to the use of the differentiation rules.

Regarding the justification and use of the spinor polynomial tool, we refer the reader to the book [5] for detail.

We give a new point of the result of [2]. We note that we used spinor polynomials only for the justification of the properties of symbols appeared in the form of equations (1.5)–(1.7), whereas the equations themselves do not explicitly written in [2] but are replaced with relations between generating functions. That is why, the mentioned article [2] is very hard for understanding.

Another approach to the imitation of equations in the Euler coordinates was proposed in [3]. However, it is also based on the tool of generating functions (spinor polynomials) and is very hard for understanding. It would be interesting to present the results of [3] in a similar way, as above.

To complete my report, I would like to emphasize once more that the systematization of equations of the classical mathematical physics, the work concerning the formulations of necessary thermodynamical restrictions that can be taken into account do not have the final form. In my opinion, the study of these question will lead to results that will be very important in the development of numerical methods of solutions of physical problems, as well as for statements of the theoretical problems of mathematics.

As was mentioned, my work (about 40 years ago) with numerical methods for gasdynamical computations was of consequence was a reason leading to the consideration of equations of the form

$$\frac{\partial L_{q_i}}{\partial t} + \frac{\partial M_{q_i}}{\partial x} = \frac{\partial}{\partial x} b_{ik} \frac{\partial q_k}{\partial x}$$

and the construction of examples in which dissipative terms unexpectedly affect the structure of solutions.

In my opinion, new approaches appearing in the connection with the systematization of a number of examples of concrete equations and the analysis of the results obtained from the point of view of the representation theory will lead to important and useful conclusions. I would like to attract attention of specialists to these open questions.

References

1. Godunov S. K. (1960) On the concept of generalized solution. *Soviet. Math. Dokl.*, Vol. 1, pp. 1194–1196, 1960.
2. Godunov S. K. and Romenskii E. I. (1998) Elements of Continuous Media and Conservation Laws. *Universitetskaya Seriya*. Vol. 4. Scientific Books Publisher. Novosibirsk. 1998 [in Russian].

3. Godunov S. K., Romenskii E. I., and Mihailova T. Yu. (1996) Systems of thermodynamically coordinated laws of conservation invariant under rotations. *Siberian Math. J.* Vol. 37. **4**. pp. 690-705. 1996.
4. Mihailova T. Yu. (1997) Thermodynamically coordinated laws of conservation with unknowns of arbitrary weight. *Siberian Math. J.* Vol. 38. **3**. 1997.
5. Godunov S. K. and Mihailova T. Yu. (1998) Representation of the Rotation Groups and Spherical Functions. *Universitetskaya Seriya*. Vol. 3. Scientific Books Publisher. Novosibirsk. 1998 [in Russian].

A NEW LIMITER THAT IMPROVES TVD-MUSCL SCHEMES

L. GOZALO

*Department of numerical simulation on aeroacoustic,
ONERA,
29, av de la division Leclerc, 92322 Châtillon cedex
Email: gozalo@onera.fr*

AND

R. ABGRALL

*Department of Applied Mathematics,
University of Bordeaux, France.
Email: Remi.Abgrall@math.u-bordeaux.fr*

Abstract.

In order to compute unsteady compressible flows with shocks or strong discontinuities, an improvement of the TVD-MUSCL approach is needed. Two methods are proposed : the “triad” approach that switches from one limiter to another according to the local behaviour of the solution, and a new limiter that annihilates as many error terms in the Equivalent System as possible. Some results on a 1-D and a 2-D test cases are presented.

1. Introduction

Research on numerical schemes for compressible unsteady flows, especially problems involving shocks, shear layers, or any strong discontinuities, is of great interest. In many unsteady applications, the flow physics is so complex so that only numerical predictions are reachable, therefore accurate schemes are mandatory. However, many existing high resolution schemes (Yee, 1989) are disappointing because they have a tendency to smear too much extrema. Some schemes have already been constructed to overcome this difficulty, such as ENO and WENO, but they need too many computer resources to be competitive for industrial applications.

The starting point of the present work is a second order TVD-MUSCL type scheme. Two methods are investigated : first, a switch between carefully chosen limiters depending on the local structure of the solution; second a new limiter designed from a study of the equivalent equations. Some results on two relevant test cases are then displayed.

2. The “triad” approach

We solve the Euler equations. The problem is first simplified to a 1-D scalar transport equation. It is discretized with a MUSCL type approach :

$$\begin{cases} \frac{u_i^{n+1} - u_i^n}{\Delta t} + \frac{1}{\Delta x} \left(F_{i+\frac{1}{2}}(u_{i+\frac{1}{2}}^L, u_{i+\frac{1}{2}}^R) - F_{i-\frac{1}{2}}(u_{i-\frac{1}{2}}^L, u_{i-\frac{1}{2}}^R) \right) = 0 \\ u_i^n = u(i\Delta x, n\Delta t) \in \mathbb{R} \end{cases} \quad (1)$$

The “triad” limiter was an idea of Billet (Billet and Louedrin, 1999) that we sketch next. The starting scheme was Dubois’ second order TVD-MUSCL scheme (Dubois, 1991). It is known that some limiters can give excellent results in some configurations and are very disappointing in some others. For each time step, at each point i , a limiter φ_i is determined according to the variation of the solution on the five-points stencil, that is from point $i-2$ to $i+2$. The limiter at node i is φ_1 , at node $i-1$: φ_2^L and at node $i+1$: φ_2^R . Then,

$$\begin{cases} u_{i+1/2}^R = u_{i+1} - \varphi_2^R \left(\frac{1}{r_{i+1}} \right) \frac{u_{i+1} - u_i}{2} \\ u_{i+1/2}^L = u_i + \varphi_1(r_i) \frac{u_{i+1} - u_i}{2} \\ u_{i-1/2}^R = u_i - \varphi_1 \left(\frac{1}{r_i} \right) \frac{u_i - u_{i-1}}{2} \\ u_{i-1/2}^L = u_{i-1} + \varphi_2^L(r_{i-1}) \frac{u_i - u_{i-1}}{2} \end{cases} \quad (2)$$

There are eight possibilities since there are at most three local extrema on $[i-2, .., i+2]$. They can be reduced into six using symmetry considerations on points $i-1$ and $i+1$. They are detailed on Figure 1.

The case the stencil fits in is determined by the sign of the slopes on $i-1$, i and $i+1$. We impose that $\varphi \equiv 0$ if $r < 0$, thus φ_1 is automatically zero in cases 4, 5 and 6. Case 1 describes a locally monotonic solution. In that case, we choose the limiter that maximizes the order of interpolation (Van Leer’s k -limiter with $k = 4$ (Anderson et. al., 1986) or minmod function with $\eta = 1/3$ and $\omega = 4$ (Yee, 1989)). Cases 2 and 3 may be seen as transitional cases before or after an extremum. Here, to avoid too much dissipation, we

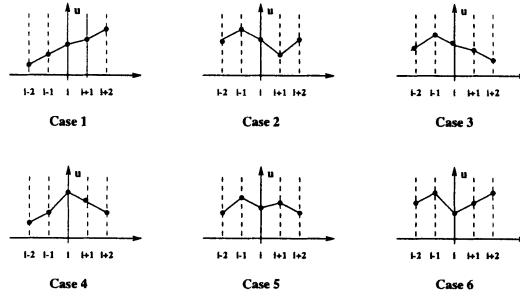


Figure 1. Possible cases of local behaviour for the discrete solution

consider a central extrapolation (or Van Leer's k -limiter with $k = 1$), i.e : $\varphi_1 \equiv 1$ ($r > 0$).

The scheme can be summarized by :

$$\left\{ \begin{array}{l} r_i \geq 0, r_{i-1} \geq 0, r_{i+1} \geq 0 \text{ (case 1)} \\ r_i \geq 0, \begin{cases} r_{i-1} \leq 0 \\ \text{and/or} \\ r_{i+1} \leq 0 \end{cases} \text{ (cases 2 - 3)} \\ r_i \leq 0 \text{ (cases 4 - 5 - 6)} \end{array} \right. \quad \left\{ \begin{array}{l} u_{i+\frac{1}{2}}^L = u_i + \varphi_{\frac{1}{3}}(r_i) \frac{u_{i+1} - u_i}{2} \\ u_{i-\frac{1}{2}}^R = u_i - \varphi_{\frac{1}{3}}(\frac{1}{r_i}) \frac{u_i - u_{i-1}}{2} \\ u_{i+\frac{1}{2}}^L = u_i + \varphi_1(r_i) \frac{u_{i+1} - u_i}{2} \\ u_{i-\frac{1}{2}}^R = u_i - \varphi_1(\frac{1}{r_i}) \frac{u_i - u_{i-1}}{2} \\ u_{i+\frac{1}{2}}^L = u_i \\ u_{i-\frac{1}{2}}^R = u_i \end{array} \right.$$

where $\varphi_k(r) = \frac{1}{2} \left[(1-k) \min(r, \frac{3-k}{1-k}) + (1+k) \min(1, \frac{3-k}{1-k} r) \right]$ which verifies monotonicity and convexity properties (Dubois, 1991)

3. A new limiter

Following once more Billet (Billet and Louedrin, 1999), we can construct a single limiter by studying the equivalent system of equations, with the aim of minimizing the error.

The spatial error term ϵ is given by :

$$\epsilon(i, .) = (f_x(u))(i\Delta x, .) - \frac{1}{\Delta x} \left(F_{i+\frac{1}{2}}(u_{i+\frac{1}{2}}^L, u_{i+\frac{1}{2}}^R) - F_{i-\frac{1}{2}}(u_{i-\frac{1}{2}}^L, u_{i-\frac{1}{2}}^R) \right) \quad (3)$$

To obtain the Taylor expansion of $\frac{F_{i+\frac{1}{2}} - F_{i-\frac{1}{2}}}{\Delta x}$ it is necessary to compute the expansions of r_{i-1} , r_i , r_{i+1} , then those of $\varphi(\frac{1}{r_{i+1}})$, $\varphi(r_i)$, $\varphi(\frac{1}{r_i})$, $\varphi(r_{i-1})$ around x_i . Finally, one obtains :

– for $u_x(i) \neq 0$:

$$\frac{F_{i+\frac{1}{2}} - F_{i-\frac{1}{2}}}{\Delta x} = F_u u_x + \Delta x^2 [a u_{xxx} \underbrace{(1 - 3\varphi'(1))}_{coeff_1} + b u_x u_{xx} \underbrace{(1 - 2\varphi'(1)) + c u_x^3}_{coeff_2} + O(\Delta x^3)] \quad (4)$$

– and for $u_x(i) = 0$:

$$\begin{aligned} \frac{F_{i+\frac{1}{2}} - F_{i-\frac{1}{2}}}{\Delta x} &= \Delta x \tilde{a} \underbrace{(2 - \varphi(-1) - \varphi(3))}_{coeff_3} u_{xx} \\ &\quad + \Delta x^2 \tilde{b} \underbrace{\left(1 + \frac{1}{2}\varphi(-1) - \varphi(3) - 2\varphi'(3) + \varphi'(-1)\right)}_{coeff_4} u_{xxx} \\ &\quad + O(\Delta x^3) \end{aligned} \quad (5)$$

with a , b , c , \tilde{a} , \tilde{b} coefficients depending on the derivatives of F with respect to u^L , u^R .

As Dubois (Dubois, 1991) mentionned, it is impossible to have better than second-order accuracy in space with a classical limiter that verifies : $\varphi(r) = 0$, $\forall r < 0$. Yet he proved that for any φ in the gray domain of the Figure 2 with parameter α , $1 < \alpha < 2$, the scheme is T.V.D. For any $\alpha < 2$, the domain allows non zero values for φ when $r < 0$. So the thing to do is to find a limiter φ which would both annihilate as many terms in (4) and (5) as possible and verify the constraints defined by the TVD diagram of the Figure 2.

When $1 < \alpha < 2$, it is easy to verify that the function in bold lines on Figure 2 cancels the $coeff_1$ in (4) and the $coeff_3$ and $coeff_4$ terms in (5). It is also possible to cancel the $coeff_2$ term instead of $coeff_1$ in (4), but the latter is a purely dissipative error term and is therefore more interesting to remove.

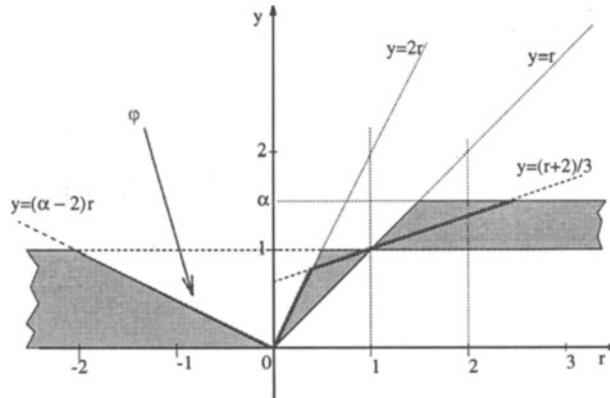


Figure 2. TVD area for φ and new family of limiters

4. Numerical results

4.1. SHU AND OSHER TEST CASE

We test both methods on Shu and Osher sine wave test case (Shu and Osher, 1989). The initial conditions are :

$$\begin{aligned} x < -4 \quad \rho &= 3.857143; & v &= 2.629369; & P &= 10.33333 \\ x \geq -4 \quad \rho &= 1 + 0.2 \sin 5x; & v &= 0; & P &= 1 \end{aligned} \quad (6)$$

Here we use Roe's flux instead of AUSM as in ref. (Billet and Louedin, 1999). It is interesting to note that the triad approach gives similar results with the different splitting methods.

From Figure 3, it appears that the triad approach is far less dissipative than any other method. The $k = 1/3$ limiter is the most compressive among the classical limiters employed in the TVD-MUSCL approach, but in this test case, the new limiter gives almost as accurate a result.

4.2. SHOCK/VORTEX INTERACTION

We now test our methods on a 2-D viscous shock/vortex interaction case (Tenaud et. al., to appear). The purpose of this test is to show the ability of the scheme to capture the acoustic wave generated by the interaction.

On this rather difficult test case (see Figure 4), we show that the new limiter, though compressive enough as verified on the previous test, is not too much compressive, unlike the $k = 1/3$ limiter. It is also very good at

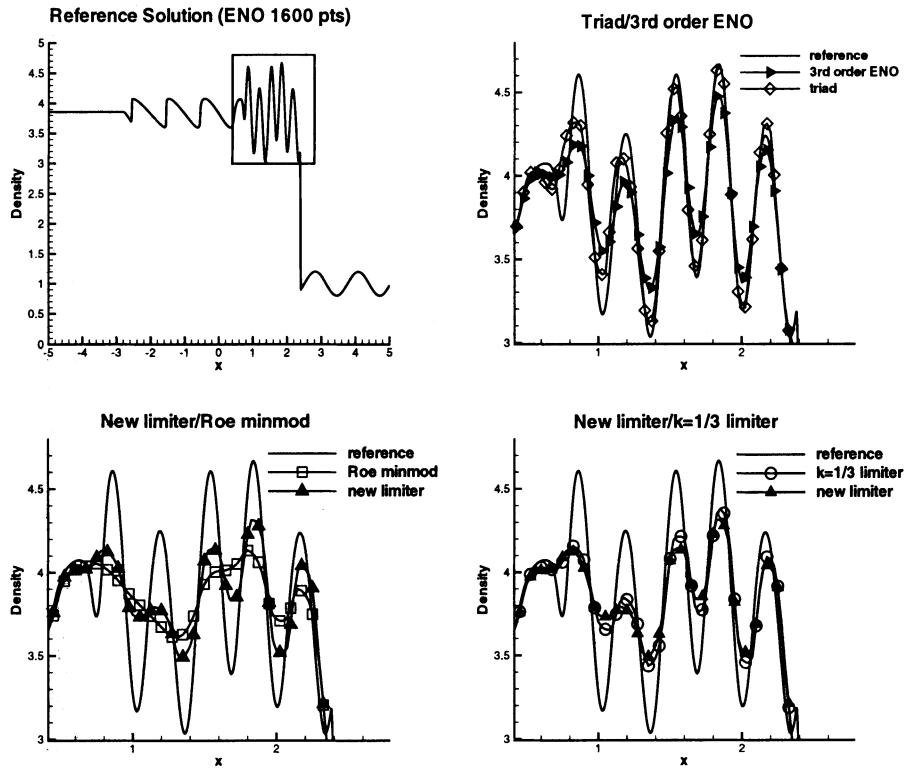


Figure 3. Shu and Osher test case : comparison of results with different methods at $t=1.8$ on a 400 points grid.

capturing the acoustic wave. As for the triad approach, though its compressiveness proved in the former case, it is also able to capture the acoustic wave, with better accuracy than most conventional limiters.

5. Conclusion

Both methods presented here reach their aim : an improvement of the TVD-MUSCL approach. The triad approach gives remarkable results, especially on the first case. Considering the new limiter, though parameter dependent which makes it also flow dependent (Yee, 1989), it appears that with the α parameter fixed near 1, the limiter seems very appropriate for any acoustic problem. This has already been verified on two others 2-D viscous cases, but it still needs to be tested further.

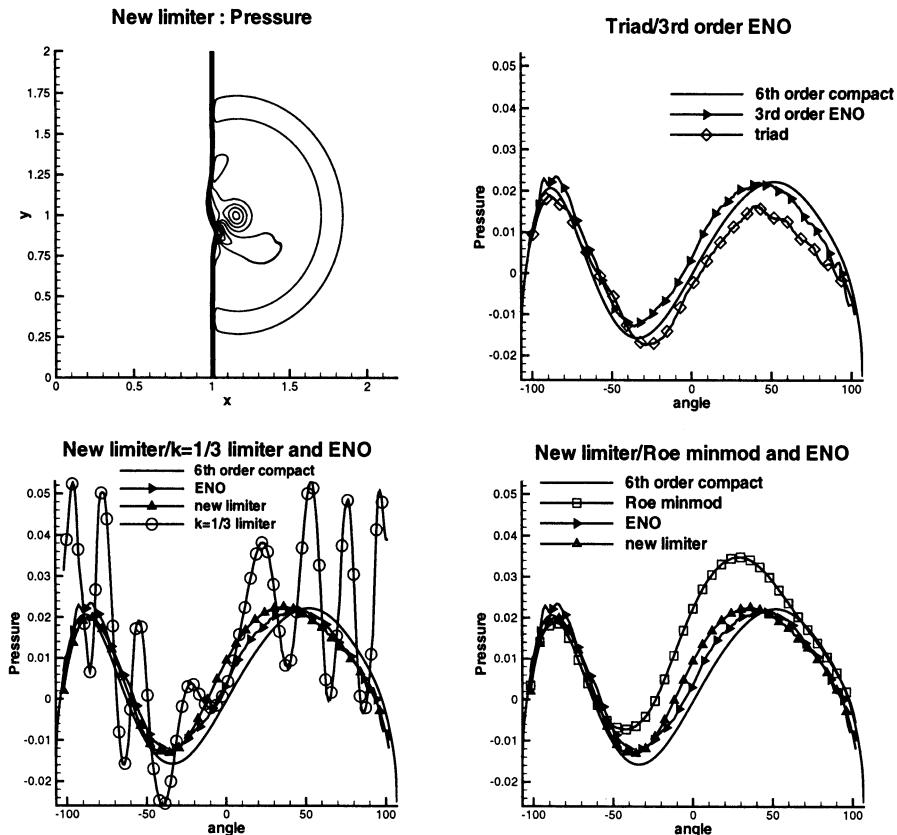


Figure 4. Shock/vortex interaction: 2-D field of pressure with new limiter and comparison of results over circumferencial profiles of pressure with different methods at $t=0.7$ on a 200×200 points grid. Reference solution : 6th order compact scheme on a 800×500 points grid.

References

- Anderson W K, Thomas J L and Van Leer B (1986). Comparison of Finite Volume Flux Vector Splittings for the Euler Equations. *AIAA J.* **24**, pp 1453–1458.
- Billet G and Louedon O (1999). A Simple Algorithm to Improve the Accuracy of TVD-MUSCL Schemes. *Int. Ser. Num. Math.* **129**, pp 65–75. Birkhäuser Verlag.
- Dubois F (1991). Nonlinear Interpolation and Total Variation Diminishing Schemes. Proc. Third International Conference on Hyperbolic Problems. studentlitteratur, pp 351–359.
- Shu C-W and Osher S (1989). Efficient implementation of essentially non-oscillatory shock-capturing schemes, II. *J. Comput. Phys.* **83**, pp 32–78.
- Tenaud C, Garnier E and Sagaut P (to appear). Evaluation of some high-order shock capturing schemes for the direct numerical simulation of unsteady, 2D free flows. *Int. J. Numer. Methods Fluids*.
- Yee H C (1989). A Class of High Resolution Explicit and Implicit Shock-Capturing Methods. NASA Technical memorandum N° 10 1088.

EXACT ROE LINEARIZATION FOR VAN DER WAALS' GAS

A. GUARDONE AND L. QUARTAPELLE

Dipartimento di Ingegneria Aerospaziale

Politecnico di Milano

Via La Masa 34, 20158 Milano, Italy

Email: guardone@aero.polimi.it

Abstract. An extension of Roe's method to the van der Waals gas is presented. The introduction of a convenient supplementary equation in the Roe linearization problem for the intermediate state allows to evaluate the velocity and total specific enthalpy separately from the determination of the density that is needed to compute the Roe matrix when thermodynamic equations of state different from those of the ideal gas are considered.

1. Introduction

The computation of compressible flows under the van der Waals equations of state is a very active research area in the study of dense gas flows near the critical point. In particular, the van der Waals gas is often taken as a simple model of BZT (Bethe, Zel'dovich and Thompson) fluid to investigate the behaviour of negative shock waves and of other phenomena related to negative nonlinearity. Although predicted theoretically in the early 1940's, negative nonlinearities are believed to have been observed experimentally for the first time only in 1983 by Borisov and collaborators (Borisov et al., 1983).

In the framework of Godunov-type schemes, different approaches have been followed to compute this kind of flows: for instance, Letellier and Forestier (1993) solved the exact Riemann problem near the critical point for the water vapour, for which only classical shocks are admissibles, while Argrow (1996) used a predictor-corrector scheme based on the Davis–Roe flux limited method without introducing any numerical approximation of the thermodynamic properties. Rider and Bates (2000) developed a Riemann solver with an explicit treatment of the non convexity of isentropic

curves. In the present work, a new method is presented that extends the well known upwind scheme proposed by Roe for the ideal polytropic gas to the (polytropic) van der Waals gas, without introducing any linear approximation for the equations of state of the gas.

2. An equation for the intermediate density

As well known, in Roe's scheme the Riemann problem associated with the left and right states \mathbf{w}_ℓ and \mathbf{w}_r , $\mathbf{w} \in \mathbb{R}^p$, is approximated by a linearized counterpart; the linearization is accomplished by a suitable matrix $\mathbf{A}^*(\mathbf{w}_\ell, \mathbf{w}_r)$ of dimension $p \times p$ defined as follows (Roe, 1981):

Matrix $\mathbf{A}^(\mathbf{w}_\ell, \mathbf{w}_r)$ is called a **Roe linearization** of the hyperbolic system with flux $\mathbf{f}(\mathbf{w})$ and Jacobian matrix $\mathbf{A}(\mathbf{w}) = \partial \mathbf{f}(\mathbf{w}) / \partial \mathbf{w}$ if $(\mathbf{w}_\ell, \mathbf{w}_r) \rightarrow \mathbf{A}^*(\mathbf{w}_\ell, \mathbf{w}_r)$ is a mapping from $\mathbb{R}^p \times \mathbb{R}^p$ into the set of $p \times p$ matrices with the following properties*

- i) *Conservation:* $\mathbf{A}^*(\mathbf{w}_\ell, \mathbf{w}_r)(\mathbf{w}_r - \mathbf{w}_\ell) = \mathbf{f}(\mathbf{w}_r) - \mathbf{f}(\mathbf{w}_\ell)$,
- ii) *Hyperbolicity:* $\mathbf{A}^*(\mathbf{w}_\ell, \mathbf{w}_r)$ has real eigenvalues and a corresponding set of eigenvectors that form a basis of \mathbb{R}^p ,
- iii) *Consistency:* $\mathbf{A}^*(\mathbf{w}_\ell, \mathbf{w}_r) \rightarrow \mathbf{A}(\mathbf{w})$ as \mathbf{w}_ℓ and $\mathbf{w}_r \rightarrow \mathbf{w}$.

A standard assumption in Roe's linearization problem is choosing \mathbf{A}^* as the Jacobian matrix $\mathbf{A}(\mathbf{w}) = \partial \mathbf{f}(\mathbf{w}) / \partial \mathbf{w}$ to be evaluated in an *intermediate state* $\tilde{\mathbf{w}} = \tilde{\mathbf{w}}(\mathbf{w}_\ell, \mathbf{w}_r)$ obtained from condition i)

$$\mathbf{A}(\tilde{\mathbf{w}}(\mathbf{w}_\ell, \mathbf{w}_r)) (\mathbf{w}_r - \mathbf{w}_\ell) = \mathbf{f}(\mathbf{w}_r) - \mathbf{f}(\mathbf{w}_\ell), \quad (1)$$

of p equations in the p unknowns \tilde{w}_i , $1 \leq i \leq p$. The Jacobian form assures that the qualitative condition ii) is automatically fulfilled and implies that condition iii) assumes the form $\tilde{\mathbf{w}}(\mathbf{w}_\ell, \mathbf{w}_r) \rightarrow \mathbf{w}$ as \mathbf{w}_ℓ and $\mathbf{w}_r \rightarrow \mathbf{w}$.

Let us now consider the one-dimensional Euler equations, with \mathbf{w} representing the vector of the conservative variables (ρ, m, E^t) density, momentum density and total energy per unit volume, so that $\mathbf{f}(\mathbf{w}) = (m, m^2/\rho + P, (E^t + P)m/\rho)$. In this case, the first equation is identically satisfied and system (1) reduces to two equations, so that the intermediate state is actually a one parameter family of solutions.

Considering the particular case of an *ideal* gas [$P = \rho RT$], the flux $\mathbf{f}(\mathbf{w})$ is a function homogeneous of degree one with respect to \mathbf{w} and the unknown of system (1) can be chosen to be a vector with only two components, the usual choice being the velocity and total enthalpy per unit mass (u, h^t) . In this case for any equation of state $P(E, \rho)$, E denoting the internal energy per unit volume, that is compatible (up to a linear function in ρ) with the assumption of gas ideality, it turns out that the intermediate state (\tilde{u}, \tilde{h}^t) is uniquely determined as solution of the two equations system. Under

the further assumption of a *polytropic* gas [$E = \rho RT/(\gamma - 1)$], $P(E, \rho) = (\gamma - 1)E$ and the unique solution is provided by the celebrated average (Roe, 1981):

$$\tilde{u} = \frac{\sqrt{\rho_\ell} u_\ell + \sqrt{\rho_r} u_r}{\sqrt{\rho_\ell} + \sqrt{\rho_r}}, \quad \tilde{h}^t = \frac{\sqrt{\rho_\ell} h_\ell^t + \sqrt{\rho_r} h_r^t}{\sqrt{\rho_\ell} + \sqrt{\rho_r}}. \quad (2)$$

On the contrary, for a *nonpolytropic* ideal gas, i.e., an ideal gas with a nonlinear function $e = e(T)$, the intermediate state is still defined uniquely as the solution of system (1), but in a way that cannot be reduced to the Roe average (2) and depends on the function $e(T)$.

Coming now to the nonideal gases of interest here, we propose to exploit the available degree of freedom of the one-parameter family by fixing the intermediate state so as to reduce the complexity of the nontrivial part of system (1). This can be achieved by augmenting the system with the introduction of the *supplementary equation*

$$\nabla_w \Pi(\tilde{\mathbf{w}}) \cdot \Delta \mathbf{w} = \Delta P, \quad (3)$$

where $\Pi(\mathbf{w}) = P(E^t - m^2/(2\rho), \rho)$ and $\Delta(\cdot) = (\cdot)_r - (\cdot)_\ell$. Within the one-parameter family of intermediate states, equation (3) selects a unique $\tilde{\mathbf{w}}$ and, at the same time, uncouples the determination of the *density* $\tilde{\rho}$ from that of the other two unknowns. In fact, the latter can be expressed by the aforementioned Roe's solution through the change of variable $\mathbf{w} = (\rho, m, E^t) \rightarrow \mathbf{v} = (\rho, u, h^t)$.

In conclusion, for any gas different from the ideal gas¹, the solution of the augmented linearization problem is obtained from (2) and from the subsequent solution of the uncoupled equation

$$\nabla_w \Pi(\mathbf{w}(\tilde{\rho}, \tilde{u}, \tilde{h}^t)) \cdot \Delta \mathbf{w} = \Delta P, \quad (4)$$

for the single unknown $\tilde{\rho}$. Note that in the present method no averaging of the pressure derivatives is introduced, since their *analytical expressions* are evaluated exactly at the intermediate state.

3. Intermediate density for the van der Waals gas

Let us now consider the polytropic van der Waals model (Callen, 1985) defined by the equations of state

$$P(T, \rho) = \frac{RT\rho}{1 - b\rho} - a\rho^2 \quad \text{and} \quad e(T, \rho) = \frac{R}{\delta} T - a\rho, \quad (5)$$

¹For the ideal polytropic gas, the proposed supplementary equation is a truism within the Roe linearization problem, since it is automatically satisfied as a consequence of the fulfillment of the momentum equation of Roe's system.

where e is the internal energy per unit mass, R is a (gas dependent) constant, a and b are the van der Waals constants and $\delta \equiv R/c_v$ is constant.

For the van der Waals gas, the supplementary equation (4) is found to be a third order polynomial in the dimensionless intermediate density $r = \tilde{\rho}/\rho_c = 3b\tilde{\rho}$, in the form

$$r^3 + Ar^2 + Br + C = 0, \quad (6)$$

with coefficients defined in terms of the left and right states as follows:

$$\begin{aligned} A &= \frac{1}{6} \frac{\rho_c}{P_c} \frac{\Delta P}{\Delta \rho} - 3(2 + \delta), \\ B &= \frac{1}{2} \frac{\rho_c}{P_c} \left[-(2 + \delta) \frac{\Delta P}{\Delta \rho} + \delta \left(\frac{\Delta E}{\Delta \rho} - \tilde{h}^t + \frac{\tilde{u}^2}{2} \right) \right] + 9(1 - \delta^2), \\ C &= \frac{3}{2} \frac{\rho_c}{P_c} (1 + \delta) \left[\frac{\Delta P}{\Delta \rho} - \delta \frac{\Delta E}{\Delta \rho} \right]. \end{aligned} \quad (7)$$

Therefore, the supplementary equation (4) can be solved analytically by standard formulas. In the particular case $\Delta \rho = 0$, (4) is linear in $\tilde{\rho}$ and gives $r = 3 - 3\delta(\Delta E/\Delta P)$, the case $\Delta \rho = 0$ and $\Delta P = 0$ being trivial since one has $\Delta \rho = \Delta P = 0 \Rightarrow \Delta E = 0$. We notice that in the limit $b \rightarrow 0$ the supplementary equation reduces to a linear equation in $\tilde{\rho}$ which gives $\tilde{\rho} = (\rho_\ell + \rho_r)/2$.

4. Numerical results

For the numerical tests, we have considered two reference shock-tube problems (Argrow, 1996). The diaphragm is located at $x = 0.5$ and separates the following constant initial states, made dimensionless by critical values:

	ρ_ℓ	u_ℓ	P_ℓ	ρ_r	u_r	P_r
Case 1	0.879	0	1.09	0.562	0	0.885
Case 2	0.879	0	1.09	0.275	0	0.575

These two problems explore the nonclassical behaviour of a dense gas in the presence of negative nonlinearity, which is associated with the negative values of the fundamental derivative

$$\Gamma = -\frac{v}{2} \left(\frac{\partial^2 P}{\partial v^2} \right)_s \Big/ \left(\frac{\partial P}{\partial v} \right)_s, \quad (8)$$

where $v \equiv 1/\rho$. [For an ideal polytropic gas $\Gamma = (\gamma + 1)/2$.] Both tests are run with $\delta = 0.0125$, which indicates a fluid with a large specific heat; for instance, perfluorodecane, $C_{10}F_{22}$ ($\delta = 0.0132$), or fluorinated ether E-5, $C_{17}F_{35}HO_5$ ($\delta = 0.0074$).

The numerical solutions, reported in figures 1 and 2, are found in excellent agreement with the reference solutions (Argrow, 1996), thus demonstrating the validity of the proposed method. The computations have been performed by means of a high-resolution flux-limiter method which takes advantage of the proposed linearization near discontinuities and uses a Lax-Wendroff scheme in smooth flow regions, see, for instance, LeVeque (1992). We emphasize that this scheme has proven successful in all computations we have attempted, with the only exception of a test case of negative/positive transonic rarefaction that has required to replace LeVeque's entropy fix by the standard entropy fix of Harten. This difficulty is caused by the nonsimple character of the rarefaction waves in that Riemann problem.

Numerical results (not shown) for the water vapor near the critical point have been compared successfully with calculations reported in the literature (Letellier and Forestier, 1993).

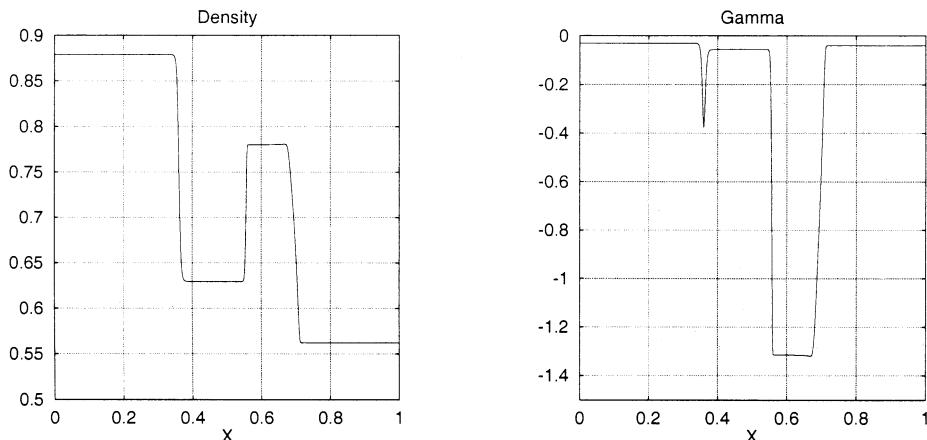


Figure 1. Numerical solution of the Riemann problem for test case 1. Dimensionless time $t^* \equiv (t/L)(P_c/\rho_c)^{1/2} = 0.25025$.

5. Conclusions

In the present work, the linearization procedure of Roe for the Euler equations has been extended from the ideal gas to a gas governed by the van der Waals equations of state. The proposed method assumes an intermediate state as the unknown of the linearization problem but, differently from standard procedures for the ideal gas, requires the determination of an *intermediate density* in addition to the intermediate velocity and total enthalpy of the original method (Roe, 1981). Such a density is needed to evaluate the eigenstructure of the Jacobian matrix, due to the nonideal form of the equations of state employed.

The originality of the proposed method lies in the introduction of a convenient supplementary condition which decouples the determination of

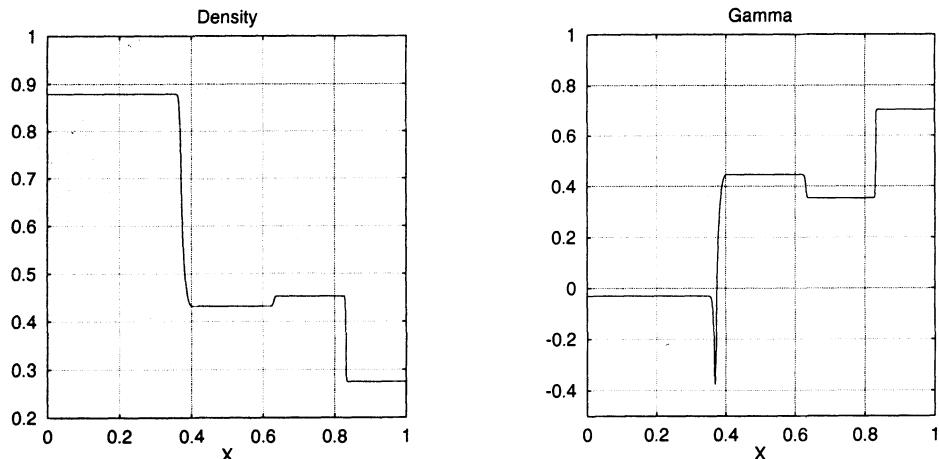


Figure 2. Numerical solution of the Riemann problem for test case 2. ($t^* = 0.2255$).

the intermediate velocity and enthalpy from the determination of the density. Thanks to the analytical form of the van der Waals thermodynamics, a third order algebraic equation for the intermediate density is obtained which gives directly the solution of the linearization problem in terms of the Roe averaged velocity and total enthalpy and of the jumps in the density, pressure and internal energy per unit volume.

By virtue of the segregation of all the aspects dependent on the thermodynamic equations of state into the equation for the intermediate density, the present method can be easily extended to deal with more complex physical systems such as, for instance, chemically reacting gases in local thermodynamic equilibrium.

References

- ARGROW B M (1996). Computational analysis of dense gas shock tube flow. *Shock Waves*, **6**, pp 241–248.
- BORISOV A A, BORISOV AL A, KUTATELADZE S S AND NAKARYAKOV V E (1983). Rarefaction shock waves near the critic liquid-vapour point. *J. Fluid. Mech.*, **126**, pp 59–73.
- CALLEN H B (1985). Thermodynamics and an Introduction to Thermostatistics, 2nd Ed. Wiley.
- GUARDONE A, SELMIN V AND VIGEVANO L (1999). An investigation of Roe's linearization and average for ideal and real gases. *Scientific Report DIA-SR 99-01*, Politecnico di Milano.
- LETELLIER A AND FORESTIER A (1993). Le problème de Riemann en fluide quelconque. *Rapport DMT/93.451 C.E.A.*, Direction de Réacteur Nucléaires.
- LEVEQUE R J (1992). Numerical Methods for Conservation Laws. Birkhäuser Verlag.
- RIDER W J AND BATES J W (2000). A high-resolution method Godunov method for modeling anomalous fluid behaviour. *Godunov Methods: Theory and Applications*. Toro E F Ed., Kluwer/Plenum Academic Publishers.
- ROE P L (1981). Approximate Riemann solvers, parameter vectors, and difference schemes. *J. Comput. Phys.*, **43**, pp 357–372.

THE GODUNOV-RYABENKII CONDITION: THE BEGINNING OF A NEW STABILITY THEORY

BERTIL GUSTAFSSON

Uppsala University, Sweden Email: bertil@tdb.uu.se

Dedicated to S.K. Godunov on the occasion of his 70th birthday

Abstract. The analysis of difference methods for initial-boundary value problems was difficult during the first years of the development of computational methods for PDE. The Fourier analysis was available, but of course not sufficient for non-periodic boundary conditions. The only other available practical tool was an eigenvalue analysis of the evolution difference operator Q . Actually, there were definitions presented, that defined an approximation as stable if the eigenvalues of Q were inside the unit circle for a fixed step-size h .

In the paper "Special criteria for stability for boundary-value problems for non-self-adjoint difference equations" by S.K. Godunov and V.S. Ryabenkii in 1963, the authors presented an analysis of a simple difference scheme that clearly demonstrated the shortcomings of the eigenvalue analysis. They also gave a new definition of the spectrum of a family of operators, and stated a new necessary stability criterion. This criterion later became known as the Godunov-Ryabenkii condition, and it was the first step towards a better understanding of initial-boundary value problems. The theory was later developed in a more general manner by Kreiss and others, leading to necessary and sufficient conditions for stability.

In this paper we shall present the contribution by Godunov and Ryabenkii, and show the connection to the general Kreiss theory.

1. Introduction

Half a century ago, some numerical computations of time-dependent PDE solutions had been done on the very few computers that existed. Most

methods were based on difference approximations, but very little theory was available. There was the early paper from 1928 by Courant, Friedrichs and Levy (Courant and Friedrichs and Levy, 1928), where the fundamental C-F-L condition was formulated. In the forties, the von Neumann theory was developed for the Cauchy problem and for problems with periodic solutions. However, for initial-boundary value problems no theory at all was available.

At the Moscow University there was a group of famous mathematicians, with I.M. Gelfand as one of the most prominent. This group was concerned with a great variety of mathematical problems spanning from the most abstract pure mathematical theory to very applied problems. One of the areas in applied mathematics that caught their interest, was the theory for difference approximations of partial differential equations, and in particular, the stability theory for initial-boundary value problems. The definition of stability could be stated as a straightforward generalization of the definition for the Cauchy problem. But the main question was how to find stability criteria that didn't lead to conditions that were impossible to apply for real world problems. If stability is defined by requiring that all powers of the evolution difference operator Q are bounded in norm, then it is a long way to find easily applicable sufficient conditions. The obvious condition $\|Q\| \leq 1$ is in most cases too restrictive, besides the fact that even this one is not trivial to check.

One natural way to go, is to investigate the eigenvalues of Q . For the Cauchy problem, these are easy to compute by considering the Fourier transform \hat{Q} . Even for the initial-boundary value problem, where the Fourier transform cannot be used, it is a much easier task to compute (or estimate) the eigenvalues than it is to compute the norm, in particular the norm of Q^n for all n . This is where the research was directed at this time in Moscow.

Gelfand was running a very active seminar, and in the early fifties, a very young bright student joined in. His name was Sergei Godunov; his talent had been demonstrated very clearly when he wrote his first scientific paper already at the age of nineteen. As well as Gelfand, he had early a very wide area of interest, and among other topics, he began taking interest in the stability theory. He learned a lot from the more senior researchers, and soon began to develop his own theories. Later he was joined by another young coworker Victor Ryabenkii, and together they laid the foundations for the development of a general theory for initial-boundary value problems.

In this paper we shall first present the contribution of Godunov and Ryabenkii, mainly as it was presented in (Godunov and Ryabenkii, 1963). We shall then show the connection to later work by H.-O. Kreiss and his coworkers, and give a short summary of the state of the theory of today.

2. The Godunov-Ryabenkii condition

We consider here linear difference schemes in its simplest form

$$u_j^{n+1} = Q u_j^n , \quad j = 0, 1, \dots, N , \quad (1)$$

where Q can be viewed as a matrix operating on the vector of grid-values u_j^n . It was well known at the time referred to above, that for a fixed value of the grid-size h , the norm of the solution tends to zero if all the eigenvalues satisfy

$$\lambda(Q) < 1 . \quad (2)$$

One possibility of defining stability was therefore to require (2), or the weaker condition

$$\lambda(Q) \leq 1 , \quad (3)$$

where the eigenvalues on the unit circle must be simple. However, in the Moscow seminar one was aware that this was not a good condition. If a certain computation with a fixed h is not accurate enough, one would like to have a better result with a smaller h . This is not necessarily the case under the condition (2), see Figure 1.

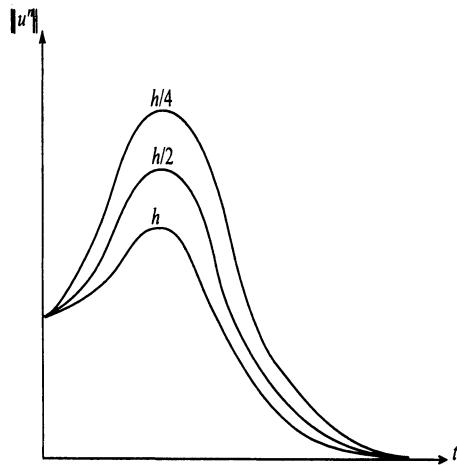


Figure 1. Norm of the solution for different step-size

That is where the difficulty enters: the constants in the estimates of the solution must be independent of h . In 1954 Godunov constructed the following very nice and simple counter-example that illustrated the essential point.

Consider the initial-boundary value problem

$$\begin{aligned} u_t + u_x &= 0, \quad 0 \leq x \leq 1, \quad 0 \leq t, \\ u(0, t) &= 0, \\ u(x, 0) &= f(x), \end{aligned}$$

and the difference approximation

$$\begin{aligned} u_j^{n+1} &= u_j^n - r(u_j^n - u_{j-1}^n), \quad r = \Delta t / h, \\ u_0^{n+1} &= 0, \\ u_j^0 &= f_j. \end{aligned}$$

(This is actually the later very famous original Godunov method when applied to this simple equation.) The corresponding evolution matrix Q is

$$Q = \begin{bmatrix} 0 & 0 & 0 & & 0 \\ r & 1-r & 0 & & 0 \\ 0 & r & 1-r & & \\ & & \ddots & \ddots & \\ & & & \ddots & \ddots \\ & & & & r & 1-r \end{bmatrix} \quad (4)$$

Obviously, there are only two eigenvalues $\lambda(Q) = 0, 1-r$, and the condition $|\lambda(Q)| \leq 1$ is therefore satisfied for

$$0 \leq r \leq 2. \quad (5)$$

However, the true stability condition is

$$0 \leq r \leq 1. \quad (6)$$

This condition follows from the C-F-L condition, that states that the domain of dependence for the differential equation must be contained in the domain of dependence for the difference scheme.

As mentioned above, there is no contradiction in the two different conditions. Even if the $\|u^n\|$ tends to zero as $n \rightarrow \infty$, there is no uniform bound on the norm if $1 < r \leq 2$.

The fundamental question for the Moscow team was how to predict the instability based on the eigenvalue distribution of Q . In 1963, the paper (Godunov and Ryabenkii, 1963) appeared, and the new concept of a *family of operators* Q_h was introduced. The new definition of the spectrum was given as

Definition 1 A point λ is a spectral point of $\{Q_h\}$ if for any $\epsilon > 0$ and $h_0 > 0$ we can give a number h , $h < h_0$, such that the inequality $\|Q_h u - \lambda u\| < \epsilon \|u\|$ has a solution u . We call the aggregate of all spectral points the spectrum of $\{Q_h\}$.

(The solution u will be called a quasi-eigenvector.)

The stability condition, that was later to be called the *Godunov-Ryabenkii condition*, was given as

Theorem 1 For the stability of a problem of the form

$$u^{n+1} = Q_h u^n , \quad n = 0, 1, \dots \quad (7)$$

it is necessary that the spectrum of $\{Q_h\}$ should lie in the unit disc.

For the example above, where the notation Q is retained for the difference operator, the quasi-eigenvectors have the form

$$\phi = [(s/r)^N \ (s/r)^{N-1} \ \dots \ 1]^T . \quad (8)$$

If $|s| < r$, then all λ with $\lambda = 1 - r + s$ belong to the spectrum.

Therefore the necessary Godunov-Ryabenkii condition for stability is

$$|1 - r + s| \leq 1, \quad |s| < r \quad (9)$$

which is equivalent to $r \leq 1$, see Figure 2.

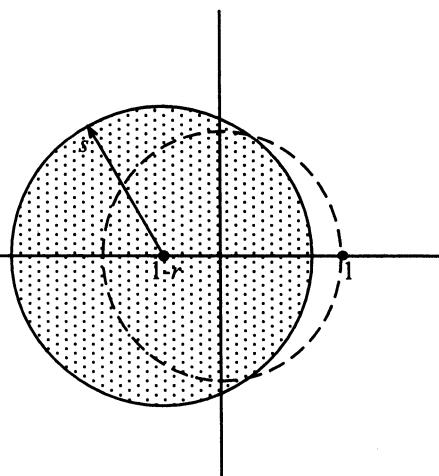


Figure 2. Location of spectrum

Note that the quasi-eigenvector satisfies the genuine eigenvalue/eigenvector criterion at all points except at $j = 0$, i.e.,

$$\begin{aligned}(Q\phi)_j &= \lambda\phi_j, \quad j = 1, 2, \dots, N, \\ (Q\phi)_0 &\neq \lambda\phi_0,\end{aligned}$$

since $\phi_0 = (s/r)^N \neq 0$. But for any s with $|s| < r$, $\phi_0 \rightarrow 0$ as $N \rightarrow \infty$, i.e., $h \rightarrow 0$, which shows that ϕ is a quasi-eigenvector.

In general, it is not so easy to find the spectrum of a family of operators Q_h . Godunov and Ryabenkii made use of an earlier observation by Gelfand, that a substantial simplification is obtained by partitioning the problem into three separate problems:

I. *The right quarter space problem*

$$u_j^{n+1} = Q_R u_j^n, \quad j = 0, 1, \dots \quad (10)$$

II. *The left quarter space problem*

$$u_j^{n+1} = Q_L u_j^n, \quad j = N, N-1, \dots \quad (11)$$

III. *The Cauchy problem*

$$u_j^{n+1} = Q_C u_j^n, \quad j = 0, \pm 1, \dots \quad (12)$$

In I, the boundary condition at $j = N$ is disregarded (if there is any), in II, the boundary condition at $j = 0$ is disregarded, and in III, both boundary conditions are disregarded.

The spectrum of each one of these problems should be investigated. For the example above, it turns out that Problem II determines the whole spectrum, and we begin with that one:

$$u_j^{n+1} = (1-r)u_j^n + ru_{j-1}^n, \quad j = N, N-1, \dots \quad (13)$$

The question is now if there is a nontrivial solution of the form $u_j^n = z^n \phi_j$, where $\|\phi\| < \infty$ and $|z| > 1$? This leads to the equation

$$z\phi_j = (1-r)\phi_j + r\phi_{j-1}, \quad j = N, N-1, \dots \quad (14)$$

There is no explicit h -dependence, and the solution is

$$\phi_j = \sigma \kappa^{j-N}, \quad \kappa = \frac{r}{z-1+r}, \quad (15)$$

where σ is a constant. The condition $\|\phi\| < \infty$ implies $|\kappa| > 1$, which is equivalent to

$$|z - 1 + r| < r. \quad (16)$$

Hence, nontrivial solutions for $|z| > 1$ exist if $r < 0$ or $r > 1$.

Problems I and III do not introduce any new spectral points, and therefore the Godunov-Ryabenkii condition is $0 \leq r \leq 1$.

This example illustrates that the analysis of these three problems separately is much easier than for the original one.

3. Sufficient conditions for stability

In this section we shall give a very brief review of the theory leading to sufficient conditions for stability, and we will emphasize the connection to the Godunov-Ryabenkii theory. We limit ourselves to approximations of hyperbolic first order systems, and discuss first the semi-discrete case. The approximation is

$$\begin{aligned} \frac{du_j}{dt} &= Qu_j + F_j, \quad j = 0, 1, \dots, N, \\ B_0 u_0 &= g_0, \\ B_N u_N &= g_N, \\ u_j(0) &= f_j. \end{aligned} \quad (17)$$

Here B_0 and B_N are boundary operators connecting neighbouring points, Q is a difference operator of the form

$$Q = \frac{1}{h} \sum_{\nu=-r}^p \alpha_\nu E^\nu, \quad Eu_j = u_{j+1}. \quad (18)$$

Without restriction we can assume that the matrix coefficients α_ν are independent of h , i.e., we are considering only the principal part of the difference operator.

In analogy with the discussion above, we partition the problem into the three different problems

I. The right quarter space problem:

$$\begin{aligned} \frac{du_j}{dt} &= Qu_j + F_j, \quad 0 \leq x_j < \infty, \\ B_0 u_0 &= g_0, \\ \|u\| &< \infty, \\ u_j(0) &= f_j. \end{aligned} \quad (19)$$

II. The left quarter space problem:

$$\begin{aligned}\frac{du_j}{dt} &= Qu_j + F_j, \quad -\infty < x_j \leq 1, \\ B_N u_N &= g_N, \\ \|u\| &< \infty, \\ u_j(0) &= f_j.\end{aligned}\tag{20}$$

III. The Cauchy problem:

$$\begin{aligned}\frac{du_j}{dt} &= Qu_j + F_j, \quad -\infty < x_j < \infty, \\ u_j(0) &= f_j.\end{aligned}\tag{21}$$

We keep the notation Q for the difference operator at inner points for all three problems, the boundary conditions are stated explicitly for problems I and II.

The norm is defined as the discrete l_2 -norm, which for the right quarter space problem is

$$\|u\| = \left(\sum_{j=0}^{\infty} |u_j|^2 h \right)^{1/2},\tag{22}$$

and similarly for the left quarter space problem. In what follows, we will mainly discuss the right quarter space problem.

Next we assume that $f = F = 0$. After Laplace transformation and multiplication by h in the main equation, we obtain with $\tilde{s} = sh$

$$\begin{aligned}\tilde{s}\hat{u}_j &= hQ\hat{u}_j, \quad j = 0, 1, \dots, \\ B_0\hat{u}_0 &= \hat{g}, \\ \|\hat{u}\| &< \infty.\end{aligned}\tag{23}$$

Considering \tilde{s} as a given parameter, this problem is independent of h . The Godunov-Ryabenkii condition was originally formulated for fully discrete problems; for the semi-discrete problem above it is

Lemma 1 *Consider the problem (23) with $\hat{g} = 0$. A necessary condition for stability of (19), and consequently of (17), is that there is no nontrivial solution \hat{u} for $\operatorname{Re} \tilde{s} > 0$.*

If a particular complex number \tilde{s} is not an eigenvalue, we can estimate the solution to the inhomogeneous problem (23) in terms of the boundary data. For any fixed index j we get the estimate

$$|\hat{u}_j| \leq K(\tilde{s})|\hat{g}|,\tag{24}$$

where the constant $K(\tilde{s})$ depends on \tilde{s} . Therefore we can formulate the Godunov-Ryabenkii condition in the following way:

Lemma 2 *The Godunov-Ryabenkii condition is satisfied if there is a unique solution to (23) that for every fixed j satisfies*

$$|\hat{u}_j| \leq K(\tilde{s})|\hat{g}| \text{ for all } \tilde{s} \text{ with } \operatorname{Re} \tilde{s} > 0. \quad (25)$$

Here $K(\tilde{s})$ is a constant that depends on \tilde{s} .

In order to derive sufficient conditions for stability, it is necessary to study the behavior of the solutions to (23) as \tilde{s} approaches the imaginary axis. This leads to

Definition 2 *The Kreiss condition is satisfied if (23) has a unique solution that for every fixed j satisfies the estimate*

$$|\hat{u}_j| \leq K|\hat{g}|, \quad \operatorname{Re} \tilde{s} > 0. \quad (26)$$

Here the constant K is independent of \tilde{s} .

One can prove that if the Cauchy problem is stable, then the roots κ of the characteristic equation

$$\operatorname{Det}(\tilde{s}I - \sum_{\nu=-r}^p \alpha_\nu \kappa^\nu) = 0 \quad (27)$$

split into two sets $|\kappa| < 1$ and $|\kappa| > 1$ for $\operatorname{Re} \tilde{s} > 0$. The solution \hat{u} is expressed in terms of powers κ^j of these roots (powers κ^{j-N} for the left quarter space problem). Therefore, the requirement of a bounded norm implies that the second set cannot be present in the general form of the solution to (23). It is important to note that the solution \hat{u} has this reduced form even when considering the Kreiss condition. Then we make the special analysis of the limit for this solution as $\operatorname{Re} \tilde{s}$ approaches zero.

Formulated like this, the connection between the Kreiss condition and the Godunov-Ryabenkii condition is very easy to describe: There is an estimate of the solution to (23) for all \tilde{s} in the right half of the complex plane in both cases. But the Kreiss condition requires the estimate to be uniform in \tilde{s} .

Different difference approximations have different stability properties, and it is necessary to work with more than one definition of stability. The strongest one is the following:

Definition 3 *The approximation (19) is strongly stable if there is a unique solution that satisfies*

$$\|u(t)\|^2 \leq K(\|v(0)\|^2 + \max_{0 \leq \tau \leq t} \|F(\tau)\|^2 + \max_{0 \leq \tau \leq t} |g(\tau)|^2). \quad (28)$$

The same definition can of course be made also for the left quarter space problem as well as for the original approximation with two boundaries. We just change the domain of the summation index j in the definition of the norm in (22).

A difference operator Q is semi-bounded if

$$(u, Qu) \leq \alpha \|u\|^2, \quad (29)$$

for all grid-functions u satisfying the boundary conditions (if there are any). We have

Theorem 2 *Assume that Q is semi-bounded for the Cauchy problem (21). Then the approximation (19) is strongly stable if $r \geq p$ and if the Kreiss condition is satisfied.*

If we want to apply the result for the original two-point boundary value problem (17), then the case $r = p$ is the only interesting one. We have

Theorem 3 *Assume that Q is semi-bounded for the Cauchy problem (21) and that $r = p$. Then the approximation (17) is strongly stable if the Kreiss condition is satisfied for both quarter space problems (19) and (20).*

In order to remove the semi-boundedness condition and the condition $r = p$, the stability definition has to be modified. The starting point is the resolvent operator $(sI - Q)^{-1}$ corresponding to the difference operator Q in (17). The well known resolvent condition is

$$\|(sI - Q)^{-1}\| \leq \frac{\text{const}}{\text{Re } s} , \quad \text{Re } s > 0. \quad (30)$$

Consider now the problem (19) with $f = 0$ and $g = 0$. The Laplace transformed problem is

$$\begin{aligned} (sI - Q)\hat{u}_j &= \hat{F}_j , \quad j = 0, 1, \dots , \\ B_0 \hat{u}_0 &= 0, \\ \|\hat{u}\| &< \infty . \end{aligned} \quad (31)$$

The resolvent condition then leads to the estimate

$$\begin{aligned} \|\hat{u}\|^2 &\leq K(\eta) \|\hat{F}\|^2 , \quad s = i\xi + \eta , \\ \lim_{\eta \rightarrow \infty} K(\eta) &= 0 . \end{aligned} \quad (32)$$

By Parseval's relation we obtain a corresponding estimate in the original space, and this one is used as the first alternative stability definition:

Definition 4 *The approximation (19) is stable in the generalized sense if for $f = 0, g = 0$ there is a unique solution satisfying*

$$\begin{aligned} \int_0^\infty e^{-2\eta t} \|u(t)\|^2 dt &\leq K(\eta) \int_0^\infty e^{-2\eta t} \|F(t)\|^2 dt, \\ \eta > \eta_0, \quad \lim_{\eta \rightarrow \infty} K(\eta) &= 0. \end{aligned} \quad (33)$$

The constant $\eta_0 > 0$ is introduced to cover the general case with variable coefficients, lower order terms and two boundaries. For the principal part, constant coefficients and only one boundary, one can choose $\eta_0 = 0$.

A stronger form of the definition above is obtained by introducing nonzero boundary data in the estimate:

Definition 5 *The approximation (19) is strongly stable in the generalized sense if for $f = 0$ there is a unique solution satisfying*

$$\begin{aligned} \int_0^\infty e^{-2\eta t} \|u(t)\|^2 dt &\leq K(\eta) \int_0^\infty e^{-2\eta t} (\|F(t)\|^2 + |g(t)|^2) dt, \\ \eta > \eta_0, \quad \lim_{\eta \rightarrow \infty} K(\eta) &= 0. \end{aligned} \quad (34)$$

The restrictions $f = 0$ and/or $g = 0$ are done only for the formal definition of stability. The actual computation should of course be carried out for the original problem with non-zero initial and boundary data. It can be shown that for these general problems, there may be a growth in the solution of order $1/h$, which is not very severe. The important fact is that even if we introduce another boundary, or if we have variable coefficients, the growth rate remains of the order $1/h$. Even if neither one of the stability definitions is satisfied, there may still be cases where the growth rate is no stronger than $1/h$ for the quarter space problem with constant coefficients. However, when introducing a second boundary and/or variable coefficients in the differential equation, the growth rate becomes worse. This will be illustrated with an example below.

The concept of stability in the generalized sense allows for simpler stability conditions. We have

Theorem 4 *Assume that (19) is a consistent and dissipative approximation of a strictly hyperbolic system. Then the approximation is strongly stable in the generalized sense if the Kreiss condition is satisfied.*

We shall now treat a simple example to illustrate the various stability concepts. Starting from the differential equation $u_t + u_x = 0$ we consider the approximation

$$\begin{aligned} \frac{du_j}{dt} + \frac{u_{j+1} - u_{j-1}}{2h} &= 0, \quad j = 1, 2, \dots, \\ au_0 - u_1 &= 0, \\ \|u\| &< \infty, \\ u_j(0) &= f_j, \end{aligned} \tag{35}$$

where $a \neq 0$ is a complex parameter. The corresponding eigenvalue problem is

$$\begin{aligned} \tilde{s}\phi_j + \phi_{j+1} - \phi_{j-1} &= 0, \quad j = 1, 2, \dots, \\ a\phi_0 - \phi_1 &= 0, \\ \|\phi\| &< \infty. \end{aligned} \tag{36}$$

The general form of the solution for $\operatorname{Re} \tilde{s} > 0$ is

$$\phi_j = \sigma \kappa^j, \tag{37}$$

where $|\kappa| < 1$ satisfies the characteristic equation

$$\tilde{s}\kappa + \kappa^2 - 1 = 0. \tag{38}$$

(Only one root κ is part of the solution, since the other one is larger than one in magnitude.) The condition for a non-trivial solution is

$$a - \kappa = 0. \tag{39}$$

If this condition is satisfied for $\operatorname{Re} \tilde{s} > 0$, then \tilde{s} is an eigenvalue. Since the Kreiss condition requires a uniform estimate of \hat{u}_0, \hat{u}_1 corresponding to ϕ_0, ϕ_1 above, we must consider the case $\operatorname{Re} \tilde{s} = 0$. If (39) is satisfied for a value \tilde{s}_0 on the imaginary axis, then there are two possibilities:

- i) $|\kappa(\tilde{s}_0)| < 1$. Then \tilde{s}_0 is an eigenvalue in the true sense.
- ii) $|\kappa(\tilde{s}_0)| = 1$. Then the condition $\|\phi\| < 0$ is not satisfied, and we say that \tilde{s}_0 is a *generalized eigenvalue*.

For general approximations, we require that all the roots κ are inside the unit circle for case i), and at least one root κ is on the unit circle for case ii).

With these concepts, we can formulate the Kreiss condition by requiring that there is no eigenvalue or generalized eigenvalue \tilde{s} with $\operatorname{Re} \tilde{s} \geq 0$. Note that in all cases we are considering only the part of the solution where $\|\phi\| < 0$ for $\operatorname{Re} \tilde{s} > 0$, and the properties of these solutions in the limit as $\operatorname{Re} \tilde{s} \rightarrow 0$. The part of the solution containing the roots κ outside the unit circle is eliminated from the beginning.

In order to find out how the parameter a influences the stability, we define the domain

$$\Omega = \{|\kappa| \leq 1, \operatorname{Re} \kappa \geq 0\}, \tag{40}$$

see Figure 3.

From (39) we easily derive the following four cases:

1. $a \notin \Omega$: Strongly stable
2. $a \in \text{Interior}(\Omega)$: Eigenvalue $\text{Re } \tilde{s} > 0$, unstable
3. $a = i\tau$, $-1 < \tau < 1$, $\tau \neq 0$: Eigenvalue $\text{Re } \tilde{s} = 0$
4. $|a| = 1$, $\text{Re } a \geq 0$: Generalized eigenvalue $\text{Re } \tilde{s} = 0$

The two cases 1 and 2 are clear: case 1 is the best of all situations, case 2 is a useless approximation. In case 3 and 4 the Kreiss condition is violated, but it turns out that the degree of violation is different for different values of a .

It can be shown that in case 3 the condition (32) is satisfied, i.e., the approximation is stable in the generalized sense. As we have discussed above, with this type of approximation and with nonzero initial data f , there may be a growth of the type $\|f\|/h$. By explicit calculation of the norm of the solution to (35), one can show exactly this. However, the growth rate stays like that even if there are two boundaries. The explanation for that is that the eigensolutions decay very quickly away from the boundary, and before the wave hits the other boundary it is annihilated.

For case 4 we distinguish between two sub-cases:

- 4a. $|a| = 1$, $\text{Re } a > 0$
- 4b. $a = \pm i$

In case 4a, we have a generalized eigenvalue of the standard form. The approximation is not stable in any sense that we have defined. Again the growth of the right quarter space problem is of the order $1/h$. However, one can show that the solution to the problem with two boundaries has a growth rate of order $(1/h)^{\alpha t}$, where $\alpha > 0$. The explanation is that the wave corresponding to the eigensolution doesn't get damped before hitting the other boundary, and when it reaches the left boundary again, it picks up another growth factor $1/h$.

In case 4b we have a very special situation. One can prove that the approximation is stable in the generalized sense, and just like in case 3, the growth rate remains $1/h$ even for the two boundary case. Without doing the full analysis, we point out the main reason for this nice behavior despite the fact that we have a generalized eigenvalue.

Consider for a moment nonzero data in the boundary condition, i.e., we have

$$a\phi_0 - \phi_1 = \hat{g} \quad (41)$$

in (36). For the constant σ in the solution (37), we get

$$\sigma = \frac{\hat{g}}{a - \kappa}. \quad (42)$$

Hence, the size of $1/|a - \kappa|$ is a measure of the strength of the singularity. Consider next the characteristic equation (38). The critical values are $\kappa_0 = a = \pm i$, which correspond to the generalized eigenvalues $\tilde{s}_0 = \mp i$. At these particular points, κ_0 is a double root of the characteristic equation. Therefore, when \tilde{s} approaches \tilde{s}_0 , $\kappa(\tilde{s})$ approaches κ_0 according to the inequality

$$|\kappa(\tilde{s}) - \kappa_0| \geq \text{const } |\tilde{s} - \tilde{s}_0|^{1/2}. \quad (43)$$

Hence, σ in (42) doesn't grow as fast when \tilde{s} approaches $\mp i$ as it does for all other generalized eigenvalues, i.e., the singularity of our operator is not so severe as in the normal case.

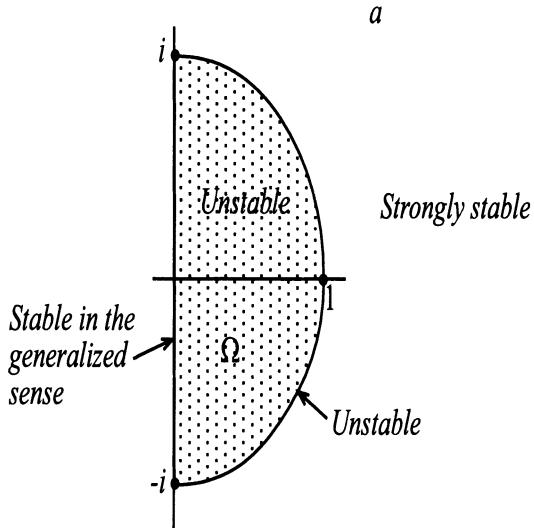


Figure 3. Stability properties as a function of a

4. Fully discretized approximations

For convenience we consider only one-step schemes, and only the right quarter space problem:

$$\begin{aligned}
u_j^{n+1} &= Pu_j^n + \Delta t F_j , \quad j = 1, 2, \dots , \\
L_0 u_0^{n+1} &= g^{n+1} , \\
\|u^n\| &< \infty , \\
u_j^0 &= f_j .
\end{aligned} \tag{44}$$

Here P is a difference operator with constant coefficients, and L_0 is a boundary operator. The theory for these approximations are very similar to the one presented in the previous section for semi-discrete approximations. By defining the discrete solution u_j^n also in between the time-levels t_n , for example as a piecewise constant function, the Laplace transform technique can be used. By assuming $f_j = 0$ in (44), and using the transformation $z = e^{s\Delta t}$, we get the z -transformed problem

$$\begin{aligned}
z \tilde{u}_j &= P \tilde{u}_j + \Delta t \tilde{F}_j , \quad j = 1, 2, \dots , \\
\tilde{L}_0 \tilde{u}_0 &= \tilde{g} , \\
\|\tilde{u}\| &< \infty .
\end{aligned} \tag{45}$$

The right half of the complex s -plane of interest for the semi-discrete case, is now transformed into the domain outside the unit circle: $|z| \geq 1$. The eigenvalue problem is

$$\begin{aligned}
z \phi_j &= P \phi_j , \quad j = 1, 2, \dots , \\
\tilde{L}_0 \phi_0 &= 0 , \\
\|\phi\| &< \infty .
\end{aligned} \tag{46}$$

The Godunov-Ryabenkii condition was defined already in Section 2: No eigenvalues z with $|z| > 1$ are allowed.

The Kreiss condition is: There are no eigenvalues or generalized eigenvalues with $|z| = 1$.

The stability definitions are the same as for the semi-discrete case, except that integrals are substituted by sums:

Definition 6 *The approximation (44) is stable in the generalized sense if for $f = 0$, $g^n = 0$ there is a unique solution satisfying*

$$\sum_{n=1}^{\infty} e^{-2\eta t_n} \|u^n\|^2 \Delta t \leq K(\eta) \sum_{n=1}^{\infty} e^{-2\eta t} \|F^{n-1}\|^2 \Delta t , \tag{47}$$

$\eta > \eta_0$, $\lim_{\eta \rightarrow \infty} K(\eta) = 0$.

The concept of strong stability in the generalized sense is defined by an obvious modification. For dissipative approximations, a simple stability condition is obtained as for the semi-discrete case. Indeed, Theorem 4 holds exactly word by word, only with (19) substituted by (44).

The method of lines.

Assume that an elimination has been made by using the boundary conditions, such that the semi-discrete approximation is formulated as

$$\begin{aligned}\frac{du_j}{dt} &= Qu_j + F_j, \quad j = 1, 2, \dots, \\ u_j(0) &= f_j.\end{aligned}\tag{48}$$

This is a system of ODE, and an ODE-solver can be directly applied. This solution procedure is called *the method of lines*. Consider now the class of Runge-Kutta methods, which can be written in the form

$$\begin{aligned}u_j^{n+1} &= P(\Delta t Q)u_j^n + P_1(\Delta t Q)F_j^n, \quad j = 1, 2, \dots, \\ u_j(0) &= f_j,\end{aligned}\tag{49}$$

where P and P_1 are polynomials in $\Delta t Q$. We assume that $\Delta t = \text{const} \cdot h$, such that $\|\Delta t Q\|$ is bounded. For the test equation $y' = \lambda y$, the stability domain Ω is defined by

$$\Omega = \{ z / |P(z)| \leq 1 \}.\tag{50}$$

Under a few technical assumptions one can prove, see (Kreiss and Wu, 1993)

Theorem 5 *Assume that one can inscribe a semi-circle $C = \{z / |z| < R, \operatorname{Re} z < 0\}$, see Figure 4. If the semi-discrete approximation is stable in the generalized sense, then the Runge-Kutta approximation (49) is stable in the generalized sense if*

$$\|\Delta t Q\| \leq R.\tag{51}$$

A similar theorem holds for linear multi-step methods.

The sufficient condition (51) for stability may be restrictive. Consider the Cauchy problem where the boundary conditions are removed, and denote by $Q = Q_C$ the corresponding difference operator. Then the von Neumann condition requires the spectrum of $\Delta t \hat{Q}_C$ to be contained in the stability domain Ω . A corresponding slightly stronger condition is

$$\rho(\Delta t Q_C) \leq R,\tag{52}$$

which would be an acceptable restriction on the timestep. However, the norm of Q for the initial-boundary value problem may be much larger than $\rho(Q_C)$. Therefore, it is natural to ask if it possible to derive another theorem, where stability follows under a weaker condition than (51). In particular, conjectures have been made that stability follows if

- i) The fully discrete Cauchy problem is stable
- ii) The Kreiss condition is satisfied for the semi-discrete approximation

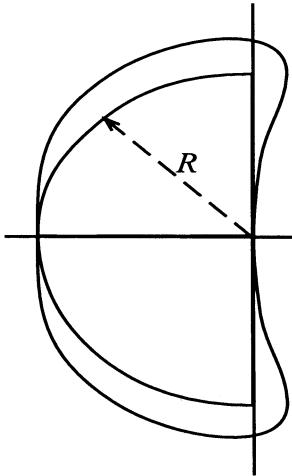


Figure 4. Stability domain and semi-circle \mathcal{C}

Since the semi-discrete scheme has no time-step involved, no extra time-step restriction would be introduced. However, the conjecture is false. There is a counterexample (Gustafsson, 1998) based on a 6th order accurate approximation of a simple 2×2 system, where $h\rho(Q_C) = 1.58$, resulting in a stable Runge-Kutta approximation for the Cauchy problem under the condition

$$\frac{\Delta t}{h} 1.58 \leq R. \quad (53)$$

However, one can show that

$$h\rho(Q) \rightarrow 2.26 \quad (54)$$

for the initial boundary value problem, which implies the necessary and more restrictive stability condition

$$\frac{\Delta t}{h} 2.26 \leq R. \quad (55)$$

5. Conclusion and related work

In this paper we have demonstrated the fundamental importance of the work by Godunov and Ryabenkii, and the connection to the later theory by Kreiss and others. We have referred mainly to the paper (Godunov and Ryabenkii, 1963), but that one was a partial result of the work on the book (Godunov and Ryabenkii, 1964), which contains a number of interesting,

and for that time, new theory for difference approximations. We have also pointed out the significance of the Moscow school in the early fifties, where Gelfand and others gave important contributions.

The Godunov-Ryabenkii theory was originally developed for the fully discrete case, and so was the theory by Kreiss and others that followed. A good deal of this paper is devoted to the semi-discrete case, just because it is a little easier to handle. The first general treatment of this class of problems was given by Strikwerda (Strikwerda, 1980). But the presentation in the present paper is based mainly on the material in the book by Gustafsson et.al., (Gustafsson and Kreiss and Oliger, 1995).

The first complete theory containing sufficient conditions for stability was presented by Kreiss in (Kreiss, 1968), but the class of difference schemes was restricted to dissipative one-step schemes. The concept of generalized eigenvalues was introduced, but the proofs were based on a somewhat different theory than the one presented here. Osher (Osher, 1969) was able to relax some of the conditions. In 1970, Kreiss published the famous paper (Kreiss, 1970), where the initial-boundary value problem for hyperbolic differential systems of PDE was given a complete treatment based on Laplace transform technique. This technique could be modified in a way such that difference approximations of general type could be treated as well. That work resulted in the paper (Gustafsson et. al., 1972), which sometimes has been called the GKS-theory. Michelson (Michelson, 1983) gave a full generalization to the multidimensional case.

In the papers mentioned above, the theory doesn't make use of any symmetry assumptions on the coefficient matrices. By assuming symmetry, the theory simplifies a lot, and it is possible to prove theorems like Theorems 2 and 3. This technique was first presented in (Gustafsson and Kreiss and Oliger, 1995).

Even when having the theory available, it is not trivial to check if a certain difference approximation is stable or not. The main difficulty is to check the Kreiss condition, and for approximations of systems of PDE, one may have to rely on numerical verification. The software system IB-STAB constructed by Thuné (Thuné, 1990) uses this technique. In order to simplify the analysis, Goldberg and Tadmor came up with a number of conditions that in some cases were more restrictive, but much simpler to verify. The first paper (Goldberg and Tadmor, 1981) in a longer series occurred in 1981; the last one (Goldberg and Tadmor, 1991) occurring in 1991 contained most of the earlier results in more compact form.

Several important contributions have also been made by Nick Trefethen. He interpreted the general theory in a new way, by relating the Kreiss condition to the group-velocity (Trefethen, 1983). In a joint work with Reddy (Reddy and Trefethen, 1992), the same author also developed a

different theory based on pseudo-eigenvalues for the method of lines.

References

- R. Courant and K.O. Friedrichs and H. Levy (1928). Über die partielle differentialgleichungen der matematischen physik. *Mathematische Annalen* **100**, pp 32-74.
- S.K. Godunov and V.S. Ryabenkii (1963). Spetral stability criteria for boundary value problems for non-self-adjoint difference equations. *Uspekhi Mat. Nauk.* **18**, pp 1-12
- H.-O. Kreiss and L. Wu (1993). On the stability definition of difference approximations for the initial boundary value problem. *Appl. Num. Math.* **12**, pp 213-227.
- B. Gustafsson (1998). On the implementation of boundary conditions for the method of lines *BIT* **38**, pp 293-314.
- S.K. Godunov and V.S. Ryabenkii (1964). Theory of Difference Schemes. Northe-Holland.
- J.C. Strikwerda (1980). Initial boundary value problems for the method of lines. *J. Comput. Phys.* **34**, pp 94-110.
- B. Gustafsson and H.-O. Kreiss and J. Oliger (1995). Time Dependent problems and Difference Methods. Wiley and Sons.
- H.-O. Kreiss (1968). Stability theory for difference approximations of mixed initial boundary value problems. *I. Math. Comp.* **22**, pp 703-714.
- S. Osher (1969). Stability pf difference approximations of dissipative type for mixed initial boundary value problems. *I. Math. Comp.* **23**, pp 335-340.
- H.-O. Kreiss (1970). Initial boundary value problems for hyperbolic systems. *Comm. Pure Appl. Math.* **23**, pp 277-298.
- B. Gustafsson and H.-O. Kreiss and A. Sundström (1972). Stability theory of difference approximations for mixed boundary value problems. *II. Math. Comp.* **26**, pp 649-686.
- D. Michelson (1983). Stability theory of difference approximations for multidimensional initial-boundary value problems. *II. Math. Comp.* **40**, pp 1-46.
- M. Thune (1990). A numerical algorithm for stability analysis of difference methods for hyperbolic systems. *SIAM J. Sci. Stat. Comput.* **11**, pp 63-81.
- M. Goldberg and E. Tadmor (1981). Scheme-independent stability criteria for difference approximations of hyperbolic initial-boundary value problems. *II. Math. Comp.* **36**, pp 605-626.
- M. Goldberg and E. Tadmor (1991). Simple stability criteria for difference approximations of hyperbolic initial-boundary value problems. II. In *Third International Conference on Hyperbolic Problems, Studentlitteratur*.
- L1.N. Trefethen (1983). Group velocity interpretation of the stability theory of Gustafsson, Kreiss and Sundström. *J. Comp. Phys.* **49**, pp 199-217.
- S.C. Reddy and L1.N. Trefethen (1992) Stability of the method of lines. *Num. Math.* **62**, pp 235-267.

A FRONT TRACKING METHOD FOR HYBRID GRIDS

D. HANEL, L. TRAN, R. VILSMEIER

Institute for Combustion and Gasdynamics,

University of Duisburg,

47057 Duisburg, Germany

E-mails: {hj454ha, linhba, hj000vi}@vug.uni-duisburg.de

WWW: http://www.vug.uni-duisburg.de/

Abstract.

A discrete treatment of discontinuities with a split flux formulation on static meshes is discussed. The approach does not require subcell resolution and the formulation is relatively simple on any mesh type. In this first version conservation is not ensured, but this is a drawback to overcome. Both sides of the discontinuities are coupled by internal boundary conditions. For the description of the location and motion of the discontinuities level set functions are used, which are defined in the vicinity of the discontinuities only. This approach allows the treatment of problems with multiple embedded discontinuities, however the problem of crossing fronts is not yet addressed.

1. Introduction

Discontinuous flow features, as material interfaces, shocks or detonations, appear in many flow problems. In general, such flow features are resolved numerically by capturing methods. Thus discontinuities become smeared out by artificial intermediate states, which requires high resolution to restrict the physical space of intermediate states (adaption) or even falsifies the solution in problems of strong interaction. In such cases an exact treatment of discontinuities by front tracking is of advantage.

Front tracking enables numerically a representation of discontinuities as jumps and tracks them in space and time, thus avoiding intermediate states and their consequences. These methods can therefore improve the accuracy and efficiency of numerical simulations, e.g. in free surface flows or

reactive detonative flows. On the other hand front tracking introduces a special treatment of solution parts with additional algorithmic difficulties. Although the mathematical and physical principles of weak solutions of conservation laws are well known, the numerical realization is complex and has lead to different approaches.

In this paper a front tracking concept based on the level-set approach (Kerstein et. al., 1988; Mulder, Osher, Sethian, 1992; Sussman, Smereka, Osher, 1994) was found to be well suited, however having in mind that some severe topological restrictions apply to the classical variants. The method is combined with a relatively simple flux separation scheme on the interfaces across discontinuities and is applicable on arbitrary, unstructured grids. This concept is integrated as a module in a finite-volume solution framework for systems of conservation laws on unstructured and hybrid grids (MOUSE). Object oriented programming allows modular and flexible program developments, reusable for arbitrary physical problems.

2. Governing equations and solution concepts

Consider the following general system of equations modelling a physical problem:

$$\frac{\partial}{\partial t} \int_V \mathbf{Q} dV + \oint_A \vec{\mathbf{H}} \cdot \vec{n} dA = \int_V \mathbf{S} dV \quad (1)$$

where Q , H and S represent the variables, fluxes and source terms respectively. The system of equations is solved on arbitrary grids employing a nodal finite volume method, Fig. 1. A discrete formulation of the system reads as follows:

$$\frac{\Delta \mathbf{Q}}{\Delta t}|_{Vd} + Res_{\Delta, Vd} = 0 \quad \text{with: } Res_{\Delta, Vd} = \frac{1}{V_{Vd}} \sum_{i=1}^{nv} \mathbf{H}_i \vec{n}_i \Delta A_i \quad (2)$$

where Vd is a control volume with the size V_{Vd} and the sum is carried out over all bounding segments. As underlying software, the MOUSE package (Gloth, Vilsmeier, Hänel, 1997), currently in development at the site of the authors, is used. The following methods for front-tracking are, however, not yet available in the present public release.

3. Location and motion of discontinuities

In the present work a level-set formulation is used on a fixed computational grid. Since the method has previously been introduced in more detail (Tran,

Vilsmeier and Hänel, 1999), the following description is intended to provide only a short overview.

The level-set method introduces a continuous function $G(x, [y, z], t)$, while the actual position of the discontinuity is given by a discrete iso-value, typically 0.

The motion of the discontinuity can thus be described by a scalar convection equation:

$$\frac{\partial G}{\partial t} + \vec{c} \cdot \nabla G = 0 \quad (3)$$

where \vec{c} is the propagation speed.

For the present paper, discrete values for G are stored at the nodes of the mesh. It is however not required to define the function G in the whole computational domain, instead it can be described locally in two neighbourhood levels. Figure 2 illustrates the neighbourhood levels. Points and edges are marked with P and K , while their neighbourhood level is given by the number appended.

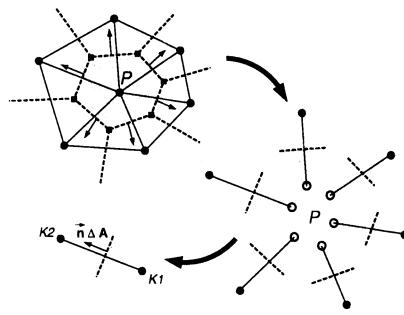


Figure 1. Top: Control volume

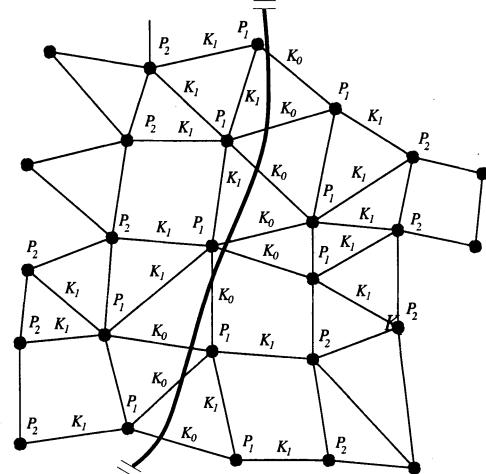


Figure 2. Right: Discontinuity on a 2-D mesh and membership of nodes and edges to corresponding levels

The restriction to two neighbourhood levels imposes, that for each step of motion, only nodes of the first neighbourhood level (P_1) can be overrun, if any. In this case the function G can still describe the position between the previous first and second level. After each time step, the definition zone is adjusted accordingly.

The above equation (3) is discretized in space and integrated in time. Although same physical time steps are employed, the solution is decoupled from the actual system of conservation equations considered. Since, for active discontinuities a propagation speed \vec{c} is only available at their actual position, a meaningful physical transport can only be evaluated on the

edges $K0$ crossing these. Therefore a finite difference type discretization is chosen.

The accuracy of the method is directly related to the curvature of the G -function normal to the discontinuities. Therefore it is useful to provide a constant slope as far as possible. This is the aim of normalization. It is desired, that the function G satisfies the normalizing condition $|\nabla G| = C$, where C is a constant $\neq 0$, typically chosen $C = 1$. At present a least squares based approach is chosen and much care is taken to minimize influences on the actual position of the discontinuity, please see (Tran, Vilsmeier and Hänel, 1999) for details and validation.

In the above description a propagation speed \vec{c} was used, without stating how this is obtained. Considering a passive transport (e.g. material interfaces, contact discontinuities, etc.,) the carrier speed is easy to access. For active discontinuities the propagation speed depends on the corresponding values at both sides of the discontinuity. Consider as example a shock, whose propagation speed can be computed upon the pressures at both sides. For the edges $K0$ crossing the discontinuity, left and right states can be computed upon projected variables from the ending nodes.

4. Flux separation method for back influence

Since the position and motion of discontinuities can be treated as described, it is now interesting, how the discontinuity affects the surrounding fields. In the present work, a flux separation scheme was used, not requiring a sub-cell resolution. In this method the original control volumes of the mesh are preserved. Referring to Fig. 1, each edge of the mesh carries a segment of two adjacent control volumes. In smooth regions, a projection from the nodes storing the variables to the cell interface is performed and a flux computed in a central or upwind manner. The flux is then used for both the adjacent control volumes.

For all edges $K0$ crossing the discontinuity this is no longer the case. The flux H_L for the "left" side is computed using the "left" side projection Q_L , and the flux for the "right" side with the right side projection accordingly, see Fig. 4. This method is very simple and can be employed on any grid.

4.1. INNER BOUNDARY CONDITIONS

The above described flux separation is not sufficient to couple both sides of the discontinuity. The reason can easily be seen from the characteristics. For simplification, consider the situation on a single edge as a one-dimensional problem. On both sides, representing the both states alongside the discontinuity, there are different sets of characteristics, including incoming and

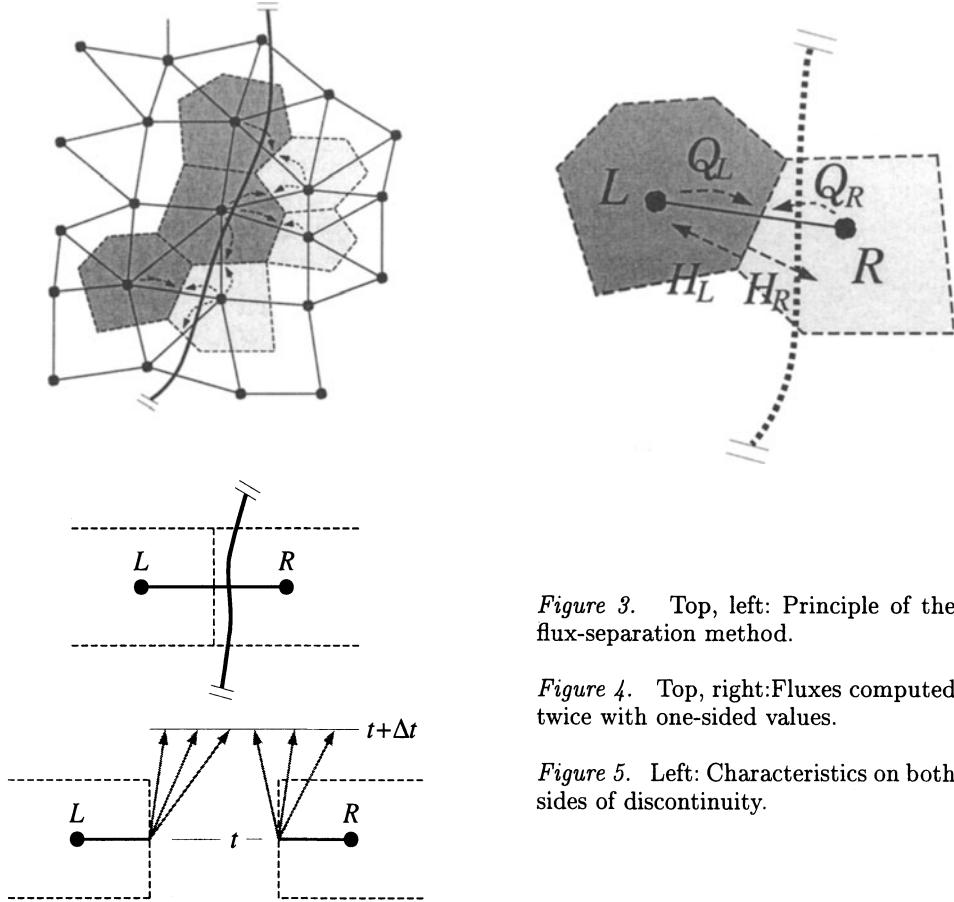


Figure 3. Top, left: Principle of the flux-separation method.

Figure 4. Top, right: Fluxes computed twice with one-sided values.

Figure 5. Left: Characteristics on both sides of discontinuity.

outgoing ones. Figure 5 illustrates that. As for an outer boundary of a computation, conditions must be imposed for incoming characteristics.

In the case of discontinuities, jump relations are evaluated accordingly. Consider for example a stationary shock in 1D: On the "cold" side there are 3 outgoing characteristics, thus no correction is required. On the "hot" side, there are two incoming and one outgoing characteristics. Accordingly, two conditions are imposed evaluating the jump relations. In practice, the pressure is kept at the "hot" side.

4.2. OVERRUNNING NODES

As a node of the mesh gets overrun by a front, the variables stored at the corresponding nodes switch to the other side. At present these new values are obtained by extrapolation from the the surrounding neighbours at the new side. The set of edges hit by the discontinuity is adjusted accordingly.

4.3. CONSERVATION

The above described method does, in the present version, not ensure conservation. First, the different flux used for the update of the "left" and "right" residual of two cells neighbouring the discontinuity is a problem. Second, the movement of the discontinuity shows a truncation error and the contents of conserved quantities in both adjacent cells depends on the position of the discontinuities. Last, also the very simple overrun formulation impairs conservation, since it relies on simple interpolation.

However, a fully conservative variant of the method is currently in development and expected to be available soon.

5. Test Cases and Results

The above presented methods have been developed without any preference for a specific application. However, the examples below are related to fluid dynamics. The figures 6 and 7 show computed results for a shock tube problem where both, the front shock and the shear layer are tracked. The computation was performed on a 2-D unstructured mesh with triangular elements.

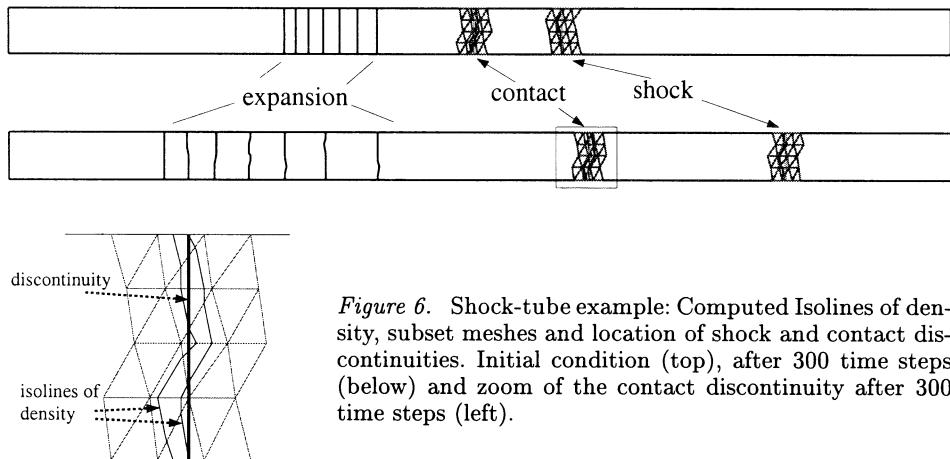


Figure 6. Shock-tube example: Computed Isolines of density, subset meshes and location of shock and contact discontinuities. Initial condition (top), after 300 time steps (below) and zoom of the contact discontinuity after 300 time steps (left).

As a second example, the inviscid computation for a cylinder at $Ma_\infty = 2$ is shown in Fig. 8. Shown are the computational mesh, tracked position of the bow shock, corresponding definition zone, and isolines of density.

Figure 9 is a radially symmetric expanding shock wave. Shown here are computational mesh, isolines of density, and density profiles along $y = 0.5$ of initial, intermediate, and final state. The shock strength and density profiles found compare favourably to the result, previously shown by Leveque and Shyue (Leveque and Shyue, 1996).

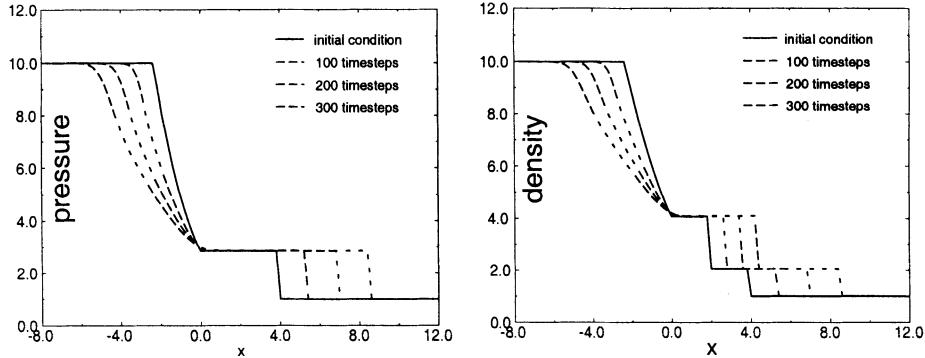


Figure 7. Shock-tube example: cut through a 2-D solution at different time levels.

Another interesting case to test the robustness of the present method is the reverse of the above test case, a focusing shock wave. Using the same mesh, Fig. 10 shows final state before collapsing of the level set (definition zone, isolines of density, $G = 0$ line) and the density profiles at initial, intermediate, and final state. One can see the sharpness of the result until the very end. The ratio of density of the final state reaches 5.73, approaching the expected value of 6, the theoretical value when the Mach number goes to infinity. The difference is due to the coarse mesh used. The non-symmetry of the final state profile is due to interpolation.

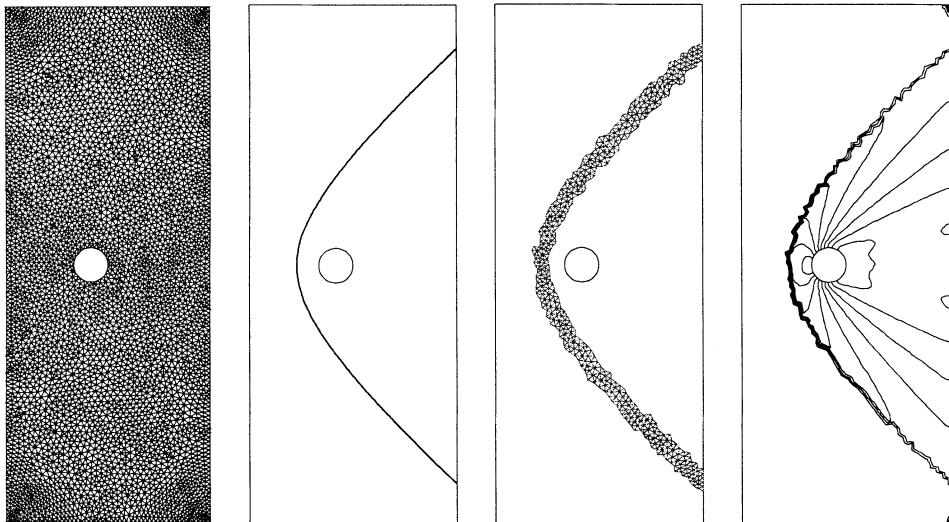


Figure 8. Flow past a cylinder (inviscid, $Ma_\infty = 2$): Mesh, final position of tracked shock, definition zone and density contours.

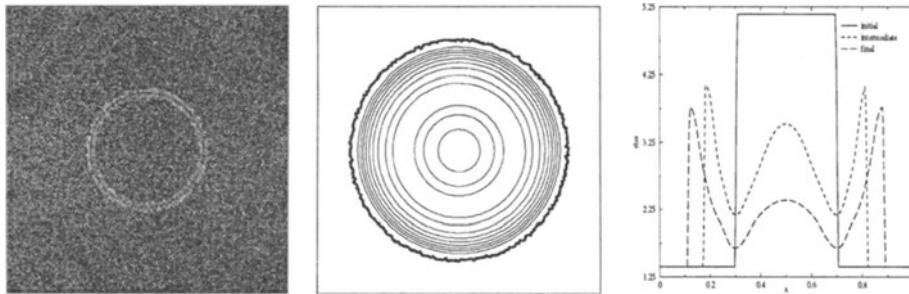
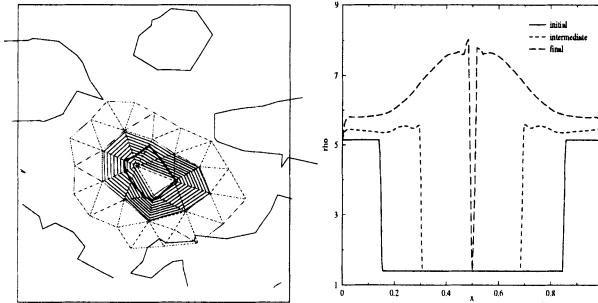


Figure 9. Top: Radially symmetric expanding shock wave (mesh, density contour, density profile along $y = 0.5$).

Figure 10. Right: Radially symmetric focusing shock wave (zoom-in of near final state, density profile along $y = 0.5$).



References

- Kerstein A., Ashurst W., Williams F. (1988). Field Equation for Interface Propagating in an Unsteady Homogeneous Flow Field. *Phys. Rev. A*, vol. 37, pp 2728-2731.
- Mulder, Osher, Sethian (1992). Computing Interface Motion in Compressible Gas Dynamics. *J. Comp. Phys.*, vol. 100, pp 209-228.
- Sussman M., Smereka P., Osher S. (1994). A Level Set Approach for Computing Solutions to Incompressible Two-Phase Flow. *J. of Comp. Physics*, vol. 114, pp 146-159.
- Gloth O., Vilsmeier R., Hänel D. (1997). Object oriented programming for computational fluid dynamics. *Proceedings of HiPer' 97*, Krakow, Poland.
- Tran L., Vilsmeier R., Hänel D. (1999). A local level set method for the treatment of discontinuities on unstructured grids. *Proceedings of FVCA-II 99*, Duisburg, Germany.
- Leveque R. J. and Shyue K. M. (1996). Two-dimensional front tracking based on high resolution wave propagation methods. *J. of Comp. Physics*, vol. 123, pp 354-368.

A PROBLEM OF CLASSICAL SHOCK CAPTURING FINITE VOLUME SCHEMES IN HYPERSONIC FLOWS

V. HANNEMANN

*Institute of Fluid Mechanics,
German Aerospace Center (DLR)
Bunsenstraße 10, D-37073 Göttingen,
Email: Volker.Hannemann@dlr.de*

1. Introduction

Usually shock waves as well as contact discontinuities are resolved by classical shock-capturing finite volume schemes by several interior cells. The best value of a conservative variable in such a cell is the integral of its exact solution divided by the volume of the cell. Although these mean values do not exist in the exact solution, they can normally be handled by the numerical method without spoiling the solution on either side of the discontinuity. The situation changes when the temperature and the ratio of the specific heats γ have different values on each side of the discontinuity. The standard procedure to calculate the cell averaged pressure starts with the internal energy and then determines one temperature and an effective γ for the mean state. This information provides a mean pressure which can deviate from the sum of the partial pressures of the exact solution for a discontinuity within a cell.

Examples of situations where numerical errors are introduced into the flow field by such a process are:

1. a contact discontinuity between fluids with different γ ,
2. discontinuities with high temperature jumps in a calorically non perfect gas.

Both cases occur in hypersonic flows, but only the first one is extensively discussed in the literature (Ton, 1996), (Shyue, 1998). Therefore the second case will be examined here. Although we can not offer a modification of the method to prohibit the error, awareness of it should be useful for the user of such a method.

2. Numerical Method

The system of one dimensional, compressible Euler equations is closed with different assumptions for the temperature dependence of the ratio of the specific heats:

- calorically perfect gas:

$$\gamma = 1.4,$$

- calorically non perfect gas:

$$\gamma(T) = \begin{cases} 1.4 & T \leq 600\text{ K} \\ 1.4 + (\gamma_2 - 1.4) \frac{T - 600\text{ K}}{1400\text{ K}} & 600\text{ K} < T < 2000\text{ K} \\ \gamma_2 & T \geq 2000\text{ K} \end{cases}$$

- $\gamma_2 = 9/7 = 1.28\dots$: simplified air with linear influence of the vibrational excitation
- $\gamma_2 = 1.1$: extreme case to study γ -dependence

The finite volume method to solve these equations consists of an equidistant spatial discretization, a forward Euler time discretization and the AUSMDV (Wada and Liou, 1994) Riemann solver as a numerical flux function at the cell interfaces. To achieve higher order accuracy a two step Runge-Kutta time integration is combined with a TVD-MUSCL-approach (reconstruction of piecewise linear primitive variables and minmod-limiter). A CFL-number of 0.9 is sufficient for the shock wave test cases, but is decreased to 0.5 to avoid high frequency oscillations when dealing with contact discontinuities.

3. Test cases

The two types of discontinuities - shocks and contacts - are investigated as steady state and as moving solutions with respect to the grid. The initial conditions are chosen in a way to get a temperature increase from 300 K on one side of the discontinuity to more than 2000 K on the other side. To avoid additional error sources most calculations are carried out with the first order scheme.

3.1. STEADY STATE SHOCK WAVE

For the calorically perfect gas a pressure ratio of $p_2/p_1 = 50$ is chosen which corresponds to a pre-shock Mach number of $M_1 \approx 6.6$, a density ratio of $\rho_2/\rho_1 = 5.375$ and velocities of $u_1 \approx 7.76 \sqrt{p_1/\rho_1}$ and $u_2 \approx 1.44 \sqrt{p_1/\rho_1}$.

A pressure ratio of $p_2/p_1 = 150$ is chosen for the calorically non perfect gas with $\gamma_2 = 1.1$. This corresponds to a pre-shock Mach number of $M_1 \approx$

10.6, a density ratio of $\rho_2/\rho_1 \approx 20.2$ and velocities of $u_1 \approx 12.5 \sqrt{p_1/\rho_1}$ and $u_2 \approx 0.62 \sqrt{p_1/\rho_1}$.

Both calculations converge to a steady state solution. Figure 1 shows the γ -jump and the preservation of the total enthalpy in the case of the calorically non perfect gas. The mass flux is constant away from the discontinuity, but the values in the interior volumes deviate about 20% in the calorically perfect case and about 60% for the non perfect case. Although the error increases in the non perfect gas case, this error is not important, because the values of the interior volumes are not part of the exact solution anyway.

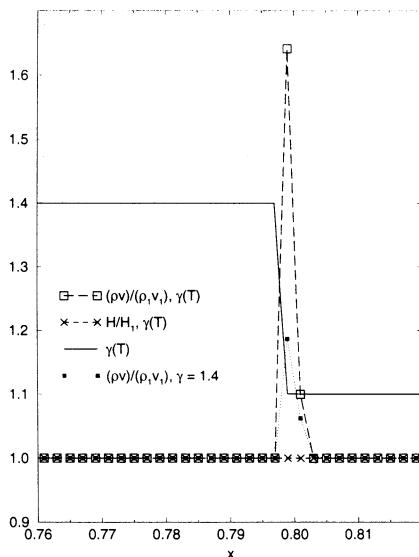


Figure 1. Steady state shock waves

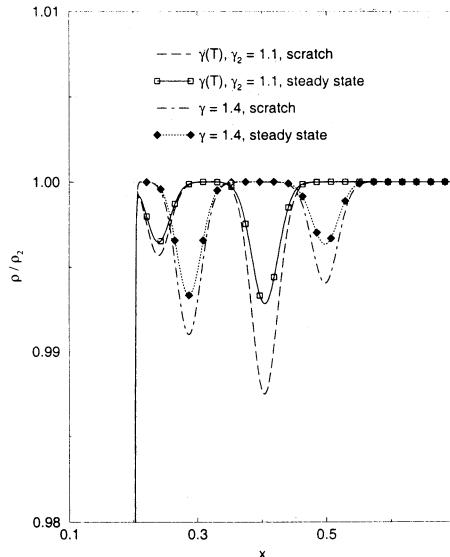


Figure 2. Moving shock wave, density normalized by the post-shock density

3.2. MOVING SHOCK WAVE

This test case is created by a superposition of the steady state shock wave described above and a shock speed of $v_s = -10 \sqrt{p_1/\rho_1}$. The calculations are started from two different initial conditions called *scratch* and *steady state*. The *steady state* condition means that the shock speed is superposed to the steady state solution of the previous test case. The *scratch* condition, also used for the steady state investigation, sets the conservative values to the cell mean values of the exact solution.

In figure 2 each of the four solutions shows two disturbances of the density profile with a magnitude in the order of one percent. The right peak

travels with the characteristic speed of $v_2 + c_2$ and is accompanied by disturbances in the pressure and velocity profiles. The left peak appears in the density profile only and is convected with the local velocity v_2 . Starting from *scratch* the disturbances are up to 75% higher, but comparing the *steady state* solutions no significant difference shows up between the calorically perfect and non perfect gas.

3.3. STEADY STATE CONTACT DISCONTINUITY

A steady state solution of a contact discontinuity in one dimensional flow means zero velocity in every cell. In contrast to some other Riemann solvers the AUSMDV is able to handle this situation. To gain the temperature increase a density ratio of $\rho_2/\rho_1 = 0.1$ is chosen. Figure 3 depicts the preservation of the exact solution when the jump is collocated with a cell interface. If the density and energy profiles jump within a cell, the *scratch* solution is spread over some cells as soon as the ratio of the specific heats is not constant across the discontinuity. A higher γ -jump enforces this behavior. Even then the numerical algorithm converges to a steady state solution.

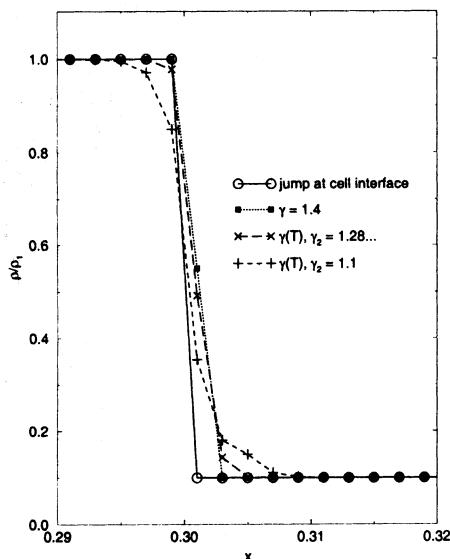


Figure 3. Steady state contact

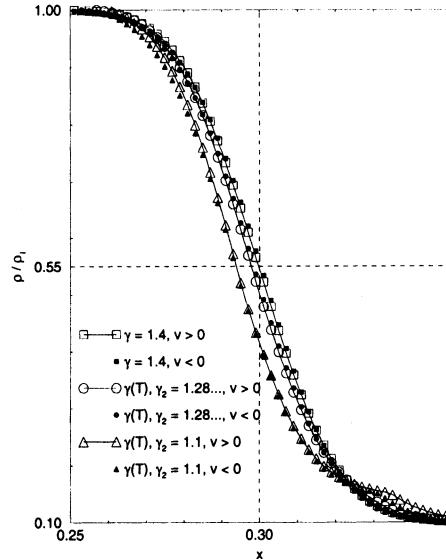


Figure 4. Moving contact discontinuity

3.4. MOVING CONTACT DISCONTINUITY

A superposition of the previous test case and convective velocities of $v_c = \pm 10\sqrt{p_1/\rho_1}$ defines the different tests of this case. All computations are

started from *scratch* due to the equality with the steady state solution when matching a cell interface.

Figure 4 shows the density profiles for the three different γ -models. The initial location of the discontinuity and the convective velocities are chosen in such a way that after the same time interval the exact solution for both directions lies at $x = 0.3$. The calorically perfect gas computations reach this location from both sides with the arithmetic mean value of the initial densities. The calorically non perfect gas solutions with the same γ_2 collocate with each other, but deviate from the exact solution in the higher density (lower temperature) region. The non perfect gas contact discontinuity runs slower into the higher temperature region and faster into the lower temperature region than the perfect gas solution.

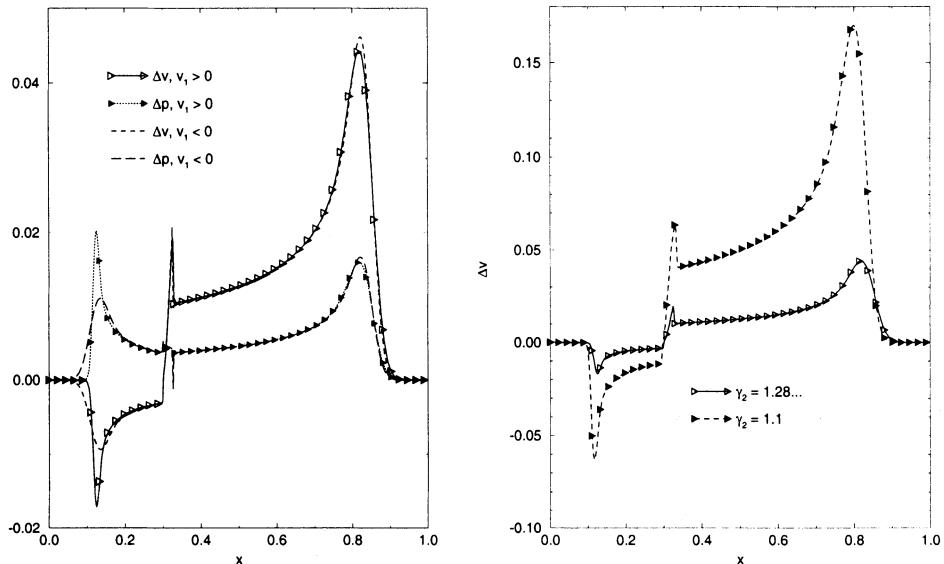


Figure 5. Errors of the moving contact in non perfect gas, $\gamma_2 = 1.28\dots$

Figure 6. Moving contact discontinuity, γ -dependence of velocity error

Figure 5 shows the errors of the velocity and pressure fields $\Delta v = (v - v_1)/v_1$, $\Delta p = (p - p_1)/p_1$. Comparing the right and left traveling solutions the error distributions of the pressure and the velocity give the same tendencies. With the exception of the right and left running peaks, both solutions are nearly the same. This explains the collocation of the density profiles. The smoother appearance of the peaks running ahead of the discontinuity is understandable by considering the much higher number of cells they have passed till reaching the same position. The velocity error behaves similar to the pressure error with two differences; the change of sign within the contact discontinuity and the about 2.5 times higher level

in the higher temperature region. Therefore, the pressure error is not shown for the following tests.

The dependence of the solution on the γ -jump of the initial conditions is shown in figure 6. While the velocity error of the simplified air solution stays below 5%, it exceeds 15% in the extreme case ($\gamma_2 = 1.1$). Except of the level of the error, both non perfect gas models behave very similar.

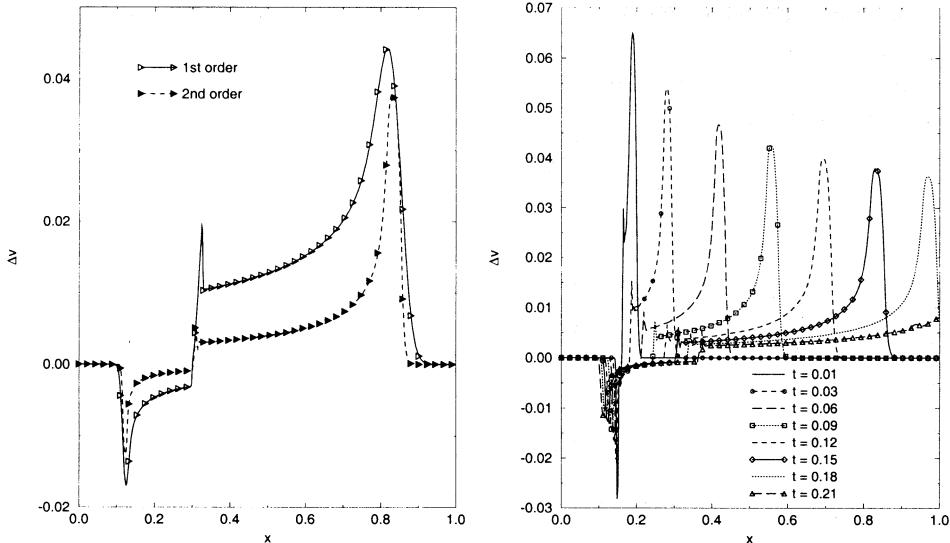


Figure 7. Moving contact velocity error, *Figure 8.* Moving contact discontinuity, order of accuracy, $\gamma_2 = 1.28\dots$

temporal evolution of velocity error

Figure 7 depicts the improvement by using a second order method. The error level is remarkably decreased with the exception of the peak values which are only slightly decreased.

Another interesting detail of the velocity error profile is the change of sign near the contact discontinuity indicating that the error is small there. This leads to the assumption that the main part of the displacement is induced during the very beginning of the computation when the γ -jump between neighboring cells is still high. Figure 8 displays the temporal evolution of the velocity error using the second order scheme for the simplified air model. The damping of the error is obvious. After a rapid change in the beginning the damping slows down. Calculations on a 10 times denser grid, which is equal to a 10 times longer running time on the coarse grid, only halves the error peaks compared to the error at a dimensionless time of $t = 0.15$. In other regions the error drops down by a factor of up to 3.5. Keeping in mind that a 10 times denser grid means a 10 times smaller time

step and therefore a computational cost increase of 100, grid refinement seems to be an expensive way to reduce the error.

4. Conclusions

No significant influence of the γ -modeling on the solution has been observed in the steady state cases and in the moving shock case. But, especially in the case of the steady state contact discontinuity, the convergence behavior is changed. In the case of a moving contact discontinuity the calorically non perfect gas introduces numerical error of the order of a few percent. The errors occur in the location of the discontinuity as well as in the pressure and velocity profiles. The position of the moving contact discontinuity is shifted into the region of lower temperature. All errors are diminished by lower γ -jumps over the cell interfaces. Therefore,

- a higher γ -jump in the initial conditions results in a larger disturbance in the solution,
- a higher order scheme decreases the overall error, but only marginally decreases the disturbance peaks,
- higher spatial resolutions as well as longer run times smear the solution over many cells and diminish the numerical error.

References

- Shyue K M (1998). An Efficient Shock-Capturing Algorithm for Compressible Multicomponent Problems. *Journal of Computational Physics*, **142**, pp 208-242.
Ton V T (1996). Improved Shock-Capturing Methods for Multicomponent and Reacting Flows. *Journal of Computational Physics*, **128**, pp 237-253.
Wada Y and Liou M-S (1994). A Flux Splitting Scheme With High-Resolution and Robustness for Discontinuities. *AIAA-94-0083*.

ORIENTATION EFFECTS ON BENT EXTRAGALACTIC JETS

STEPHEN HIGGINS

*Manchester Metropolitan University, Chester Street, Manchester,
UK*

TIM O'BRIEN

University of Manchester, Jodrell Bank, Macclesfield, UK

AND

JAMES DUNLOP

*Institute for Astronomy, University of Edinburgh, Blackford
Hill, Edinburgh*

Abstract. We have investigated how varying several parameters affects the results of a collision between an extragalactic jet and a dense, intergalactic cloud, through a series of hydrodynamic simulations. We have produced synthetic radio images for comparison with observations. These show that a variety of structures may be produced from simple jet-cloud collisions. Moderate Mach numbers and density contrasts are needed to produce observable bends. We investigate the effect of viewing from various angles on the appearance of such sources.

1. Observations of distorted jets

The jets and hotspots of radio galaxies and quasars often show complex structure. Jets can bend by over 90° and remain collimated for several jet radii (Bridle and Perley, 1984), despite the expectation that the oblique shock causing the bend should decelerate the jet (Icke, 1991). Barthel *et al.* (Barthel et al., 1988) present a large sample of quasars in which 25% showed bending greater than 20°. Explanations for these complex structures include collision with dense clouds in the ambient medium (Stocke, Burns and Christiansen, 1985; Lonsdale and Barthel, 1998).

In studies of astrophysical fluid dynamics we can only measure the emission properties of sources as projected onto the sky. We must infer flow properties from such observations. In numerical simulations we must attempt to invert this by estimating the emission properties of our solutions and considering the effect of the projection of three-dimensional solutions onto two-dimensional observations. We present the results of such studies, along with some tentative conclusions, in section 4.

2. Jets and their interaction with environment

Much of the understanding of the fluid dynamics of extragalactic jets, and of the shape and structures found in observations, has come from numerical simulations (Williams, 1991). Fully three-dimensional simulations have only become possible fairly recently (Norman, 1993), but many important results have been obtained from axisymmetric calculations. Several models for the production of complex and distorted structures in radio jets and lobes have been explored through numerical simulations including: variations in the direction of the jet at its source (Williams and Gull, 1985; Scheuer, 1982); cross-winds (Leahy, 1984); the source axis is not parallel to the axis of a spheroidal gas distribution, or the source galaxy moves through the cluster medium (Leahy and Williams, 1984); helical instabilities (Steffen et al., 1997); oblique magnetic fields in the intra-cluster medium (Koide et al., 1996). Cloud collisions are particularly applicable in cases where these bends are very sharp. Loken *et al.* (Loken et al., 1995) show that the necessary gas velocities can arise in cluster mergers, as can shocks that will bend the jet. These simulations still do not explain the sharpness of the bend.

The first investigation of the effect of off-axis jet-cloud collisions was by De Young (De Young, 1991) using the 'beam scheme' (Sanders and Prendergast, 1974). De Young observed that the jet was decelerated by the cloud. A similar interaction was investigated at higher resolution by Balsara and Norman using their RIEMANN code (Norman, 1993). They argued that a De Laval nozzle was formed which re-accelerated the jet in a new direction after impact. They did not present any results at later time to show the formation of a deflected flow pattern.

More recently Raga and Canto (Raga and Canto, 1996) have published analytical calculations and two-dimensional simulations showing bending by clouds. They conclude that slower jets will be bent more, and clouds will be eroded as jets bore through them.

3. Numerical methods

We have extended this work through a series of simulations using various sets of parameters (Higgins, 1998; Higgins, O'Brien and Dunlop, 1999). The parameters are: Mach number, and the density contrasts between of the jet and the cloud with the ambient medium. Details of the simulations are given in table 1. We have assumed conditions in the ambient medium consistent with observations: a temperature of 5×10^7 K and a particle number density of 0.01 cm^{-3} . These values are used to form dimensionless units in the computation so that model values for the ambient density and pressure in the code are set to 1.0. The jet and cloud are both taken to be in pressure balance with the ambient medium.

To calculate the synchrotron emission we need to express the magnetic field and the energy distribution in terms of the results of our hydrodynamic simulations. We can then produce synthetic radio maps by integrating this through the grid. We used the data visualization package PV-Wave to examine the simulations. This has the facility to rotate three-dimensional data sets, and hence integrate along any chosen line of sight.

4. Results

The interactions produce a variety of structures depending on the values of these parameters, so this model can be applied to many radio structures. Different structures can also be produced by a single set of parameters as the interaction progresses. Strong deflections ($\sim 90^\circ$) are difficult to sustain, producing transient structures with complex features such as double hotspots. Deflection may be easier to produce or detect in lower power jets close to the plane of the sky. This is the case for simulations 3 and 4. As the jet breaks through the cloud there are two hotspots within a boot-shaped lobe. This is a similar radio structure to 4C 29.50 (Lonsdale and Barthel,

TABLE 1. The values of the parameters for different simulations.
There is no cloud in simulations 9 – 12.

Simulation number	Jet density contrast	Cloud density contrast	Jet mach number	Jet speed
1, 2, 9	0.01	50, 200, –	2	0.07c
3, 4, 10	0.01	50, 200, –	10	0.36c
5, 6, 11	0.2	50, 200, –	2	0.02c
7, 8, 12	0.2	50, 200, –	10	0.08c

1986), with the close double hotspot. Although the cloud impact is the cause of the bending, the deflection and secondary hotspot is actually produced as the jet bends inside the distorted cocoon that has been formed during the interaction.

It is difficult to reach firm conclusions on the basis of these simulations of the kind of sources, and how many of any kind, we would expect to observe. We need to know how sources are distributed over the ranges of parameters, how the environments vary in clumpiness and density and hence what the probability of colliding with a density enhancement is. However we can make crude estimates of the likely distribution of sources by assuming fairly uniform distribution of the parameters characterising the jets and their environments. Figure 1 shows contour plots of the radio intensity of simulation 4 at two epochs (at $t = 4$ and 8). These show that the secondary hotspot that forms after impact is about one order of magnitude fainter than the primary hotspot at the impact. The line connecting the two hotspots is about 90° to the axis of the initial jet direction, which would be interpreted in observations as a 90° bend. It is about one jet radius long, but the observed width of the jet is smaller than the real width, due to limb darkening, so in observations this might be interpreted as a few jet radii. By the next epoch the secondary hotspot has faded by an order of magnitude or so, while the angle has increased to 120° .

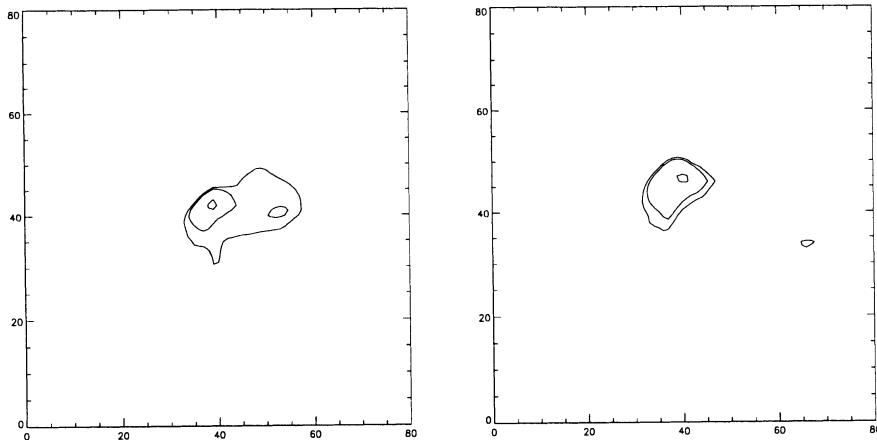


Figure 1. Contour plots of the radio intensity of simulation 4 at $t = 4$ and 8.

Clearly this would be difficult to detect without huge dynamic range (signal to noise), and the 90° structure only lasts for at most a single epoch of the interaction. This is no more than 15% of the lifetime of the interaction. This interaction is itself short lived, perhaps 10% of the typical

lifetime of a radio source (10^8 years), so we would expect only 1–2% of sources with these parameters, *viewed in the plane of the sky*, to show 90° bends.

Figure 2 shows the radio emission from simulation 4 at $t = 4$ at several orientations. Each column shows the source rotated by 30° intervals, and each row is tilted 30° toward the line of sight. The 90° bend is only visible for a few orientations. Assuming such sources are distributed isotropically with viewing angle, we would only expect to detect about 20%. Thus we would only expect to see between a third and a half a percent of sources with these parameters.

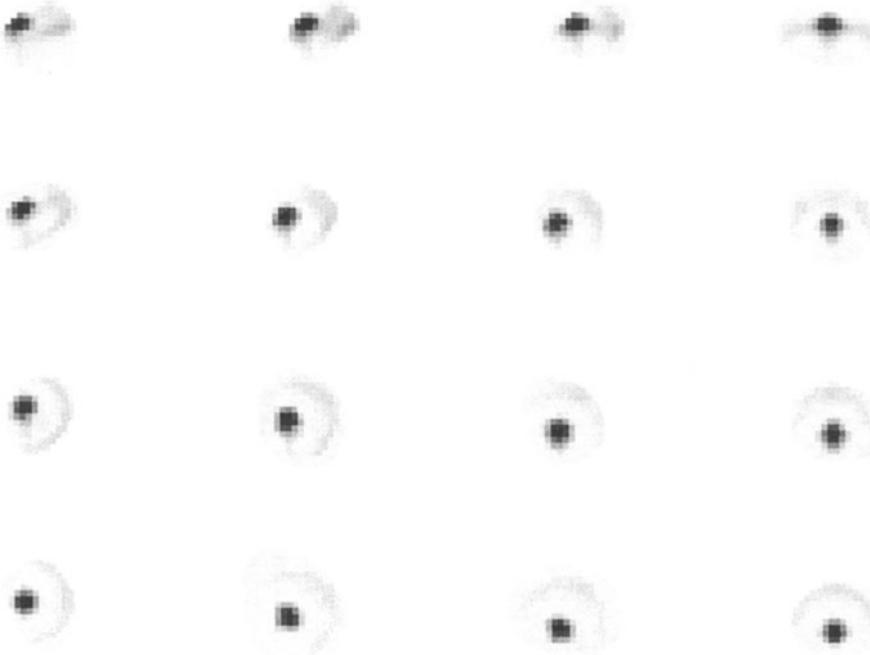


Figure 2. Integrated radio emission from simulation 4 ($M = 2, \eta_j = 0.01, \eta_c = 200.0$) at various angles.

These simulations are the only one of the four sets of jet parameters that show a 90° deflection with a secondary hotspot. If this is a representative sampling of jet parameters then we would expect at most a quarter of all sources to fall into this region. Thus we expect a total of one-tenth of a percent of all radio sources to show this sort of extreme bend.

Bent jets seem to be more common than this. It would appear that the conditions to produce bends are more common than we have assumed. The jet may interact with more than one cloud, extending the lifetime and the likelihood of observation. More detailed simulations and better statistics of such bends may allow us to estimate the number of sources with a sufficiently clumpy medium to make collisions likely. An observational study of the true statistics of bent jets may allow us to predict the number of sources with a sufficiently clumpy medium to make collisions likely.

We have simulated the passage of a jet through a medium containing an ensemble of clouds in (Higgins, O'Brien and Dunlop, 1999; Higgins, 1998). As the jet progresses through the grid it is deflected several times where it has encountered clouds, but can clearly be defined through the chain of knots, whose total lifetime is much longer.

Acknowledgments

SWH acknowledges the PPARC for receipt of a studentship. Computing was performed using the Liverpool John Moores University Starlink node. We thank Prof. Sam Falle and Dr Alan Heavens for valuable suggestions.

References

- Barthel P, Miley G, Schilizzi R and Lonsdale C. (1988). Observations of the Large-scale Radio structure in high redshift quasars. *A&AS* **73**, pp 515-547.
- Bridle A and Perley R (1984). Extragalactic jets. *ARA&A* **22**, pp 319-358.
- De Young D S (1991). The Deflection of Cosmic Jets. *ApJ* **371**, pp 69-81.
- Falle S A E G (1991). Self-similar Jets. *MNRAS* **250**, pp 581-596.
- Higgins S W, O'Brien T J and Dunlop J S (1999). Structures Produced by the Collision of Extragalactic Jets with Clouds in the Inter-galactic Medium. *MNRAS* **309**, pp 273-286.
- Higgins S W (1998). Numerical Simulations of Jet-cloud collisions and the structure of Extragalactic radio sources. Ph.D. thesis, Liverpool John Moores University
- Icke V (1991). From Nucleus to Hotspot: Nine Powers of Ten. Beams and Jets in Astrophysics. Hughes (Editor). Cambridge University Press, pp 232-277.
- Koide S, Sakai J-I, Nishikawa K-I and Mutel R (1996). Numerical Simulation of Bent Jets: Propagation into an Oblique Magnetic Field. *ApJ* **464**, pp 724-737.
- Leahy J P (1984). 3C 465: dynamics of a wide-angle-tail radio source. *MNRAS* **208**, pp 323.
- Leahy J P and Williams A G (1984). The bridges of classical double radio sources. *MNRAS* **210**, pp 929-952.
- Loken C, Roettiger K, Burns JO and Norman M (1995). Jet propagation and wide-angle tailed radio sources in merging galaxy clusters. *ApJ* **445**, pp 80-97L.
- Lonsdale C J and Barthel P D (1986). Distorted structure in the small high-redshift radio source 4C29.50. *ApJ* **303**, pp 617-623.
- Lonsdale C J and Barthel P D (1998). The Anatomy of a Radio Source Hot Spot: Very Large Baseline Array Imaging of 3C 205. *AJ* **115**, pp 895-908.
- Matthews A P and Scheuer P A G (1990). Models of radio galaxies with tangled magnetic

- fields - II: Numerical simulations and their interpretation. *MNRAS* **242**, pp 623-635.
- Norman M L (1993). Numerical simulations of astrophysical jets. *Astrophysical Jets, STScI Symp. Ser. Vol. 6*. Burgarella, Livio and O'Dea (Editors). Cambridge University Press, pp 211.
- Raga A C and Canto J (1996). The steady structure of a jet/cloud interaction - II. The case of a spherically symmetric stratification. *MNRAS* **280**, pp 567-571.
- Sanders R H and Prendergast K H (1974). The Possible Relation of the 3-KILOPARSEC Arm to Explosions in the Galactic Nucleus *ApJ* **188**, pp 489-500.
- Scheuer P A G (1982). Extragalactic Radio Sources, IAU Symp. 97. Heeschen and Wade (Editors). Reidel, Dordrecht, pp 163.
- Steffen W S, Gómez J L, Raga A C and Williams R J R (1997). Jet-Cloud Interactions and the Brightening of the Narrow-Line Region in Seyfert Galaxies *ApJL* **491** pp 73.
- Stocke J T, Burns J O and Christiansen W A (1985). VLA observations of quasars with "dogleg" radio structure. *ApJ* **299**, pp 799-813.
- Williams A G (1991). Numerical Simulations of Radio Source Structure. Beams and Jets in Astrophysics. Hughes (Editor). Cambridge University Press, pp 342-??.
- Williams A G and Gull S F (1985). Multiple Hotspots in Extragalactic Radio Sources. *Nature* **313**, pp 34-36.

OPERATOR SPLITTING FOR CONVECTION-DOMINATED NONLINEAR PARTIAL DIFFERENTIAL EQUATIONS

HELGE HOLDEN

*Department of Mathematical Sciences,
Norwegian University of Science and Technology,
N-7491 Trondheim, Norway.
e-mail: holden@math.ntnu.no*

KENNETH HVISTENDAHL KARLSEN

*Department of Mathematics,
University of Bergen,
Johs. Brunsgr. 12,
N-5008 Bergen, Norway.
e-mail: kennethk@mi.uib.no*

KNUT-ANDREAS LIE

*Department of Informatics,
University of Oslo,
P.O. Box 1080 Blindern,
N-0316 Oslo, Norway.
e-mail: kalie@ifi.uio.no*

AND

NILS HENRIK RISEBRO

*Department of Mathematics,
University of Oslo,
P.O. Box 1053 Blindern,
N-0316 Oslo, Norway.
e-mail: nilshr@math.uio.no*

1. Introduction

We describe an efficient solution strategy for nonlinear systems of partial differential equations of the form

$$U_t + \sum_i F_i(U)_{x_i} = \sum_{i,j} D_{ij}(U)_{x_i x_j} + G(U), \quad U|_{t=0} = U_0. \quad (1)$$

We explicitly allow for degeneracy of the viscous term in the sense that we only require $\sum_{i,j} D'_{ij}(u) \xi_i \xi_j \geq 0$. The solution strategy is based on operator splitting where an abstractly defined Cauchy problem $U_t + \mathcal{A}(U) = 0$, is split into simpler problems $V_t^l + \mathcal{A}^l(V^l) = 0$, by writing $\mathcal{A} = \mathcal{A}^1 + \dots + \mathcal{A}^\ell$. If the solution of subproblem l is written $V^l(t) = S_{\Delta t}^l V_0^l$ then the idea is that an approximate solution of the original problem reads

$$U(n\Delta t) \approx U^n = [S_{\Delta t}^\ell \circ \dots \circ S_{\Delta t}^1]^\ell U^0, \quad t = n\Delta t.$$

From a numerical point of view, the idea behind operator splitting is to combine efficient and accurate numerical methods for each of the subproblems to build an overall solution strategy. This allows for a solver with great modularity, where each component can easily be replaced. Operator splitting has therefore been a common strategy over the last decades.

Operator splitting may introduce new numerical difficulties and lead to the computation of nonphysical numerical artifacts. Lately, much of the development of numerical methods within the scientific computing community therefore aims at incorporating as many of the physical effects as possible into a single method. On the other hand, there is an increasing influx of modern software tools such as object-orientation into scientific computing. In object-orientation one often seeks to identify simple and generic objects that can be combined to solve complicated problems. The generic objects can then be implemented, tested and verified independently and reused in other settings. In this sense, operator splitting is an object-oriented approach to solving (1). However, equipped with a large toolbox of small parts that can be fitted together independently (like Lego pieces), the natural question is whether the new creation will work as planned?

This question is normally answered in the form of a mathematical convergence theory; an abstract and general convergence theory that includes a large number of previous specific splitting methods for (1) is presented in (Holden et al., 1999).

Theoretical convergence is not sufficient from a practical point of view; accuracy and efficiency for given discretization parameters are often more important. We therefore present several splitting methods developed especially to capture sharp gradient variations in the solutions of (1). The distinct feature of our approach is the use of a large-time-step front-tracking

method (Holden and Holden, 1998; Risebro, 1993; Bressan, 1992) to solve one-dimensional hyperbolic subproblems of the form $U_t + F(U)_x = 0$. The integrated method is unconditionally stable and delivers more than the standard resolution with surprisingly high efficiency. Our methods have been applied to specific problems arising when simulating flow in porous media, gas dynamics, shallow water waves, glaciers, traffic flow, and sedimentation.

2. Operator Splitting Methods

In the following we present some examples of efficient operator splitting methods based on front tracking, which we introduce first.

Front Tracking. The term ‘front-tracking’ is applied to a wide variety of methods with the common feature that they seek to perform explicit tracking of discontinuities in hyperbolic solutions. Our method originates from an idea by Dafermos. Consider the conservation law

$$U_t + F(U)_x = 0, \quad U(x, 0) = U_0(x). \quad (2)$$

Making the usual piecewise constant approximation to the initial data, the Cauchy problem is converted into a sequence of Riemann problems. The essence of the method is to approximate the solution of each Riemann problem by a step function, i.e., by a set of constant states separated by space-time rays of discontinuity (*fronts*), and track the discontinuities explicitly. Each time two or more fronts collide, they define a new Riemann problem which is approximated by piecewise constants, and so on.

For scalar equations, the approximation of Riemann problems is typically achieved by making a piecewise linear approximation to the flux function. In the systems case, one retains shocks and constants and approximates rarefaction waves. The front tracking method is unconditionally stable and has first order convergence with respect to the approximation of the initial data and the Riemann problems, see (Holden and Holden, 1998).

Figure 1 shows the front tracking approximation for a scalar problem (2) with flux function $f(u) = 2u^2(1 - u^2)$ and initial data $u_0(x) = \sin(\pi x)\chi_{[-1,1]}(x)$.

The front tracking method is easily extended to quasilinear equations with variable coefficients $u_t + V(x, t)f(u)_x = 0$ by introducing e.g., a polynomial approximation to the velocity field (Lie, 1999).

Example 1. The natural extension of front tracking to multidimensions is by dimensional splitting on a Cartesian grid (Holden and Risebro, 1993). Since front tracking is unconditionally stable, the step size in the splitting is not limited by a stability condition. Instead, the limiting factor lies in the two error mechanisms; temporal splitting errors that increase with Δt

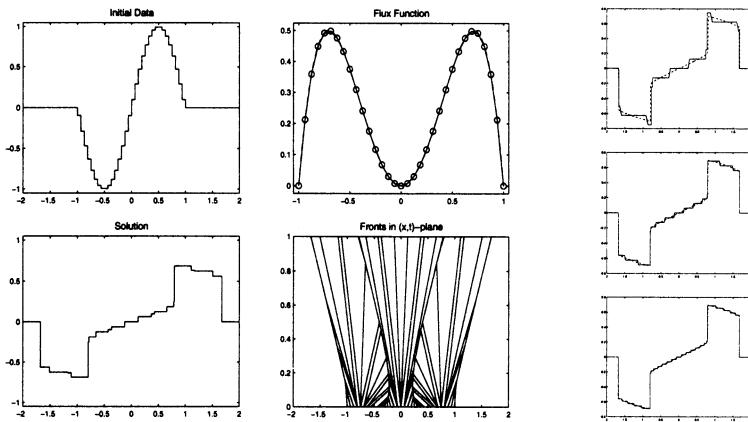


Figure 1. (Left) Front tracking solution for a scalar 1-D problem with $n = 64$ intervals in approximation of u_0 and $m = 32$ intervals for f . (Right) Refinement of flux approximation: $m = 16, 32$, and 64 from top to bottom for $n = 2048$.

and spatial errors from the projections that decrease with Δt , see Figure 2. The theoretical convergence order is one half (Lie, Haugse, and Karlsen, 1998) and is observed for *linear* problems, see (Lie, 1999). For nonlinear problems, convergence of order one is observed (Lie, Haugse, and Karlsen, 1998; Lie, 1999) due to nonlinear self-sharpening effects that counteract the smearing from the projections.

For many problems one can therefore use surprisingly large CFL numbers. Figure 3 illustrates this for Burgers' equation with a variable velocity field $V = (\cos(\pi(y+t)), \sin(\pi(x+t)))$. Although not all details are resolved accurately, the qualitative representation is quite good. Extensive numerical experiments indicate that the method gives best performance for CFL numbers around 10-20. The large-step ability of the splitting method makes it very efficient, typically a factor 40-50 faster than standard high-resolution method when comparing runtime versus accuracy, see (Lie, Haugse, and Karlsen, 1998; Lie, 1999).

Example 2. The next example illustrates the use of dimensional splitting to solve multidimensional systems by front tracking, see (Holden, Lie, and Risebro, 1999) for more examples. Here we consider the Euler equations of gas dynamics describing an ideal, polytropic gas with gas constant 1.4. Our test case is a cylindrical Riemann problem between two horizontal walls, see Figure 4. The initial Riemann problem leads to an inward moving rarefaction and a strong outward moving shock followed by a contact. The first shock reflects at the lower wall. At time $t = 0.2$ the reflected shock has passed through the contact and into the low-density region behind, where the shock speed increases. The rarefaction wave implodes on the cylinder center and produces an outward moving shock.

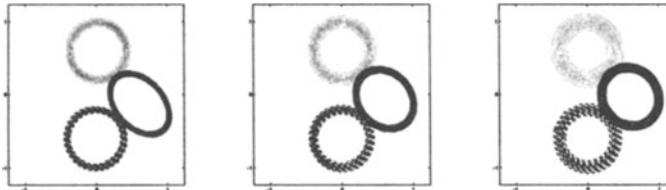


Figure 2. Solid body rotation around the origin ($u_t + yu_x - xu_y = 0$) of a cylinder of radius 0.4 centred at (0,0.6). Solutions at time $t = \pi/2$ (solid), π (dashed), and 2π (dotted) computed with 16 (left), 32, and 64 splitting steps (right) and $\Delta x = 0.025$. The corresponding CFL numbers are 20, 10, and 5.

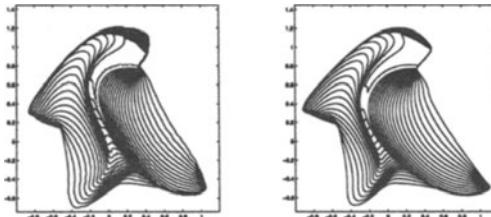


Figure 3. Solution of $u_t + \nabla \cdot (u^2 V) = 0$ at time $t = 1.0$ computed with 8 splitting steps and $\Delta x = 0.01$ (left) corresponding to CFL number $\nu \approx 25$ compared with a fine grid solution (right). The initial data equals one inside a square with sides lengths 0.5 centered at the origin and zero outside.

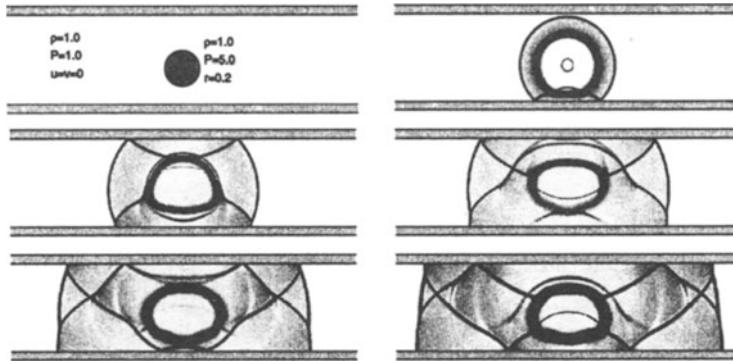


Figure 4. Evolution of a high pressure cylinder between two infinite walls; initial setup and emulated Schlieren images at times $t = 0.2, 0.4, 0.6, 0.8$, and 1.0.

For systems, the generation of small-scale oscillations prevents the use of very large time steps and the method performs best at CFL numbers moderately above unity (typically 1-4), see (Holden, Lie, and Risebro, 1999).

Example 3. Operator splitting methods are often applied to solve

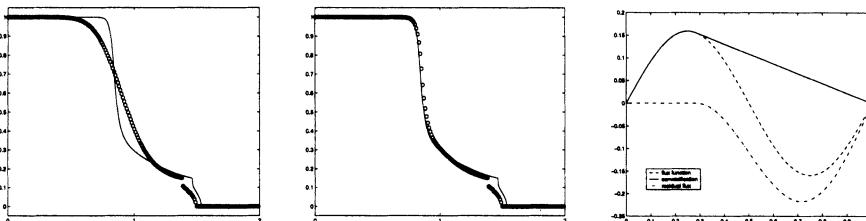


Figure 5. Degenerate convection-diffusion equation with $f(u) = \sin(2\pi u)/2\pi$, $D'(u) = \chi_{[0,15,\infty)}(u)$, and $u_0(x) = \chi_{(-\infty,1]}(u)$. (Left) Operator splitting solution at time $t = 0.5$ computed with one step. (Middle) Corrected operator splitting. (Right) Convexification of flux function and the corresponding residual.

convection-diffusion equations on the form $u_t + f(u)_x = D(u)_{xx}$. A straightforward splitting decomposes the equation into a hyperbolic part $u_t + f(u)_x = 0$ and a diffusive part $u_t = D(u)_{xx}$. For convection dominated problems this may lead to significant smearing of sharp gradients unless the splitting step is chosen very small. This viscous splitting error comes from the entropy loss imposed by local linearization of f during the hyperbolic step (Karlsen and Risebro, 2000). The entropy loss can be identified in the form of local residual fluxes that ensure the correct amount of self sharpening in shock layers. By including these residuals in the splitting, e.g., in the form of a modified diffusive step $u_t + f_{\text{res}}(u; x)_x = D(u)_{xx}$, one obtains a robust splitting strategy, called *corrected operator splitting*, see (Karlsen and Risebro, 2000; Espedal and Karlsen, 2000). This is illustrated in Figure 5, where the shock layer is resolved correctly by the modified splitting. However, the interface between the hyperbolic and the parabolic region, which depends on $f(u)$ and $D'(u)$, is not resolved by one splitting step. An overview of general splitting methods are given in (Espedal and Karlsen, 2000).

Example 4. The same splitting strategy can be applied in multidimensions, either using first dimensional splitting and then viscous splitting for the one-dimensional problems, or first viscous splitting and the dimensional splitting for the multidimensional hyperbolic problem. We adopt the latter strategy and consider two-phase flow in a reservoir. This process is governed by an elliptic pressure equation $\nabla(\lambda\nabla p) = 0$ that is coupled to a saturation equation $s_t + V \cdot \nabla f(s) = \varepsilon \nabla(Kd(s)\nabla s)$ through the Darcy velocity $V = -\lambda\nabla p$. The system is typically solved by a sequential splitting; first solve for p and compute V , then V is held fixed as s is updated, and so on. Figure 6 shows the quarter five-spot test case with a heterogeneous permeability field K , i.e., the first quadrant of a periodic well configuration with a water injection well in the origin and production wells at $(\pm 1, \pm 1)$.



Figure 6. Operator splitting solution with CFL number 4.0 and 16 sequential steps. The flux function is $f(s) = s^2/(s^2 + 0.125(1-s)^2)$, the capillary diffusion function $d(s) = 4s(1-s)$, and $\varepsilon = 0.005$. Water breaks through in the production well after the injection of 0.306 pore volumes of water.

References

- Bressan A (1992). Global solutions of systems of conservation laws by wave-front tracking. *J. Math. Analysis and App.*, **170**: 414–432.
- Espedal M S and Karlsen K H. Numerical solution of reservoir flow models based on large time step operator splitting algorithms. To appear in A. Fasano and H. van Duijn, editors, *Filtration in Porous Media and Industrial Applications*, Lecture Notes in Mathematics, Springer Verlag.
- Holden H and Holden L (1998). On scalar conservation laws in one-dimension. In *Ideas and Methods in Mathematics and Physics*, pages 480–509. Cambridge University Press.
- Holden H, Karlsen K H, Lie K-A, and Risebro N H (1999). Operator splitting for nonlinear partial differential equations: An L^1 convergence theory. In preparation.
- Holden H, Lie K-A, Risebro N H (1999). An unconditionally stable method for the Euler equations. *J. Comput. Phys.*, **150**(1): 76–96.
- Holden H and Risebro N H (1993). A method of fractional steps for scalar conservation laws without the CFL condition. *Math. Comp.*, **60**(201): 221–232.
- Karlsen K H and Risebro N H (2000). Corrected operator splitting for nonlinear parabolic equations. *SIAM J. Numer. Anal.*, **37**(3): 980–1003.
- Lie K-A, Haugse V, and Karlsen K H (1998). Dimensional splitting with front tracking and adaptive local grid refinement. *Numer. Methods Partial Differential Equations*, **14**(5): 627–648.
- Lie K A (1999). A dimensional splitting method for quasilinear hyperbolic equations with variable coefficients. *BIT*, **39**(4): 683–700.
- Risebro N H (1993). A front-tracking alternative to the random choice method. *Proc. Amer. Math. Soc.*, **117**(4): 1125–1139.

BALANCING SOURCE TERMS AND FLUX GRADIENTS IN FINITE VOLUME SCHEMES

M. E. HUBBARD

DAMTP, Silver Street, Cambridge, CB3 9EW, U.K.

Email: M.E.Hubbard@damtp.cam.ac.uk

AND

P. GARCIA-NAVARRO

Fluid Mechanics, CPS, University of Zaragoza, 50015, Spain.

Email: pigar@posta.unizar.es

1. Introduction

In the field of computational hydraulics the modelling can be dominated by the effects of source terms and in some cases, quantities which vary spatially but independently of the flow variables. This paper is concerned with the shallow water equations and how the additional terms should be discretised, given that Roe's scheme has been used to approximate the flux terms, extending recent research by other authors (Glaister, 1992; Vázquez-Cendón, 1999; Bermúdez and Vázquez, 1994). In each of these papers the discrete form of the source terms has been deliberately constructed along similar lines to the numerical fluxes. This is done to ensure that equilibria which occur in the mathematical model are retained by the numerical model, and that in the absence of additional terms, the conservative fluxes are retrieved for accurate modelling of discontinuous solutions. However, all previous work deals only with the first order scheme. In this paper the extension of these ideas to higher order Total Variation Diminishing (TVD) versions of Roe's scheme (using both flux limiting and slope limiting techniques) is described. It is then possible to construct a source term approximation which has each of the above properties on all types of regular and irregular grids in any number of dimensions; see (Hubbard and Garcia-Navarro, 1999) for details. Furthermore, following on from (Garcia-Navarro and Vázquez-Cendón, 1997), a new formulation is presented for the discretisation of the flux in the case where it depends on a spatially varying

quantity which is independent of the solution. The one-dimensional shallow water equations have been chosen to demonstrate the effectiveness of these new techniques, by modelling the effects of a sloping bed and the inclusion of breadth variation in open channel flows.

2. The General Discretisation

The one-dimensional equations representing a general system of conservation laws with source terms may be written

$$\underline{U}_t + \underline{F}_x = \underline{S}, \quad (1)$$

where \underline{U} is the vector of conservative variables, \underline{F} is the conservative flux vector and \underline{S} includes all of the source terms. The flux is assumed to depend not only on the conservative variables but also another independent, spatially varying quantity, denoted here by $B(x)$, *i.e.* $\underline{F} = \underline{F}(\underline{U}, B(x))$.

Using the standard (cell centre) finite volume approximation of the flux terms in (1) with forward Euler time-stepping gives

$$\underline{U}_i^{n+1} = \underline{U}_i^n - \frac{\Delta t}{\Delta x_i} \left(\underline{F}_{i+\frac{1}{2}}^* - \underline{F}_{i-\frac{1}{2}}^* \right) + \frac{\Delta t}{\Delta x_i} \underline{S}_i^*, \quad (2)$$

in which \underline{F}^* represents a numerical flux evaluated at an interface between cells and $\underline{S}^* \approx \int \underline{S} dx$ is a numerical source integral over the cell.

Roe's scheme (Roe, 1981) is used here to discretise the flux derivatives, with a minor modification to take into account the dependence of the flux on $B(x)$. This approximate Riemann solver splits the flux difference at an interface into independent components, giving

$$\begin{aligned} \Delta \underline{F}_{i+\frac{1}{2}} &= (\tilde{\mathbf{A}} \Delta \underline{U} + \tilde{\underline{V}})_{i+\frac{1}{2}} = (\tilde{\mathbf{R}} \tilde{\Lambda} \tilde{\mathbf{R}}^{-1} \Delta \underline{U} + \tilde{\mathbf{R}} \tilde{\Lambda} \tilde{\mathbf{R}}^{-1} \tilde{\underline{V}})_{i+\frac{1}{2}} \\ &= \left(\sum_{k=1}^{N_w} \tilde{\alpha}_k \tilde{\lambda}_k \tilde{r}_k + \sum_{k=1}^{N_w} \tilde{\gamma}_k \tilde{r}_k \right)_{i+\frac{1}{2}}, \end{aligned} \quad (3)$$

where $\tilde{\underline{V}} \approx \frac{\partial \underline{F}}{\partial \underline{B}} \Delta B$ (so it reduces to the standard Roe flux difference splitting when B is constant). $\Delta \underline{F}$ represents the jump in \underline{F} across the edge of a grid cell, $\tilde{\mathbf{R}}$ is the matrix whose columns are the right eigenvectors \tilde{r}_k of $\tilde{\mathbf{A}}$, the approximate flux Jacobian, $\tilde{\Lambda}$ is the diagonal matrix of eigenvalues $\tilde{\lambda}_k$ of $\tilde{\mathbf{A}}$, and the components of $\tilde{\mathbf{R}}^{-1} \Delta \underline{U}$ are the 'strengths' $\tilde{\alpha}_k$ associated with each component of the decomposition. Additionally, $\tilde{\gamma}_k$, the coefficients of the decomposition of the extra term, are the components of $\tilde{\mathbf{R}}^{-1} \tilde{\underline{V}}$. The final expression in (3) indicates how the flux difference is decomposed into N_w characteristic components, N_w being the number of equations in the

system (1). Throughout $\tilde{\cdot}$ denotes the evaluation of a quantity at its Roe-average state (Roe, 1981), calculated specifically so that (3) is satisfied.

The numerical fluxes which are used in (2) are simply

$$\underline{F}_{i+\frac{1}{2}}^* = \frac{1}{2} (\underline{F}_{i+1} + \underline{F}_i) - \frac{1}{2} \left(\tilde{\mathbf{R}} |\tilde{\Lambda}| \tilde{\mathbf{R}}^{-1} \Delta \underline{U} + \tilde{\mathbf{R}} \operatorname{sgn}(\mathbf{I}) \tilde{\mathbf{R}}^{-1} \tilde{\underline{V}} \right)_{i+\frac{1}{2}}, \quad (4)$$

in the first order case, where $|\tilde{\Lambda}| = \operatorname{diag}(|\tilde{\lambda}_k|)$ and $\operatorname{sgn}(\mathbf{I}) = \tilde{\Lambda}^{-1} |\tilde{\Lambda}|$. When a flux limited high resolution scheme is being used,

$$\underline{F}_{i+\frac{1}{2}}^* = \frac{1}{2} (\underline{F}_{i+1} + \underline{F}_i) - \frac{1}{2} \left(\tilde{\mathbf{R}} |\tilde{\Lambda}| \mathbf{L} \tilde{\mathbf{R}}^{-1} \Delta \underline{U} + \tilde{\mathbf{R}} \operatorname{sgn}(\mathbf{I}) \mathbf{L} \tilde{\mathbf{R}}^{-1} \tilde{\underline{V}} \right)_{i+\frac{1}{2}}, \quad (5)$$

in which, additionally, $\mathbf{L} = \operatorname{diag}(1 - L(r_k)(1 - |\nu_k|))$, where $\nu_k = \tilde{\lambda}_k \Delta t / \Delta x$ is the Courant number associated with the k^{th} component of the decomposition, L is a nonlinear flux limiter function, and $r_k = \tilde{\alpha}_k^{\text{upwind}} / \tilde{\alpha}_k^{\text{local}}$. If the high resolution scheme being used employs slope limiters then

$$\underline{F}_{i+\frac{1}{2}}^* = \frac{1}{2} \left(\underline{F}_{i+\frac{1}{2}}^{\text{R}} + \underline{F}_{i+\frac{1}{2}}^{\text{L}} \right) - \frac{1}{2} \left(\tilde{\mathbf{R}} |\tilde{\Lambda}| \tilde{\mathbf{R}}^{-1} \Delta \underline{U} + \tilde{\mathbf{R}} \operatorname{sgn}(\mathbf{I}) \tilde{\mathbf{R}}^{-1} \tilde{\underline{V}} \right)_{i+\frac{1}{2}}, \quad (6)$$

where the superscripts $^{\text{R}}$ and $^{\text{L}}$ represent evaluation on, respectively, the right and left hand sides of the interface indicated by the associated subscript, so the averages (\cdot) are now calculated from a linear reconstruction of the solution.

Following the work of (Glaister, 1992; Bermúdez and Vázquez, 1994; García-Navarro and Vázquez-Cendón, 1997), the approximate source term integral associated with an edge of a cell is similarly projected on to the eigenvectors of the flux Jacobian, so that in its linearised form it becomes

$$\int_{x_i}^{x_{i+1}} \underline{S} \, dx \approx \tilde{\underline{S}}_{i+\frac{1}{2}} = \left(\tilde{\mathbf{R}} \tilde{\mathbf{R}}^{-1} \tilde{\underline{S}} \right)_{i+\frac{1}{2}} = \left(\sum_{k=1}^{N_w} \tilde{\beta}_k \tilde{\underline{r}}_k \right)_{i+\frac{1}{2}}, \quad (7)$$

where $\tilde{\beta}_k$, the coefficients of the decomposition, are the components of the vector $\tilde{\mathbf{R}}^{-1} \tilde{\underline{S}}$. \underline{S}_i^* will be constructed out of contributions from both ends of the cell, with consistency assured as long as the whole of each dual cell integral (7) is distributed.

In order to obtain the discrete balance which is required between the flux and source terms the numerical source term integral of (2) is approximated by

$$\underline{S}_i^* = \underline{S}_{i+\frac{1}{2}}^- + \underline{S}_{i-\frac{1}{2}}^+, \quad (8)$$

in which the edge contributions are given by

$$\begin{aligned}\underline{\mathbf{S}}_{i+\frac{1}{2}}^* - &= \frac{1}{2} \left(\tilde{\mathbf{R}}(\mathbf{I} - \text{sgn}(\mathbf{I})) \tilde{\mathbf{R}}^{-1} \tilde{\underline{\mathbf{S}}} \right)_{i+\frac{1}{2}}, \\ \underline{\mathbf{S}}_{i-\frac{1}{2}}^* + &= \frac{1}{2} \left(\tilde{\mathbf{R}}(\mathbf{I} + \text{sgn}(\mathbf{I})) \tilde{\mathbf{R}}^{-1} \tilde{\underline{\mathbf{S}}} \right)_{i-\frac{1}{2}}.\end{aligned}\quad (9)$$

It is now simple to make high resolution corrections to the numerical source terms which will maintain the required balance. In the flux limiting case this leads to replacing (9) with

$$\underline{\mathbf{S}}_{i+\frac{1}{2}}^* - = \frac{1}{2} \left(\tilde{\mathbf{R}}(\mathbf{I} - \text{sgn}(\mathbf{I})\mathbf{L}) \tilde{\mathbf{R}}^{-1} \tilde{\underline{\mathbf{S}}} \right)_{i+\frac{1}{2}}, \quad (10)$$

and a similar expression for $\underline{\mathbf{S}}_{i-\frac{1}{2}}^* +$. When slope limiters are applied an appropriate correction to the numerical source within each cell is given by

$$\underline{\mathbf{S}}_i^* = \left(\underline{\mathbf{S}}_{i+\frac{1}{2}}^* - + \underline{\mathbf{S}}_{i-\frac{1}{2}}^* + \right) - \tilde{\underline{\mathbf{S}}} \left(\underline{U}_{i+\frac{1}{2}}^L, \underline{U}_{i-\frac{1}{2}}^R \right). \quad (11)$$

The first term on the right hand side is evaluated precisely as before, in (8), except that the interface values are now those of the linear reconstruction of the solution within each cell. $\tilde{\underline{\mathbf{S}}}$ is simply the source term integral approximated over the mesh cell and hence evaluated at the Roe-average of the left and right states of the linear reconstruction of the solution within that cell.

3. Shallow water flows

In one dimension, shallow water flow through a rectangular open channel of varying breadth and bed slope is modelled by the equations

$$\begin{pmatrix} bd \\ bdu \end{pmatrix}_t + \begin{pmatrix} bdu \\ bdu^2 + \frac{1}{2}gbd^2 \end{pmatrix}_x = \begin{pmatrix} 0 \\ \frac{1}{2}gd^2b_x - gbdz_x \end{pmatrix}, \quad (12)$$

which, when compared with (1) to find \underline{U} , \underline{F} and \underline{S} , ultimately leads to

$$\frac{\partial \underline{F}}{\partial \underline{U}} = \begin{pmatrix} 0 & 1 \\ gd - u^2 & 2u \end{pmatrix}, \quad \frac{\partial \underline{F}}{\partial B} = \begin{pmatrix} 0 \\ -\frac{1}{2}gd^2 \end{pmatrix}. \quad (13)$$

In these equations d is the depth of the flow, z is the height of the bed above a nominal zero level, $b = b(x)$ is the channel breadth, u is the flow velocity, and g is the acceleration due to gravity.

The characteristic decomposition (3) for the equations (12) and (13) is completely defined by

$$\begin{aligned}\tilde{\alpha}_1 &= \frac{\Delta(bd)}{2} + \frac{1}{2\tilde{c}} (\Delta(bdu) - \tilde{u} \Delta(bd)) \\ \tilde{\alpha}_2 &= \frac{\Delta(bd)}{2} - \frac{1}{2\tilde{c}} (\Delta(bdu) - \tilde{u} \Delta(bd)) \\ \tilde{\lambda}_1 &= \tilde{u} + \tilde{c}, \quad \tilde{\lambda}_2 = \tilde{u} - \tilde{c}, \quad \tilde{\gamma}_1 = -\frac{1}{4g} \tilde{c}^3 \Delta b, \quad \tilde{\gamma}_2 = \frac{1}{4g} \tilde{c}^3 \Delta b \\ \tilde{r}_1 &= \begin{pmatrix} 1 \\ \tilde{u} + \tilde{c} \end{pmatrix}, \quad \tilde{r}_2 = \begin{pmatrix} 1 \\ \tilde{u} - \tilde{c} \end{pmatrix},\end{aligned}\quad (14)$$

and it is easily shown that (3) is satisfied exactly when

$$\tilde{u} = \frac{\sqrt{b^R d^R} u^R + \sqrt{b^L d^L} u^L}{\sqrt{b^R d^R} + \sqrt{b^L d^L}}, \quad \tilde{c}^2 = g \left(\frac{\sqrt{b^R} d^R + \sqrt{b^L} d^L}{\sqrt{b^R} + \sqrt{b^L}} \right), \quad (15)$$

which reduce to standard Roe-averages for one-dimensional shallow water flow in the absence of breadth variation (*i.e.* when $b^R = b^L$). The corresponding decomposition of the source terms (7) then leads to

$$\tilde{\beta}_1 = \frac{1}{4g} \tilde{c}^3 \Delta b - \frac{1}{2} \tilde{b} \tilde{c} \Delta z = -\tilde{\beta}_2. \quad (16)$$

In order for (9) to maintain the correct balance when the flow is quiescent, \tilde{b} is constructed so that it satisfies

$$\tilde{b} \Delta z = \Delta(bz) - \tilde{z} \Delta b \quad \text{where} \quad \tilde{z} = K - \frac{\sqrt{b^R} d^R + \sqrt{b^L} d^L}{\sqrt{b^R} + \sqrt{b^L}}, \quad (17)$$

K being the height of the still water surface above the nominal zero level.

4. Numerical results

The upwind source term discretisation described above maintains still water to machine accuracy for an indefinite period for any test case geometry for both first and higher order schemes (unlike most standard approximations) so no results of this type are presented here. Instead a ‘tidal’ flow is modelled in a channel of varying breadth and depth, and compared with an asymptotically exact solution, described fully in (Vázquez-Cendón, 1999). The comparison is made between first order, slope limited and flux limited schemes combined with pointwise and upwind source term discretisations: in all high resolution cases the Minmod limiter has been applied. The ‘exact’ and numerical solutions (all computed on the same regular 600 cell

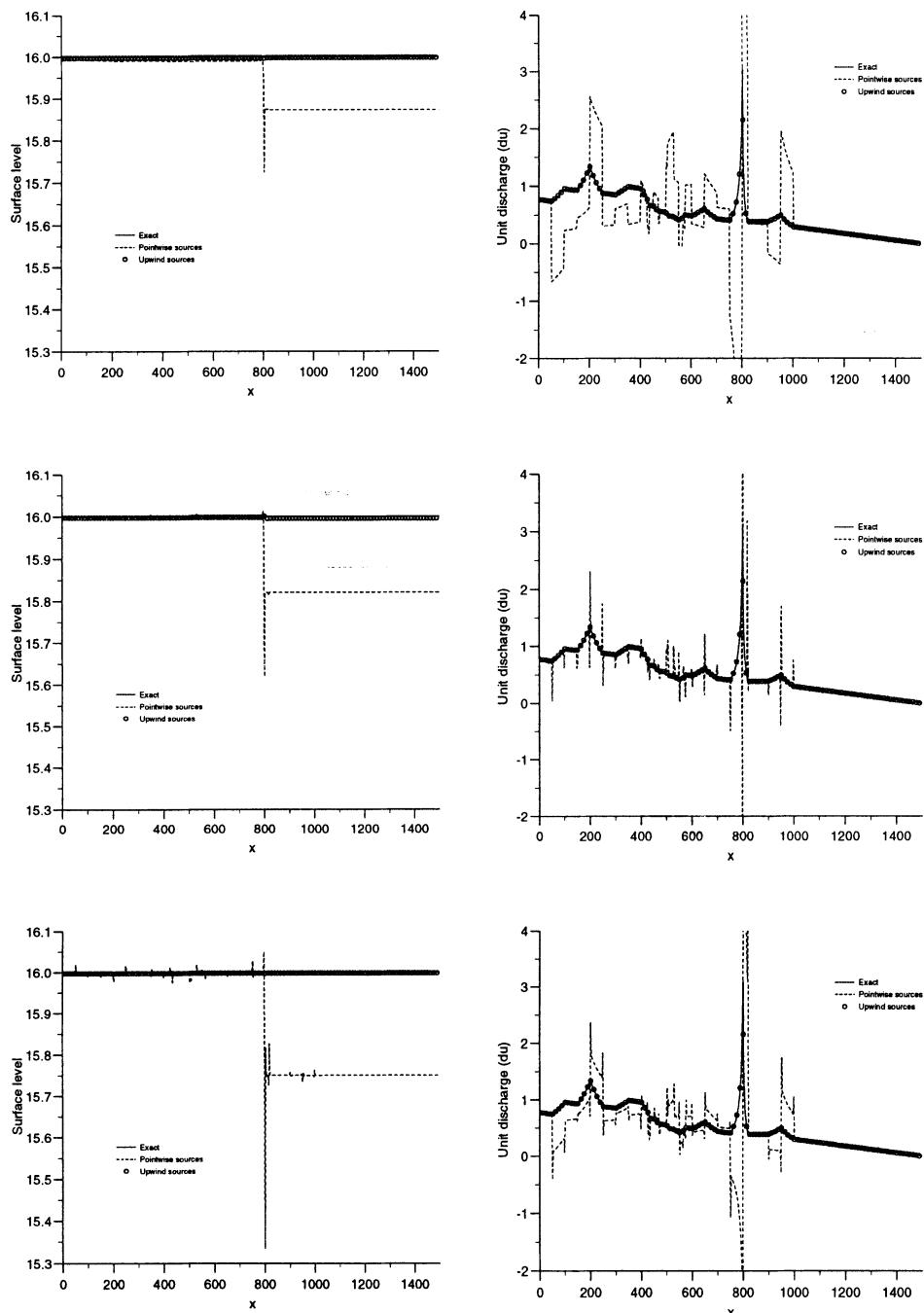


Figure 1. Water surface level and unit discharge for the tidal flow test case for first order (top) and high resolution slope limited (centre) and flux limited (bottom) schemes.

grid) to this problem when $t = 10800$ ('high tide') are compared in Figure 1. The agreement is very close, not only for the first order scheme, but also for both of the higher order schemes when the upwind source discretisation is used. However, as in the still water tests, the pointwise source discretisation gives, at best, only a reasonable approximation to the depth, and a very poor prediction of the flow velocity.

5. Conclusions

A new method has been presented for the discretisation of source terms which provide a balance with flux derivatives in nonlinear systems of conservation laws, extending the work of (Glaister, 1992; Bermúdez and Vázquez, 1994; Garcia-Navarro and Vázquez-Cendón, 1997) to high order TVD versions of Roe's scheme, using both flux and slope limiters. A technique for discretising fluxes which can vary independently of the flow variables is also suggested. The methods have been shown to be effective in the modelling of the one-dimensional shallow water equations (on the understanding that the TVD condition which underlies the limiting procedures is derived for homogeneous equations), and they also readily generalise for use on arbitrary polygonal meshes in any number of dimensions (Hubbard and Garcia-Navarro, 1999).

Acknowledgements

The work of the first author was funded by the EPSRC grant GK/K74616.

References

- Bermúdez A and Vázquez M E (1994). Upwind Methods for Hyperbolic Conservation Laws with Source Terms. *Computers and Fluids*, **23**(8), pp 1049-1071.
- Garcia-Navarro P and Vázquez-Cendón M E (1997). Some Considerations and Improvements on the Performance of Roe's Scheme for 1D Irregular Geometries. Internal Report 23, Departamento de Matemática Aplicada, Universidade de Santiago do Compostela.
- Glaister P (1992). Prediction of Supercritical Flow in Open Channels. *Comput. Math. Applic.*, **24**(7), pp 69-75.
- Hubbard M E and Garcia-Navarro P (1999). Flux Difference Splitting and the Balancing of Source Terms and Flux Gradients. Report NA-3/99, Department of Mathematics, University of Reading (submitted to *J. Comput. Phys.*).
- Roe P L (1981). Approximate Riemann Solvers, Parameter Vectors, and Difference Schemes. *J. Comput. Phys.*, **43**(2), pp 357-372.
- Vázquez-Cendón M E (1999). Improved Treatment of Source Terms in Upwind Schemes for the Shallow Water Equations in Channels with Irregular Geometry. *J. Comput. Phys.*, **148**(2), pp 497-526.

RIEMANN SOLVERS IN GENERAL RELATIVISTIC HYDRODYNAMICS

J. M^A. IBÁÑEZ, M.A. ALOY

*Departamento de Astronomía y Astrofísica,
UVEG, 46100 Burjassot, Spain.*

Emails: Jose.M.Ibanez@uv.es, Miguel.A.Aloy@uv.es

J. A. FONT

*Max-Planck-Institut für Astrophysik,
85748 Garching, Germany.
Email: font@mpa-garching.mpg.de*

AND

J. M^A. MARTÍ, J. A. MIRALLES, J. A. PONS

*Departamento de Astronomía y Astrofísica,
UVEG, 46100 Burjassot, Spain.
Emails: Jose.M.Marti@uv.es, Juan.A.Miralles@uv.es,
Jose.A.Pons@uv.es*

Abstract.

Our contribution concerns with the numerical solution of the 3D general relativistic hydrodynamical system of equations within the framework of the $\{3 + 1\}$ formalism. We summarize the theoretical ingredients which are necessary in order to build up a numerical scheme based on the solution of local Riemann problems. Hence, the full spectral decomposition of the Jacobian matrices of the system, i.e., the eigenvalues and the right and left eigenvectors, is explicitly shown. An alternative approach consists in using any of the special relativistic Riemann solvers recently developed for describing the evolution of special relativistic flows. Our proposal relies on a local change of coordinates in terms of which the spacetime metric is locally Minkowskian and permits an accurate description of numerical general relativistic hydrodynamics.

1. Introduction

Astrophysical scenarios involving relativistic flows have drawn the attention and efforts of many researchers since the pioneering studies of (May and White, 1967) and (Wilson, 1972). Relativistic jets, accretion onto compact objects (in X-ray binaries or in the inner regions of active galactic nuclei), stellar core collapse, coalescing compact binaries (neutron star and/or black holes) and recent models of formation of Gamma-ray bursts (GRBs) are examples of systems in which the evolution of matter is described within the frame of the theory of relativity (special or general).

Since 1991 (Martí, Ibáñez and Miralles, 1991) the use of Riemann solvers in relativistic hydrodynamics has proved successful in handling complex flows, with high Lorentz factors and strong shocks, superseding traditional methods which failed to describe ultrarelativistic flows (Norman and Winkler, 1986). Exploiting the hyperbolic and conservative character of the relativistic hydrodynamical equations, we proposed how to extend *modern high-resolution shock-capturing* (HRSC) methods to the relativistic case, first in one-dimensional calculations (Martí, Ibáñez and Miralles, 1991), and, later on, we extended them to multidimensional special relativistic (Font et al., 1994) and multidimensional general relativistic hydrodynamics (Banyuls et al., 1997). We made use of a linearized Riemann solver based on the *spectral decomposition* of the Jacobian matrices of the system.

Unlike the case of classical fluid dynamics the use of HRSC techniques in the frame of relativistic fluid dynamics is very recent and has yet to cover the full set of possible applications. Up to now, the most interesting astrophysical applications have involved the simulation of extragalactic relativistic jets (see (Aloy et al., 2000) and (Komissarov, 2000) for, respectively, relativistic 3D-hydro and 2D-magnetohydro calculations). Recently, some studies on the morphology of accreting flows onto moving black holes have been carried out (see (Font et al., 2000a) and references therein) using a multidimensional general relativistic hydrocode. A very promising application of HRSC techniques in the frame of general relativistic magnetohydrodynamics has been used recently to simulate the formation of relativistic jets from black holes magnetized accretion disks (Koide, Shibata and Kudoh, 1999).

At present, to develop robust and accurate (special or general) relativistic hydrocodes is a challenge in the field of Relativistic Astrophysics. A general relativistic hydrocode is a useful research tool for studying flows which evolve in a background spacetime. Furthermore, when appropriately coupled with Einstein equations, such a general relativistic hydrocode is crucial to model the evolution of matter in a dynamical spacetime. The coupling between geometry and matter arises through the sources of the corresponding system of equations. Such a marriage between numerical relativity and

numerical relativistic hydrodynamics could be useful, for example, to analyze the dynamics (and the physics) of coalescing compact binaries. These are one of the most promising sources of gravitational radiation to be detected by the near future Earth-based laser-interferometer observatories of gravitational waves.

2. The equations of general relativistic hydrodynamics as a hyperbolic system of conservation laws

The evolution of a relativistic fluid is governed by a system of equations which summarize *local conservation laws*: the local conservation of baryon number, $\nabla \cdot \mathbf{J} = 0$, and the local conservation of energy-momentum, $\nabla \cdot \mathbf{T} = 0$ ($\nabla \cdot$ stands for the covariant divergence).

If $\{\partial_t, \partial_i\}$ define the coordinate basis of 4-vectors which are tangents to the corresponding coordinate curves, then, the *current of rest-mass*, \mathbf{J} , and the *energy-momentum tensor*, \mathbf{T} , for a perfect fluid, have the components: $J^\mu = \rho u^\mu$, and $T^{\mu\nu} = \rho h u^\mu u^\nu + p g^{\mu\nu}$, respectively, ρ being the rest-mass density, p the pressure and h the specific enthalpy, defined by $h = 1 + \varepsilon + p/\rho$, where ε is the specific internal energy. u^μ is the four-velocity of the fluid and $g_{\mu\nu}$ defines the metric of the spacetime \mathcal{M} where the fluid evolves. As usually, Greek (Latin) indices run from 0 to 3 (1 to 3) – or, alternatively, they stand for the general coordinates $\{t, x, y, z\}$ ($\{x, y, z\}$) – and the system of units is the so-called geometrized ($c = G = 1$).

An equation of state $p = p(\rho, \varepsilon)$ closes, as usual, the system. Accordingly, the local sound velocity c_s satisfies: $hc_s^2 = \chi + (p/\rho^2)\kappa$, with $\chi = \partial p/\partial\rho|_\varepsilon$ and $\kappa = \partial p/\partial\varepsilon|_\rho$.

Following (Banyuls et al., 1997), let \mathcal{M} be a general spacetime described by the four dimensional metric tensor $g_{\mu\nu}$. According to the $\{3+1\}$ formalism, the metric is split into the objects α (*lapse*), β^i (*shift*) and γ_{ij} , keeping the line element in the form:

$$ds^2 = -(\alpha^2 - \beta_i \beta^i)dt^2 + 2\beta_i dx^i dt + \gamma_{ij} dx^i dx^j \quad (1)$$

If \mathbf{n} is a unit timelike vector field normal to the spacelike hypersurfaces Σ_t ($t = \text{const.}$), then, by definition of α and β^i is: $\partial_t = \alpha \mathbf{n} + \beta^i \partial_i$, with $\mathbf{n} \cdot \partial_i = 0$, $\forall i$. Observers, \mathcal{O}_E , at rest in the slice Σ_t , i.e., those having \mathbf{n} as four-velocity (*Eulerian observers*), measure the following velocity of the fluid

$$v^i = \frac{u^i}{\alpha u^t} + \frac{\beta^i}{\alpha} \quad (2)$$

where $W \equiv -(\mathbf{u} \cdot \mathbf{n}) = \alpha u^t$, the Lorentz factor, satisfies $W = (1 - v^2)^{-1/2}$ with $v^2 = v_i v^i$ ($v_i = \gamma_{ij} v^j$).

Let us define a basis adapted to the observer \mathcal{O}_E , $\mathbf{e}_{(\mu)} = \{\mathbf{n}, \partial_i\}$, and the following five four-vector fields $\{\mathbf{J}, \mathbf{T} \cdot \mathbf{n}, \mathbf{T} \cdot \partial_1, \mathbf{T} \cdot \partial_2, \mathbf{T} \cdot \partial_3\}$. Hence, the above system of equations of general relativistic hydrodynamics (GRH) can be written

$$\nabla \cdot \mathbf{A} = s, \quad (3)$$

where \mathbf{A} denotes any of the above 5 vector fields, and s is the corresponding source term.

The set of *conserved variables* gathers those quantities which are directly measured by \mathcal{O}_E , i.e., the rest-mass density (D), the momentum density in the j -direction (S_j) and the total energy density (E). In terms of the *primitive variables* $\mathbf{w} = (\rho, v_i, \varepsilon)$ ($v_i = \gamma_{ij} v^j$) they are

$$D = \rho W, \quad S_j = \rho h W^2 v_j, \quad E = \rho h W^2 - p \quad (4)$$

Taking all the above relations together, the fundamental system to be considered for numerical applications is

$$\frac{1}{\sqrt{-g}} \left(\frac{\partial \sqrt{\gamma} \mathbf{F}^0(\mathbf{w})}{\partial x^0} + \frac{\partial \sqrt{-g} \mathbf{F}^i(\mathbf{w})}{\partial x^i} \right) = \mathbf{s}(\mathbf{w}) \quad (5)$$

where the quantities $\mathbf{F}^\alpha(\mathbf{w})$ are

$$\mathbf{F}^0(\mathbf{w}) = (D, S_j, \tau) \quad (6)$$

$$\mathbf{F}^i(\mathbf{w}) = \left(D \left(v^i - \frac{\beta^i}{\alpha} \right), S_j \left(v^i - \frac{\beta^i}{\alpha} \right) + p \delta_j^i, \tau \left(v^i - \frac{\beta^i}{\alpha} \right) + p v^i \right) \quad (7)$$

and the corresponding sources $\mathbf{s}(\mathbf{w})$ are

$$\mathbf{s}(\mathbf{w}) = \left(0, T^{\mu\nu} \left(\frac{\partial g_{\nu j}}{\partial x^\mu} - \Gamma_{\nu\mu}^\delta g_{\delta j} \right), \alpha \left(T^{\mu 0} \frac{\partial \ln \alpha}{\partial x^\mu} - T^{\mu\nu} \Gamma_{\nu\mu}^0 \right) \right) \quad (8)$$

τ being $\tau \equiv E - D$, and $g \equiv \det(g_{\mu\nu})$ is such that $\sqrt{-g} = \alpha \sqrt{\gamma}$ ($\gamma \equiv \det(\gamma_{ij})$)

2.1. LINEARIZED RIEMANN SOLVERS AND CHARACTERISTIC FIELDS

Modern HRSC schemes use the characteristic structure of the hyperbolic system of conservation laws. In many Godunov-type schemes, the characteristic structure is used to compute either an exact or an approximate solution to a sequence of Riemann problems at each cell interface. In characteristic based methods the characteristic structure is used to compute the local characteristic fields, which define the directions along which the characteristic variables propagate. In both these approaches, the characteristic

decomposition of the Jacobian matrices of the nonlinear system of conservation laws is important, not only because it is one of the key ingredients in the design of the numerical flux at the interfaces, but because experience has shown that it facilitates a robust upgrading of the order of a numerical scheme.

The three 5×5 -Jacobian matrices \mathcal{B}^i associated to system (5) are

$$\mathcal{B}^i = \alpha \frac{\partial \mathbf{F}^i}{\partial \mathbf{F}^0}. \quad (9)$$

The *eigenvalues* of \mathcal{B}^x are

$$\lambda_0 = \alpha v^x - \beta^x \quad (\text{triple}) \quad (10)$$

which defines the *material waves*, and two other, λ_{\pm} , associated with the *acoustic waves*

$$\begin{aligned} \lambda_{\pm} &= \frac{\alpha}{1 - v^2 c_s^2} \left\{ v^x (1 - c_s^2) \pm \right. \\ &\quad \left. \pm c_s \sqrt{(1 - v^2)[\gamma^{xx}(1 - v^2 c_s^2) - v^x v^x (1 - c_s^2)]} \right\} - \beta^x \end{aligned} \quad (11)$$

A complete set of *right-eigenvectors* is

$$\mathbf{r}_{\pm} = \begin{bmatrix} 1 \\ hW \left(v_x - \frac{v^x - \Lambda_{\pm}^x}{\gamma^{xx} - v^x \Lambda_{\pm}^x} \right) \\ hW v_y \\ hW v_z \\ \frac{hW(\gamma^{xx} - v^x v^x)}{\gamma^{xx} - v^x \Lambda_{\pm}^x} - 1 \end{bmatrix}, \quad \mathbf{r}_{0,1} = \begin{bmatrix} \frac{\mathcal{K}}{hW} \\ v_x \\ v_y \\ v_z \\ 1 - \frac{\mathcal{K}}{hW} \end{bmatrix}$$

$$\mathbf{r}_{0,2} = \begin{bmatrix} Wv_y \\ h(\gamma_{xy} + 2W^2v_xv_y) \\ h(\gamma_{yy} + 2W^2v_yv_y) \\ h(\gamma_{zy} + 2W^2v_zv_y) \\ Wv_y(2hW - 1) \end{bmatrix}, \quad \mathbf{r}_{0,3} = \begin{bmatrix} Wv_z \\ h(\gamma_{xz} + 2W^2v_xv_z) \\ h(\gamma_{yz} + 2W^2v_yv_z) \\ h(\gamma_{zz} + 2W^2v_zv_z) \\ Wv_z(2hW - 1) \end{bmatrix}$$

The corresponding set of *left-eigenvectors* is

$$\mathbf{l}_{0,1} = \frac{W}{\mathcal{K} - 1} (h - W, Wv^x, Wv^y, Wv^z, -W)$$

$$\mathbf{l}_{0,2} = \frac{1}{h\xi} \begin{bmatrix} -\gamma_{zz}v_y + \gamma_{yz}v_z \\ v^x(\gamma_{zz}v_y - \gamma_{yz}v_z) \\ \gamma_{zz}\mathcal{W}^x + \gamma_{xz}v_zv^x \\ -\gamma_{yz}\mathcal{W}^x - \gamma_{xz}v_yv^x \\ -\gamma_{zz}v_y + \gamma_{yz}v_z \end{bmatrix}, \quad \mathbf{l}_{0,3} = \frac{1}{h\xi} \begin{bmatrix} -\gamma_{yy}v_z + \gamma_{zy}v_y \\ v^x(\gamma_{yy}v_z - \gamma_{zy}v_y) \\ -\gamma_{zy}\mathcal{W}^x - \gamma_{xy}v_zv^x \\ \gamma_{yy}\mathcal{W}^x + \gamma_{xy}v_yv^x \\ -\gamma_{yy}v_z + \gamma_{zy}v_y \end{bmatrix}$$

$$\mathbf{l}_\mp = (\pm 1) \frac{h^2}{\Delta} \begin{bmatrix} hW\mathcal{V}_\pm^x\xi + l_\mp^{(5)} \\ \Gamma_{xx}(1 - \mathcal{K}\tilde{\mathcal{A}}_\pm^x) + (2\mathcal{K} - 1)\mathcal{V}_\pm^x(W^2v^x\xi - \Gamma_{xx}v^x) \\ \Gamma_{xy}(1 - \mathcal{K}\tilde{\mathcal{A}}_\pm^x) + (2\mathcal{K} - 1)\mathcal{V}_\pm^x(W^2v^y\xi - \Gamma_{xy}v^x) \\ \Gamma_{xz}(1 - \mathcal{K}\tilde{\mathcal{A}}_\pm^x) + (2\mathcal{K} - 1)\mathcal{V}_\pm^x(W^2v^z\xi - \Gamma_{xz}v^x) \\ (1 - \mathcal{K})[-\gamma v^x + \mathcal{V}_\pm^x(W^2\xi - \Gamma_{xx})] - \mathcal{K}W^2\mathcal{V}_\pm^x\xi \end{bmatrix}$$

where the following auxiliary quantities and relations have been introduced:

$$\Lambda_\pm^i \equiv \tilde{\lambda}_\pm + \tilde{\beta}^i, \quad \tilde{\lambda} \equiv \lambda/\alpha, \quad \tilde{\beta}^i \equiv \beta^i/\alpha \quad (12)$$

$$\mathcal{K} \equiv \frac{\tilde{\kappa}}{\tilde{\kappa} - c_s^2} \quad , \quad \tilde{\kappa} \equiv \kappa/\rho \quad (13)$$

$$\mathcal{C}_{\pm}^x \equiv v_x - \mathcal{V}_{\pm}^x \quad , \quad \mathcal{V}_{\pm}^x \equiv \frac{v^x - \Lambda_{\pm}^x}{\gamma^{xx} - v^x \Lambda_{\pm}^x} \quad (14)$$

$$\mathcal{W}^x \equiv 1 - v_x v^x \quad , \quad \tilde{\mathcal{A}}_{\pm}^x \equiv \frac{\gamma^{xx} - v^x v^x}{\gamma^{xx} - v^x \Lambda_{\pm}^x} \quad (15)$$

$$\Delta \equiv h^3 W(\mathcal{K} - 1)(\mathcal{C}_+^x - \mathcal{C}_-^x)\xi \quad (16)$$

$$\xi \equiv \Gamma_{xx} - \gamma v^x v^x \quad , \quad \Gamma_{xx} = \gamma_{yy} \gamma_{zz} - \gamma_{yz}^2 \quad , \quad \dots \quad (17)$$

Symmetry arguments allow one to obtain the spectral decomposition in the other spatial directions. The corresponding expressions in Special Relativity (Donat et al., 1998) are easily covered. The above full spectral decomposition provides the user with the technical ingredients needed to develop state-of-the-art, upwind-based HRSC codes for numerical relativistic hydrodynamics. In 3D general-relativistic applications, and depending on the particular Riemann solver or flux formula used, the knowledge of the analytical values of the left-eigenvectors could be crucial in the efficiency of the code (see (Aloy et al., 1999)).

3. Special Relativistic Riemann Solvers in General Relativistic Hydrodynamics

Up to now, only a small number of papers have considered the extension of HRSC methods to GRH using linearized Riemann Solvers or flux formulae –(Martí, Ibáñez and Miralles, 1991) and (Romero et al., 1996) for 1D problems, (Banyuls et al., 1997), (Font et al., 2000b) and (Papadopoulos and Font, 1999)–. or deriving explicitly an extension of Roe’s Riemann solver to GRH (Eulderink and Mellema, 1995).

In (Pons et al., 1998) we have answered the following basic question (suggested in (Balsara, 1994) and (Martí, 1997)): Is it possible to obtain a general procedure that allows one to take advantage of Special Relativistic Riemann Solvers (SRRS) to generate numerical solutions describing the evolution of relativistic flows in strong gravitational fields?

The affirmative answer relies on a *local change of coordinates*, at each numerical interface, in terms of which the spacetime metric is locally Minkowskian. Our procedure, hence, follows analogous trends to those used in

classical fluid dynamics to solve Riemann problems in general curvilinear coordinates.

Let us consider the integral form of the system of equations (3) on a four-dimensional volume Ω , with three-dimensional boundary $\partial\Omega$, and apply Gauss theorem to obtain the corresponding balance equation

$$\int_{\partial\Omega} \mathbf{A} \cdot d^3\Sigma = \int_{\Omega} s d\Omega. \quad (18)$$

For numerical applications, we choose volume Ω as the one bounded by the coordinate hypersurfaces $\{\Sigma_{x^\alpha}, \Sigma_{x^\alpha+\Delta x^\alpha}\}$. Hence, the time variation of the mean value of A^0 ,

$$\bar{A}^0 = \frac{1}{\Delta\mathcal{V}} \int_{x^1}^{x^1+\Delta x^1} \int_{x^2}^{x^2+\Delta x^2} \int_{x^3}^{x^3+\Delta x^3} \sqrt{-g} A^0 dx^1 dx^2 dx^3, \quad (19)$$

within the spatial volume

$$\Delta\mathcal{V} = \int_{x^1}^{x^1+\Delta x^1} \int_{x^2}^{x^2+\Delta x^2} \int_{x^3}^{x^3+\Delta x^3} \sqrt{-g} dx^1 dx^2 dx^3, \quad (20)$$

can be obtained from

$$\begin{aligned} (\bar{A}^0 \Delta\mathcal{V})_{t+\Delta t} = & (\bar{A}^0 \Delta\mathcal{V})_t + \int_{\Omega} s d\Omega - \\ & \left(\int_{\Sigma_{x^1}} \mathbf{A} \cdot d^3\Sigma + \int_{\Sigma_{x^1+\Delta x^1}} \mathbf{A} \cdot d^3\Sigma + \right. \\ & \int_{\Sigma_{x^2}} \mathbf{A} \cdot d^3\Sigma + \int_{\Sigma_{x^2+\Delta x^2}} \mathbf{A} \cdot d^3\Sigma + \\ & \left. \int_{\Sigma_{x^3}} \mathbf{A} \cdot d^3\Sigma + \int_{\Sigma_{x^3+\Delta x^3}} \mathbf{A} \cdot d^3\Sigma \right). \end{aligned} \quad (21)$$

In order to advance in time, the volume and surface integrals on the right hand side have to be evaluated. We have applied HRSC to calculate the \mathbf{A} vector fields by solving local Riemann problems combined with monotonized cell reconstruction techniques.

According to the Equivalence Principle, physical laws in a *local inertial frame* of a curved spacetime have the same form as in special relativity (see, e.g., (Schutz, 1985)). Locally flat (or geodesic) systems of coordinates, in which the metric is brought into the Minkowskian form up to second order terms, are the realization of these local inertial frames. However, whereas the coordinate transformation leading to locally flat coordinates involves second order terms, locally Minkowskian coordinates are obtained by a linear transformation. This fact is of crucial importance when exploiting the

selfsimilar character of the solution of the Riemann problems set up across the coordinate surfaces.

Hence, we propose to perform a coordinate transformation to locally Minkowskian coordinates at each numerical interface assuming that the solution of the Riemann problem is one in special relativity and planar symmetry. This last assumption is equivalent to the approach followed in classical fluid dynamics, when using the solution of Riemann problems in slab symmetry for problems in cylindrical or spherical coordinates, which breaks down near the singular points (*e.g.* the polar axis in cylindrical coordinates). Analogously to classical fluid dynamics, the numerical error will depend on the magnitude of the Christoffel symbols, which might be large whenever huge gradients or large temporal variations of the gravitational field are present. Finer grids and improved time advancing methods will be required in those regions.

In the rest of this section we will focus on the evaluation of the first flux integral in Eq. (21).

$$\int_{\Sigma_{x^1}} \mathbf{A} \cdot d^3\Sigma = \int_{\Sigma_{x^1}} A^1 \sqrt{-g} dx^0 dx^2 dx^3 \quad (22)$$

To begin, we will express the integral on a basis $\mathbf{e}_{\hat{\alpha}}$ with $\mathbf{e}_{\hat{0}} \equiv \mathbf{n}$ and $\mathbf{e}_{\hat{i}}$ forming an orthonormal basis in the plane orthogonal to \mathbf{n} with $\mathbf{e}_{\hat{j}}$ normal to the surface Σ_{x^1} and $\mathbf{e}_{\hat{2}}$ and $\mathbf{e}_{\hat{3}}$ tangent to that surface. The vectors of this basis verify $\mathbf{e}_{\hat{\alpha}} \cdot \mathbf{e}_{\hat{\beta}} = \eta_{\hat{\alpha}\hat{\beta}}$ with $\eta_{\hat{\alpha}\hat{\beta}}$ the Minkowski metric (in the following, caret subscripts will refer to vector components in this basis).

Denoting by x_0^α the coordinates at the center of the interface at time t , we introduce the following locally Minkowskian coordinate system

$$x^{\hat{\alpha}} = M_{\alpha}^{\hat{\alpha}}(x^\alpha - x_0^\alpha), \quad (23)$$

where the matrix $M_{\alpha}^{\hat{\alpha}}$ is given by $\partial_\alpha = M_{\alpha}^{\hat{\alpha}} \mathbf{e}_{\hat{\alpha}}$, calculated at x_0^α . In this system of coordinates the equations of general relativistic hydrodynamics transform into the equations of special relativistic hydrodynamics, in Cartesian coordinates, but with non-zero sources, and the flux integral (22) reads

$$\int_{\Sigma_{x^1}} (A^{\hat{1}} - \frac{\beta^{\hat{1}}}{\alpha} A^{\hat{0}}) \sqrt{-\hat{g}} dx^{\hat{0}} dx^{\hat{2}} dx^{\hat{3}} \quad (24)$$

with $\sqrt{-\hat{g}} = 1 + \mathcal{O}(x^{\hat{\alpha}})$, where we have taken into account that, in the coordinates $x^{\hat{\alpha}}$, Σ_{x^1} is described by the equation $x^{\hat{1}} - \frac{\beta^{\hat{1}}}{\alpha} x^{\hat{0}} = 0$ (with $\beta^{\hat{i}} = M_i^{\hat{i}} \beta^i$), where the metric elements β^1 and α are calculated at x_0^α . Therefore, this surface can be considered as a moving surface with speed β^1/α . We draw reader's attention to the fact that since the motion of the

boundaries $\partial\Omega$ of the control volumes does not correspond, in general, to the motion of physical particles it is not constrained by causality. Consequently, the values of β^i/α (which are gauge-dependent) can be greater than one.

At this point, all the theoretical work on SRRS developed in recent years, can be exploited. The quantity in parenthesis in (24) represents the numerical flux across Σ_{x^1} , which can, now, be calculated by solving the special relativistic Riemann problem defined with the values at the two sides of Σ_{x^1} of two independent thermodynamical variables (namely, the rest mass density ρ and the specific internal energy ϵ) and the components of the velocity in the orthonormal spatial basis v^i ($v^i = M_i^{\hat{j}} v^{\hat{j}}$). Although most linearized Riemann solvers provide the numerical fluxes for surfaces at rest, it is easy to apply them to moving surfaces, relying on the conservative and hyperbolic character of the system of equations ((Harten and Hyman, 1983)).

Once the Riemann problem has been solved, by means of any linearized or exact SRRS, we can take advantage of the selfsimilar character of the solution of the Riemann problem, which makes it constant on the surface Σ_{x^1} simplifying the calculation of the above integral enormously (24):

$$\int_{\Sigma_{x^1}} \mathbf{A} \cdot d^3\Sigma = (A^{\hat{1}} - \frac{\beta^{\hat{1}}}{\alpha} A^{\hat{0}})^* \int_{\Sigma_{x^1}} \sqrt{-\hat{g}} dx^{\hat{0}} dx^{\hat{2}} dx^{\hat{3}} \quad (25)$$

where the superscript (*) stands for the value on Σ_{x^1} obtained from the solution of the Riemann problem. The integral in the right hand side of (24) is the area of the surface Σ_{x^1} and can be expressed in terms of the original coordinates as

$$\int_{\Sigma_{x^1}} \sqrt{\gamma^{11}} \sqrt{-g} dx^0 dx^2 dx^3 \quad (26)$$

which can be evaluated for a given metric.

Finally, notice that the numeriacl fluxes defined in (24) correspond to the vector fields $\mathbf{A} = \{\mathbf{J}, \mathbf{T} \cdot \mathbf{n}, \mathbf{T} \cdot e_1, \mathbf{T} \cdot e_2, \mathbf{T} \cdot e_3\}$. Thus the additional relation

$$\mathbf{T} \cdot \partial_i = M_i^{\hat{j}} (\mathbf{T} \cdot \mathbf{e}_{\hat{j}}) \quad (27)$$

has to be used for the momentum equations.

The interested reader can address reference (Pons *et al.*, 1998) for details on the testing and calibration of our procedure. The additional computational cost of the approach is completely negligible. The procedure has a large potentiality and can be applied to other systems of conservation laws, such as magneto-hydrodynamics (MHD), making possible to solve the general relativistic MHD equations using the corresponding Riemann solvers developed for the special relativistic case.

4. Conclusions

An appropriate conservative formulation for the equations, together with the knowledge of the characteristic fields associated to the system, define the starting point for using HRSC schemes. The spectral decomposition of the Jacobian matrices, corresponding to the fluxes in each spatial direction, is used in the numerical flux computation and, moreover, it is potentially interesting in allowing an extensive range of application of HRSC methods with different approximate Riemann solvers or flux formulae.

The procedure outlined in Section §3 is –from the computational point of view – very cheap, since it involves a linear change of coordinates. It has a large potentiality and can be applied to other systems of conservation laws, as magneto-hydrodynamics, giving a very useful numerical tool to solve the general relativistic MHD equations using the corresponding Riemann solvers developed for the special relativistic case. In particular, it is possible to use the exact solution of the special relativistic Riemann problem (Martí and Müller, 1994), (Pons, Martí and Müller, 2000).

The astrophysical applications foreseen in the present and near future include the study of jet formation, multidimensional stellar core collapse, gamma-ray bursts and the coalescence of compact binaries. HRSC methods can, without doubt, be used successfully to tackle these scenarios, and acquire the prestige they have already earned in the simulation of relativistic jets and accretion flows around compact objects.

Acknowledgments: This work has been supported by the Spanish DGES (grant PB97-1432). J.A.F. acknowledges financial support from a TMR fellowship of the European Union (contract nr. ERBFMBICT971902).

References

- Aloy, M.A., Ibáñez, J.M.^a, Martí, J.M.^a, Gómez and Müller, E. (2000). Simulations of relativistic jets with GENESIS. This volume.
- Aloy, M.A., Pons, J.A., and Ibáñez, J.M.^a. (1999). An Efficient Implementation of Flux Formulae in Multidimensional Relativistic Hydrodynamical Codes. *Comp. Phys. Comm.*, **120**, pp 115 - 121.
- Balsara, D.S. (1994). Riemann Solver for Relativistic Hydrodynamics. *J. Comput. Phys.*, **114**, pp 284 - 297.
- Banyuls F., Font J.A., Ibáñez J.M.^a, Martí J.M.^a, and Miralles J.A. (1997). Numerical {3+1} General-Relativistic Hydrodynamics: A Local Characteristic Approach. *ApJ*, **476**, pp 221 - 233.
- Donat, R., Font, J.A., Ibáñez, J.M.^a, and Marquina, A. (1998). A Flux-Split Algorithm Applied to Relativistic Flows. *J. Comput. Phys.*, **146**, pp 58 - 81.
- Eulderink F., and Mellemma G. (1995). General Relativistic Hydrodynamics with a Roe Solver. *A&A Suppl.*, **110**, pp 587 - 623.
- Font J.A., Ibáñez J.M.^a, Marquina A., and Martí J.M.^a. (1994). Multidimensional Relativistic Hydrodynamics: Characteristic Fields and High-Resolution Shock-Capturing Schemes. *A&A*, **282**, pp 304 - 314.

- Font J.A., Ibáñez J.M^A., and Papadopoulos P. (2000). Numerical Simulations of Relativistic Accretion onto Black Holes using Godunov-type Methods. This volume.
- Font J.A., Miller M., Suen W.-M., and Tobias M. (2000). Three Dimensional Numerical General Relativistic Hydrodynamics I: Formulations, Methods, and Code Tests. *Phys. Rev. D*, **61**, pp 044011.1 - 044011.26.
- Harten A., and Hyman J.M. (1983). Self-adjusting Grid Methods for One-dimensional Hyperbolic Conservation Laws. *J. Comput. Phys.*, **50**, pp 235 - 269.
- Koide S., Shibata K., and Kudoh T. (1999). Relativistic Jet Formation from Black Hole Magnetized Accretion Disks: Method, Tests, and Applications of a General Relativistic Magnetohydrodynamic Numerical Code. *ApJ*, **522**, pp 727 - 752.
- Komissarov S. (2000). Relativistic MHD Simulations using Godunov-Type Methods. This volume.
- Martí J.M^A. (1997). High-Order Finite-Difference Schemes. In: Relativistic Gravitation and Gravitational Radiation, pp 239 - 255. Marck J-A., and Lasota J-P., (Editors). Cambridge University Press.
- Martí J.M^A., Ibáñez J.M^A., and Miralles J.A. (1991). Numerical Relativistic Hydrodynamics: Local Characteristic Approach *Phys. Rev. D*, **43**, pp 3794 - 3801.
- Martí J.M^A., and Müller E. (1994). The Analytical Solution of the Riemann Problem in Relativistic Hydrodynamics. *J. Fluid Mech.*, **258**, pp 317 - 333.
- May M.A., and White R.H. (1967). Stellar Dynamics and Gravitational Collapse. *Math. Comp. Phys.*, **7**, pp 219 - 258.
- Norman M.L., and Winkler K-H.A. (1986). Why Ultrarelativistic Hydrodynamics is Difficult. In: Astrophysical Radiation Hydrodynamics, pp 449 - 476. Norman M.L., and Winkler K-H.A. (Editors). Reidel Publ.
- Papadopoulos P., and Font J.A. (2000). Relativistic Hydrodynamics on Spacelike and Null Surfaces: Formalism and Computations of Spherically Symmetric Spacetimes. *Phys. Rev. D*, **61**, pp 024015.1 - 024015.15.
- Pons J.A., Martí J.M^A., and Müller E. (2000). An Exact Riemann Solver for Multidimensional Special Relativistic Hydrodynamics. This volume.
- Pons J.A., Font J.A., Ibáñez J.M^A., Martí J.M^A., and Miralles J.A. (1998). General Relativistic Hydrodynamics with Special Relativistic Riemann Solvers. *A&A*, **339**, pp 638 - 642.
- Romero J.V., Ibáñez J.M^A., Martí J.M^A., and Miralles J.A. (1996). A New Spherically Symmetric General Relativistic Hydrodynamical Code. *ApJ*, **462**, pp 839 - 854.
- Schutz B.F. (1985). A First Course in General Relativity, pp 184. Cambridge University Press.
- Wilson J.R. (1972). Numerical Study of Fluid Flow in a Kerr Space. *ApJ*, **173**, pp 431 - 438.

A FULLY ADAPTIVE MULTIRESOLUTION SCHEME FOR SHOCK COMPUTATIONS

M. K. KAIBARA

Universidade Estadual Paulista - Depto. de Matemática

Av. Engenheiro Edmundo C. Coube s/n, 17033-360

Bauru - SP, Brasil

Email: kaibara@fc.unesp.br

AND

S. M. GOMES

Universidade Estadual de Campinas - IMECC

Caixa Postal 6065, 13081-970

Campinas - SP, Brasil

Email: soniag@ime.unicamp.br

Abstract. The scheme is based on Ami Harten's ideas (Harten , 1994), the main tools coming from wavelet theory, in the framework of multiresolution analysis for cell averages. But instead of evolving cell averages on the finest uniform level, we propose to evolve just the cell averages on the grid determined by the significant wavelet coefficients. Typically, there are few cells in each time step, big cells on smooth regions, and smaller ones close to irregularities of the solution. For the numerical flux, we use a simple uniform central finite difference scheme, adapted to the size of each cell. If any of the required neighboring cell averages is not present, it is interpolated from coarser scales. But we switch to ENO scheme in the finest part of the grids. To show the feasibility and efficiency of the method, it is applied to a system arising in polymer-flooding of an oil reservoir. In terms of CPU time and memory requirements, it outperforms Harten's multiresoltution algorithm.

The proposed method applies to systems of conservation laws in 1D

$$\partial_t u(x, t) + \partial_x f(u(x, t)) = 0, \quad u(x, t) \in \mathbf{R}^m. \quad (1)$$

In the spirit of finite volume methods, we shall consider the explicit scheme

$$v_{\mu}^{n+1} = v_{\mu}^n - \frac{\Delta t}{h_{\mu}} (\bar{f}_{\mu} - \bar{f}_{\mu^-}) = [\mathcal{D}v^n]_{\mu}, \quad (2)$$

where μ is a point of an irregular grid Γ , μ^- is the left neighbor of μ in Γ , $v_{\mu}^n \approx \frac{1}{\mu-\mu^-} \int_{\mu^-}^{\mu} u(x, t_n) dx$ are approximated cell averages of the solution, $\bar{f}_{\mu} = \bar{f}_{\mu}(v^n)$ are the numerical fluxes, and \mathcal{D} is the numerical evolution operator of the scheme.

According to the definition of \bar{f}_{μ} , several schemes of this type have been proposed and successfully applied (LeVeque, 1990). Godunov, Lax-Wendroff, and ENO are some of the popular names. Godunov scheme resolves well the shocks, but accuracy (of first order) is poor in smooth regions. Lax-Wendroff is of second order, but produces dangerous oscillations close to shocks. ENO schemes are good alternatives, with high order and without serious oscillations. But the price is high computational cost.

Ami Harten proposed in (Harten , 1994) a simple strategy to save expensive ENO flux calculations. The basic tools come from multiresolution analysis for cell averages on uniform grids, and the principle is that wavelet coefficients can be used for the characterization of local smoothness. Typically, only few wavelet coefficients are significant. At the finest level, they indicate discontinuity points, where ENO numerical fluxes are computed exactly. Elsewhere, cheaper fluxes can be safely used, or just interpolated from coarser scales. Different applications of this principle have been explored by several authors, see for example (G-Müller and Müller, 1998).

Our scheme also uses Ami Harten's ideas. But instead of evolving the cell averages on the finest uniform level, we propose to evolve the cell averages on sparse grids associated with the significant wavelet coefficients. This means that the total number of cells is small, with big cells in smooth regions and smaller ones close to irregularities. This task requires improved new tools, which are described next.

A. About the Grids: We shall work with embedded grids $\Gamma^l \subset \Gamma^{l+1}$. In the uniform setting, $\Gamma^l = X^l$ are dyadic grids of the unit interval $[0, 1]$. In the non uniform setting, $\Gamma^0 = X^0$, and for $l > 0$, $\Gamma^l \subset X^l$ is constructed by adding to Γ^{l-1} some points from $X^l \setminus X^{l-1}$. $\Lambda^{l-1} = \Gamma^l \setminus \Gamma^{l-1}$ is the set of these new points.

B. Multiresolution Analysis (MRA): In a multilevel setting, the relations between the discrete informations g^{l+1} and g^l at two consecutive levels, and the difference of information d^l between them, are crucial. They lead to multilevel transformations $g^L \xrightarrow{T^L} (g^0, d^0, \dots, d^{L-1})$. In wavelet analysis,

d_μ^l are the wavelet coefficients, T^L and its inverse $(T^L)^{-1}$ are the *analysis* and *synthesis* algorithms.

MRA for Point Values: The main ingredient in a MRA for point values is an interpolatory subdivision scheme. Of particular interest are those defined by polynomial Lagrange interpolation (Dubuc, 1986), which can easily be adapted to irregular grids and bounded intervals. Starting from the point values $g_\mu^l = g(\mu)$, $\mu \in \Gamma^l$, they iteratively define interpolation operators $\mathcal{I}^l(x; g^l)$ such that $\mathcal{I}^l(\mu; g^l) = g_\mu^l$, $\mu \in \Gamma^l$. The difference of information between two consecutive levels is produced by the interpolation error

$$d_\mu^l = g_\mu^{l+1} - \mathcal{I}^l(\mu; g^l), \quad \mu \in \Lambda^l. \quad (3)$$

MRA for Cell Averages: To each point $\mu \in \Gamma^l \setminus \{0\}$ there is an associated cell $[\mu^-, \mu]$, of size $h_{l,\mu} = \mu - \mu^-$, where μ^- is the left neighbor of μ in Γ^l . Let $G(x) = \int_0^x g(s)ds$ be the primitive function. For $\mu \in \Lambda^l$, the cell averages \bar{g}^l satisfy $\bar{g}_\mu^l = [G(\mu) - G(\mu^-)]/h_{l,\mu}$, and $G(\mu) = \sum_{\nu \leq \mu} h_{l,\nu} \bar{g}_\nu^l$. Therefore, $\mathcal{R}^l(x; \bar{g}^l) = \frac{d}{dx} \mathcal{I}^l(x; G^l)$ is a cell-averages interpolator. For $\mu \in \Lambda^l$, it can be used in the approximation

$$\bar{g}_\mu^{l+1} \approx \tilde{g}_\mu^{l+1} = \frac{1}{h_{l+1,\mu}} \int_{\mu^-}^\mu \mathcal{R}^l(x; \bar{g}^l) dx = \frac{\mathcal{I}^l(\mu; G^l) - \mathcal{I}^l(\mu^-; G^l)}{h_{l+1,\mu}}, \quad (4)$$

Wavelet coefficients are the errors in this approximation

$$\bar{d}_\mu^l = \bar{g}_\mu^{l+1} - \tilde{g}_\mu^{l+1}. \quad (5)$$

C. Grid Reduction: So far, we have considered MRAs for given irregular grids. However, we want to have grids adapted to a particular function at hand. For instance, if g is well represented by point values in Γ , the goal may be a more economic representation in a subgrid $\Gamma_\epsilon \subset \Gamma$, with comparable accuracy. In wavelet analysis this occurs naturally after the thresholding operation \mathcal{T}_ϵ acting after the analysis algorithm by simply removing from Γ those points corresponding to wavelet coefficients of magnitude less than ϵ . As interpolation errors, wavelet coefficients are good indicators of local smoothness. This means that the discarded coefficients correspond to smooth regions. Thus Γ_ϵ will be more sparse there, and fine just close to irregularities of g . The same procedure works for cell averages.

D. Grid Extension: From cell averages \bar{g} on a certain grid Γ , it may be of interest to have the cell averages \tilde{g} in a refined grid $\tilde{\Gamma}$. We may not have access exactly to all of them. If necessary, they are approximated using the reconstruction $\mathcal{R}(x; \bar{g})$. For our applications, the grid extension is performed before each time step, both in space and frequency domains. This

means that, for each present wavelet coefficient, we add some neighboring coefficients at the same scale and at the next finer scale. This precaution is needed to capture possible translations and appearance of high frequencies in the solution during the next time interval.

E. Adaptive Numerical Flux: The design of the numerical flux is crucial. In our tests we adopt Lax-Wendroff as basic algorithm in smooth regions, i.e., where the grid is sparse. In such case, the finite difference is performed with step of the size of the cell. Eventually, neighboring cell averages of the corresponding size may not be present. In such case, they are approximated from the coarser scales by using the reconstruction operator. This method was suggested by Mats Holmström in (Holmström, 1996) in the context of MRA for point values. Close to discontinuities, i.e., at points associated with significant wavelet coefficients at the finest scale, we switch to a second order ENO scheme.

During time evolution, there will be several grids, with corresponding cell averages, and thus several MRAs. It will be useful to keep an uniform notation for the thresholding operator \mathcal{T}_ϵ , grid extension \mathcal{E} , and time evolution operation \mathcal{D} , independently of the the grid at hand.

Suppose that the numerical solution at time t_n is represented by the cell averages v^n in an adapted grid Γ^n . The following steps will lead to the representation v^{n+1} at the next time level.

1. **Extension:** $\tilde{v}^n \leftarrow \mathcal{E}v^n$
2. **Evolution:** $\tilde{v}^{n+1} \leftarrow \mathcal{D}\tilde{v}^n$
3. **Reduction:** $v^{n+1} \leftarrow \mathcal{T}_\epsilon\tilde{v}^{n+1}$

To show the feasibility and efficiency of the method, it is applied to a system arising in polymer-flooding of an oil reservoir. The process consists of injection of water, plus a small amount of polymer, into certain wells to produce oil from others. The variables are the water saturation s , and the polymer saturation $b = cs$, where c is the volume concentration of polymer in the water, such that $0 \leq b \leq s \leq 1$. The system of conservation law becomes

$$s_t + (sg)_x = 0 \quad (6)$$

$$b_t + (bg)_x = 0 \quad (7)$$

where $g = f(s, c)/s$, $f(s, c)$ being the water particle velocity. More details can be found in (Johansen and Winther, 1988). We solve the problem with $f(s, c) = s^2[s^2 + (1 - s)^2(0.5 + c)]^{-1}$, and initial and boundary conditions

$$\begin{cases} s(x, 0) &= 0.6 & s(0, t) &= 1.0 \\ b(x, 0) &= 0.0 & b(0, t) &= 0.2. \end{cases} \quad (8)$$

ϵ	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
s	$1.1 \cdot 10^{-2}$	$6.2 \cdot 10^{-3}$	$1.5 \cdot 10^{-3}$	$5.76 \cdot 10^{-5}$	$4.97 \cdot 10^{-6}$
b	$6.3 \cdot 10^{-3}$	$2.1 \cdot 10^{-3}$	$4.92 \cdot 10^{-4}$	$1.53 \cdot 10^{-5}$	$1.27 \cdot 10^{-6}$

TABLE 1. $\max_{0 \leq t \leq 1} \|ER(t)\|_1$.

In our tests, cubic interpolation is used in the subdivision scheme, $CFL = 0.8$, and the coarsest grid always has 8 points. Figure 1 shows the numerical solution at $t = 0.109375$, and in Figure 2 the position-scale of the significant wavelet coefficients are indicated ($\epsilon = 10^{-5}$, and the number of levels is $L = 8$). Let $ER(t_n) = \bar{w}^n - \tilde{v}^n$ be the difference between

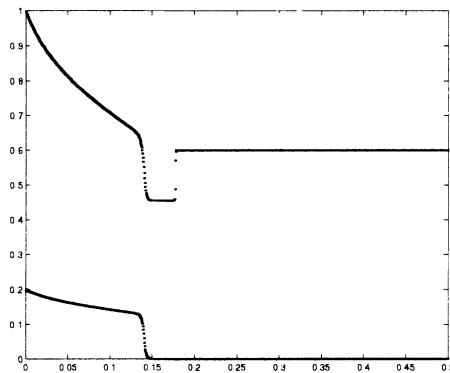
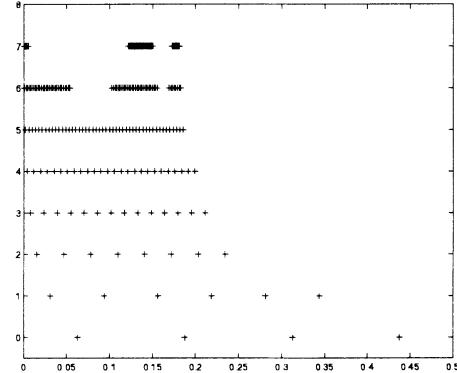
Figure 1. $t = 0.109375$ 

Figure 2. wavelet coefficients

\bar{w}^n , the solution produced by the pure ENO scheme on the finest uniform mesh X^L , and \tilde{v}^n , the solution of our adaptive scheme extended to X^L . Let $N_L = \#X^L - 1$ and define

$$\|ER(t)\|_1 = \frac{1}{N_L} \sum_{\mu \in X^L \setminus \{0\}} |\tilde{v}_\mu^n - \bar{w}_\mu^n|. \quad (9)$$

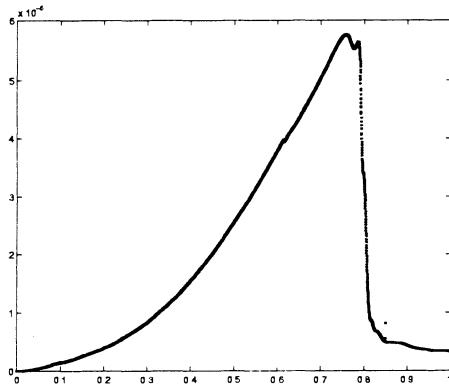
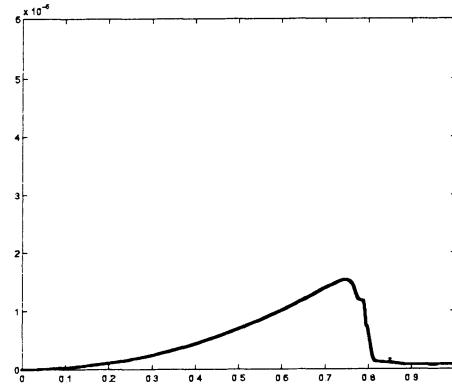
As shown in Table 1, $\max_{0 \leq t \leq 1} \|ER(t)\|_1$ stays at the same order of the truncation parameter ϵ , the same behavior of Harten's multiresolution scheme (Harten, 1995).

Table 2 lists the CPU times in seconds to evolve the solution up to $t = 0.109375$, by running the fully adaptive algorithm (FA), with different values of ϵ , and number of levels L . Also included are the same data for pure ENO and Harten's multiresolution algorithms based on X^L . The fully

ϵ	Scheme	$L = 5$	$L = 6$	$L = 7$	$L = 8$	$L = 9$	$L = 10$	$L = 11$
	ENO	1.14	4.53	17.98	71.52	285.92	1143.07	4608.98
10^{-3}	Harten	0.53	1.85	6.85	25.99	100.26	393.15	1553.52
	FA	0.60	1.78	5.26	15.24	46.34	148.75	485.45
10^{-4}	Harten	0.55	1.94	7.14	26.83	102.44	397.64	1564.93
	FA	0.64	2.00	6.16	17.96	51.73	154.45	496.64
10^{-5}	Harten	0.56	2.02	7.47	27.87	105.44	405.24	1581.95
	FA	0.66	2.15	7.08	21.70	63.40	182.04	547.01

TABLE 2. CPU time

adaptive method starts to be faster at $L = 6$. This efficiency improves as L and ϵ increase. The plots in Figures 3 and 4 correspond to the error $\|ER(t)\|_1$ for water and polymer saturations. Figure 5 shows $CPU(t)$, the time to evolve the solution up t , for both pure ENO (dashed line) and fully adaptive (continuous line) schemes. Figure 6 indicates the number of cell averages evolved at each time step.

Figure 3. $\|ER(t)\|_1$ for s Figure 4. $\|ER(t)\|_1$ for b

Conclusions

In order to accelerate the computations, we propose in this paper a fully adaptive multiresolution method for the numerical solution of conservation laws. The method is successfully tested in a model problem simulating polymer-flooding of an oil reservoir. Adaptivity is obtained by keeping control of the significant wavelet coefficients in a multiresolution analysis

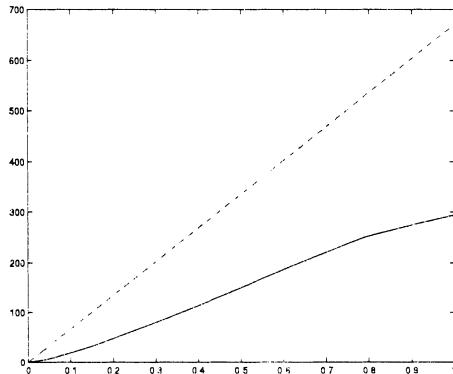


Figure 5. CPU times

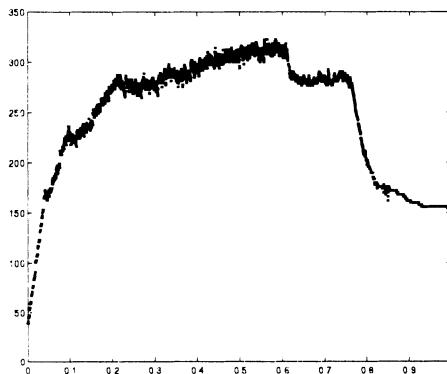


Figure 6. # cells per iteration

of the cell averages of the numerical solution. The cells are automatically adapted to the solution: big cells on smooth regions and small ones close to irregularities. The number of cells maintains small during time evolution, leading to substantial savings in CPU times. This efficiency increases if the level of resolution is increased. Savings in memory may also be of interest, specially in higher dimensions. As described in (Holmström, 1996), this can be achieved by imposing block structured sparse grids.

Acknowledgements

The work of the second author is supported by CNPq (Grant 302714/88-0).

References

- Dubuc S (1986). Interpolation through an Iterative Scheme. *J. Math. Anal. Appl.* **114**, pp 185-204.
- Harten A (1994). Adaptive Multiresolution Schemes for Shock Computations. *Journal of Computational Physics* **115** (2), pp 319-338.
- Harten A (1995). Multiresolution algorithms for the numerical solution of hyperbolic conservation laws. *Comm. Pure Appl. Math.* **XLVIII**, pp 1305-1342.
- Johansen T and Winther R (1988). The solution of the Riemann problem for a hyperbolic system of conservation laws modelling polymer flooding. *Siam J. Math. Anal.* **19** (3), pp 541-566.
- LeVeque R J (1990). Numerical Methods for Conservation Laws. Birkhäuser Verlag.
- Holmström M (1997). Wavelet based methods for time dependent PDEs. Doctoral Dissertation, Uppsala University, Sweden.
- G-Müller B and S Müller (1998). Adaptive finite volume schemes for conservation laws based on local multiresolution techniques. *Proceeding of 7th International Conference on Hyperbolic Problems*, Zurich, Birkhäuser-Verlag.

APPLICATION OF A GODUNOV-TYPE ALE-METHOD TO UNDERWATER SHOCK WAVES

A. KLOMFASS, P. NEUWALD, K. THOMA

Fraunhofer Institut für Kurzzeitdynamik

(Ernst-Mach-Institut)

Eckerstr.4, 79104 Freiburg, Germany

Email: klomfass@emi.fhg.de

Abstract.

A second order, explicit Godunov-type finite-volume method has been modified for application to underwater shock wave studies. For this purpose an adequate equation of state has been derived, which includes a simple cavitation model. The propagation of an explosively generated shock wave in a water-filled steel tube has been used as an initial test case for this method. The paper provides the relevant details of the numerical method and describes the results of the numerical and experimental investigation.

1. Introduction

The study of underwater shock waves is relevant to many areas of technology, including naval defence and off-shore engineering as well as industrial and medical applications. Thus numerical simulation is in strong demand. In the field of gasdynamics Godunov-type methods are established tools for the simulation of transient multidimensional flows. Together with a high order extension, as realized in TVD or ENO schemes, they attain both high accuracy and computational efficiency.

In contrast to a gas, the compressibility of the fluid water is very small and considerable pressure variations are already caused by minor density changes. The propagation of even strong shocks through water therefore leads to comparatively little fluid convection and thus poses nearly acoustic problems. Nevertheless, a "standard" Godunov-type method may be applied to underwater shock studies, as was done here with the APOLLO-code (Klomfass, 1999) of the Ernst-Mach-Institut, which was originally de-

veloped for gasdynamic computations. The only major modifications concerned the equation of state and the issues of cavitation and free surfaces. A special feature of the code is the ALE (Arbitrary Langrangian-Eulerian) formulation which allows computations on arbitrarily time dependent grids. This offers the capability to simulate free surfaces as well as fluid-structure interactions.

The study presented here served as the initial validation step toward the application on underwater shocks. In the investigated configuration the structural response was sufficiently small, so that the computations could be carried out under the assumption of rigid walls, i.e. on time-independent grids.

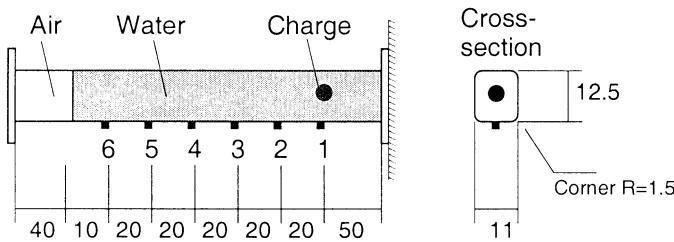


Figure 1. Geometry of the validation test (steel tube mounted vertically).

A schematic of the configuration is shown in Fig.1. The vertically mounted steel tube (wall thickness 3 cm) was filled with fresh water at a temperature of 291 Kelvin. The shock wave was generated by the detonation of a spherically shaped miniature charge of 320 mg PETN, which was positioned in the center of the cross section 50 cm above the lower end of the tube. The free water surface is covered with air at ambient pressure. The instrumentation of the tube consisted of 6 pressure transducers, which were mounted in shock-insulated suspensions in order to decouple them from the structural response.

2. Finite-Volume Method

The physical model underlying the APOLLO-Code are the conservation equations for a compressible, inviscid and non heat-conducting fluid. Applied to an arbitrary, time-dependent control volume the equations are:

$$\frac{d}{dt} \int_V \mathbf{u} dV = \oint_S (\mathbf{u} \circ (\vec{v} - \vec{w}) + \mathbf{P}) \vec{n} dS, \quad (1)$$

with:

$$\mathbf{u} = (\rho, \rho \vec{v}, \rho e^{tot})^T, \quad \mathbf{P} = (0, p\mathbf{I}, p\vec{v})^T.$$

Here d/dt is the derivative with respect to the control volume V , the surface of which, S , moves with a non-uniform velocity field \vec{w} . This is the ALE formulation of the conservation equations in integral form. The Lagrangian and the Eulerian formulation are recovered from it for $\vec{w} = \vec{v}$ or $\vec{w} = 0$, respectively.

For the numerical solution an explicit time integration of a cell-centered finite-volume approximation is used. The spatial discretization works on block-structured, body-fitted hexahedral grids and a one-step Euler-scheme is applied for the time-integration. The discretized equations are thus,

$$V_i^{n+1} \mathbf{U}_i^{n+1} = V_i^n \mathbf{U}_i^n + \Delta t \sum_{l(i)} \mathbf{F}_l^n S_l^n, \quad (2)$$

where \mathbf{U} denotes the average values of \mathbf{u} within a grid cell i . The surface areas and the normal unit-vectors of cell interfaces $l(i)$ are S_l and \vec{n}_l , respectively. The numerical fluxes at the cell interfaces are defined as

$$\mathbf{F} = \mathbf{u}(v - w) + \mathbf{p}, \quad \mathbf{p} = (0, p\vec{n}, pv)^T, \quad (3)$$

where $v = \vec{v}\vec{n}$ and $w = \vec{w}\vec{n}$ are the normal components of the material and the interface velocity. To preserve conservation and consistency in the case of time dependent grids, the cell volumina at the new time level are calculated by the same integration scheme as the flow field, i.e. from

$$V_i^{n+1} = V_i^n + \Delta t \sum_{l(i)} w_l^n S_l^n. \quad (4)$$

Upon evaluation of the \mathbf{U}_i^{n+1} , the positions of the grid nodes are updated and the cell volumina and the other metric quantities are re-evaluated exactly for the next time step. The velocities of the cell interfaces are determined by the instantaneous velocities of the grid nodes, which may either be prescribed or determined from a fluid-structure interaction model.

3. Flux Calculation

The method of flux calculation follows closely the ideas established in the HLLC-solver suggested by Toro et al., (Toro, 1997). For each cell interface a standard 3-wave Riemann problem (regions 1 - 4 separated by left pressure wave, contact, right pressure wave) is set up and evaluated in the direction of the interface normal. Both pressure waves are considered as discontinuities and thus described by the Rankine-Hugoniot relations (RHR). Together with the contact conditions the RHR render the normal velocity and the pressure across the contact wave:

$$v_{23} = \frac{p_1 - p_4 + q_1 v_1 - q_4 v_4}{q_1 - q_4}, \quad (5)$$

$$p_{23} = \frac{p_4 q_1 - p_1 q_4 + q_1 q_4 (v_4 - v_1)}{q_1 - q_4}. \quad (6)$$

With the wave speed estimates c_L and c_R for the left and the right pressure waves, the abbreviations q_1 and q_2 are defined as

$$q_1 = \rho_1(v_1 - c_L), \quad q_4 = \rho_4(v_4 - c_R). \quad (7)$$

Using the above results for the pressure related parts \mathbf{p} of the fluxes in region 2 or 3 and eliminating the unkowns \mathbf{u}_2 and \mathbf{u}_3 with the RHR, the flux at the moving cell interface can be expressed as

$$\mathbf{F} = \mathbf{u}_1 g_1 + \mathbf{u}_4 g_4 + \mathbf{p}_f \quad (8)$$

with:

$$\begin{aligned} \mathbf{p}_f &= s_1 \mathbf{p}_1 + s_2 (\mathbf{p}_{23} - \mathbf{p}_1) (c_L - w) / (c_L - v_{23}) \\ &\quad + s_4 \mathbf{p}_4 + s_3 (\mathbf{p}_{23} - \mathbf{p}_4) (c_R - w) / (c_R - v_{23}), \end{aligned}$$

$$\begin{aligned} g_1 &= s_1 (v_1 - w) + s_2 (v_{23} - v_1) (c_L - w) / (c_L - v_{23}), \\ g_4 &= s_4 (v_4 - w) + s_3 (v_{23} - v_4) (c_R - w) / (c_R - v_{23}), \end{aligned}$$

$$\begin{aligned} s_1 &= 1, \quad v_{23} > w, \quad s_2 = 1, \quad v_{23} > w \wedge c_L < w, \\ s_4 &= 1, \quad v_{23} \leq w, \quad s_3 = 1, \quad v_{23} \leq w \wedge c_R > w. \end{aligned}$$

The switch functions $s_i, i = 1, 4$ ensure that the correct region of 1 to 4 is chosen, depending on the velocity of the cell interface and the wave speeds. The switches take on unit values under the conditions specified and are otherwise zero. Wave speed estimates $c_L = v_1 - a_1$ and $c_R = v_4 + a_4$, with sound speeds a_1 and a_4 , are used here. They provide exact flux solutions in the acoustic limit and yield satisfactory results even for strong shocks and expansions.

For a second order accurate flux calculation the initial values $\mathbf{u}_1, \mathbf{p}_1$ and $\mathbf{u}_4, \mathbf{p}_4$ in the Riemann problem are obtained from extrapolations of the cell average-values to the cell interfaces. For this purpose a piece-wise linear reconstruction of \mathbf{U} is carried out, where the MINMAX limiter is applied to the linear slopes. An exception from this standard procedure has been introduced for the total energy variable. Here the limiter operator is applied separately to the slopes of ρe and $\rho \vec{v} \cdot \vec{v} / 2$. The sum of the two slopes is then used to extrapolate the total energy. This modification improves the robustness and diminishes slight pressure and velocity oscillations, which in the standard scheme usually arise at contact surfaces.

4. Equation of State

Liquid water is here described by the Mie-Grüneisen equation of state (EOS),

$$p = p_H(\rho) + \Gamma \rho_0 (e - e_H(\rho)), \quad \rho \geq \rho_0 \quad (9)$$

where $p_H(\rho), e_H(\rho)$ defines a reference curve in the $p - e - \rho$ surface. A suitable choice for the reference curve is the Hugoniot, which emerges from some reference point p_0, e_0, ρ_0 . Given an empirical relation $c = a_0 + sv$ between shock velocity c and the material velocity v behind the shock wave, the Hugoniot becomes

$$p_H(\rho) = p_0 + K \frac{\eta}{(1 - s\eta)^2}, \quad e_H(\rho) = e_0 + \frac{1}{2} (p_0 + p_H) \frac{\eta}{\rho_0}, \quad (10)$$

where $\eta = 1 - \rho_0/\rho$ and $K = \rho_0 a_0^2$ is the bulk modulus of the fluid. This type of EOS, also known as "Shock-EOS", is frequently used in Hydro-Code computations on shock waves in liquids and solids, (CDL, 1998). The three parameters a_0 , s and Γ are to be determined experimentally. The sound speed a_0 of fresh water at 291 Kelvin was here found to be 1350 m/s. The values of the other parameters were taken from (Marsch, 1980). They are: $s = 1.7$, $\Gamma = 0.28$.

To account for cavitation the EOS is extended under the following assumptions: A) cavitation occurs upon a density decrease below the reference value ρ_0 , where an instantaneous change into the gaseous phase occurs. This means that no voids are formed in the cavitation process, and that cavitation always occurs at a positive pressure. B) For the gaseous phase the equation of state

$$p = \Gamma \rho e \quad , \quad \rho < \rho_0 \quad (\eta < 0), \quad (11)$$

applies, where Γ is the same as in the liquid phase. Continuity between both phases requires, that $p_0 = \Gamma \rho_0 e_0$. Hence the reference point of the Hugoniot must be a point on the cavitation threshold. For a given cavitation density ρ_0 the cavitation pressure p_0 is therefore adjusted by the value used for the ambient internal energy e_0 of the liquid water. With e_0 fixed in that way, the ambient pressure of the liquid water is set by the density $\rho > \rho_0$. Here values of $\rho_0=1$ g/cm³ and $p_0=0.2$ bar were used.

Upon cavitation the sound speed drops discontinuously from the high values in the liquid phase to very small values (depending on e_0) in the gaseous phase. Too small a value of the cavitation pressure may thus lead to numerical difficulties, as the computed internal energy in an expanding cavitation bubble may become negative and the sound speed undefined.

5. Free Surfaces

Free surfaces are treated as a special kind of far-field boundaries, which are implemented via dummy cells placed at the outer margins of the grid. To model a free surface the ambient conditions of the medium that covers the surface are assigned to the dummy cells. For the present application this medium is air at atmospheric conditions.

Free surfaces may be either traced or regarded as standard far-field boundaries. In the first case, the velocities of the grid nodes at the free surface are adapted to the material velocities at the cell interfaces. The grid thus deforms according to the motion of the surface and no mass flux through the surface takes place. In the second alternative, which was applied here, the grid nodes at the free surface remain fixed. The fluid may therefore leave the computational domain or the ambient medium may enter it. This approach is applicable, when the deformation of the free surface is small compared to the characteristic lengths of the overall geometry. As the APOLLO-code has no multi-fluid capability, the inflow of ambient material is only possible in the case of a gas, which is for this purpose modelled as gaseous water.

6. Numerical and Experimental Results

The initial phase of the underwater detonation was computed as spatially one-dimensional problem in spherical symmetry. This phase covers the time interval up to the moment when the spherically propagating shock wave reaches a tube wall. The spherical fields were then remapped onto the spatially three dimensional grid, which covered the complete fluid volume (18x16x435 nodes on a quarter segment of the tube). The calculations for the initial phase were performed with the AUTODYNTM-Code, (CDL, 1998), which offers the required detonation model. In the remapping procedure, the gaseous detonation products of PETN were “converted” to gaseous water. This was done by altering the respective part of the internal energy profile such, that the original pressure profile was preserved.

A comparison of calculated and measured pressure records is shown in Fig.2. Three identical experiments have been carried out, which displayed an excellent reproducibility. For clarity only one set of data is shown here (dashed lines). The amplitudes and the main signature of the calculated pressure time-curves fit the experimental data reasonably well. An exception is the wave reflected at the bottom of the tube (arrival time ≈ 0.7 ms at Pos.1), the amplitude of which is less than the calculated values. This can be attributed to the elastic response of the floor to which the tube was mounted.

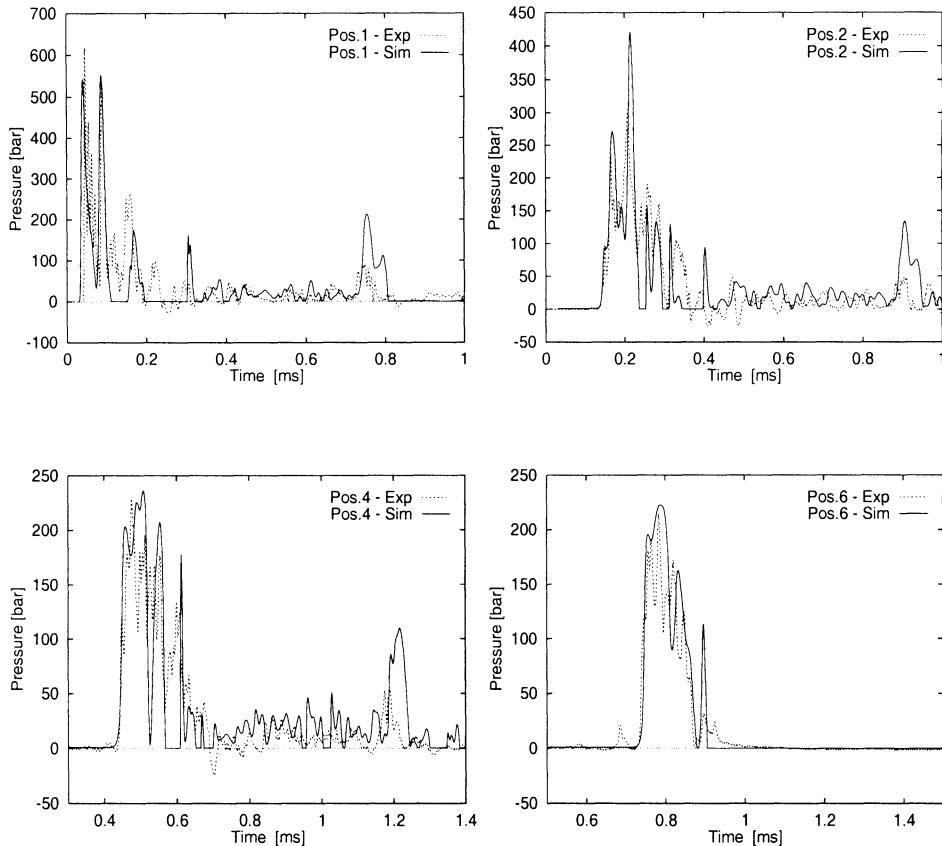


Figure 2. Experimental and numerical pressure records at gauges 1, 2, 4 and 6

7. Summary

The initial validation test presented in this paper confirmed the modification of the APOLLO-code for application to underwater shock waves. The computational model was found adequate for explosively generated underwater shocks. Further experiments and calculations were meanwhile carried out, which addressed the issue of fluid-structure interactions. The results of these tests, which could not be included here, also confirmed the numerical method.

References

- Klomfass A (1999). APOLLO 3D Flow Solver - Theory and User Manual. Internal Report, Ernst-Mach-Institut.
- Toro E F (1997). Riemann Solvers and Numerical Methods for Fluid Dynamics. First Edition, Springer-Verlag.
- N. N. (1998) Autodyn User and Theory Manual. Century Dynamics Ltd. Horsham, UK.
- Marsh S P (Editor) (1980). LASL Shock Hugoniot Data. University of California Press.

NUMERICAL SIMULATION OF 2-D TWO-PHASE FLOWS WITH INTERFACE

S. KOKH

DRN/DMT/SERMA,

CEA Saclay, 91191 Gif-sur-Yvette, France.

AND

G. ALLAIRE

Laboratoire d'Analyse Numérique,

Université Paris-VI, 75252 Paris Cedex 5, France.

Abstract

This paper is devoted to the direct numerical simulation of compressible two-phase flows, *i.e.* including material interfaces, in an Eulerian framework. Eulerian methods, such as Volume Of Fluid, are easy to handle but suffer from numerical diffusion which spreads out the precise localization of the interface. We discuss some remedies to this loss of accuracy.

1. Introduction

Modelization and simulation of bifluid and diphasic flows have become of increasing interest among the computational fluid dynamic community. So-called direct simulation, on the contrary of average models, involves the description of the interface between fluids which is a discontinuity surface for the material properties. We propose a model for compressible two-phase flows. The dynamical aspect of the problem is handled by the compressible Euler equations, written for the overall mixture, while the phase interface is captured on an Eulerian mesh. An extra equation is therefore added to the Euler system in order to advect values of a color function ψ according to the fluids motion. This type of systems has been studied by Abgrall (Abgrall R, 1988) and Karni (Karni S, 1996). This approach provides key features such as no extra complexity in dealing with “high-dimensional” problems, easy handling of drastic topological changes or complex topology of the interface

between the two phases. The other side of the coin is a lack of accuracy for the interface description in contrast to methods involving interface reconstruction such as Front Tracking. In fact, numerical diffusion tends to thicken the interface into a transition zone which is no longer a sharp discontinuity between the media. We focus on the numerical diffusion problem near contact discontinuities. Simple observations for a single transport equation lead us to propose various processes for improving the accuracy of the interface description while preserving the ease-of-use of interface capturing methods.

2. Bifluid solver: model and numerical treatment

The model used here follows the line of Abgrall in (Abgrall R, 1988). The motion of the fluids is here driven by the compressible Euler equations (1)–(3). It is supplied by the color function transport equation (4) and an equation of state (EOS) that closes the system. For the sake of simplicity we expose the method in the case of a perfect gas EOS, despite the process is still valid for more general laws such as Stiffened Gas. Numerical diffusion implies that some cells do contain both species. Upon an isothermal assumption, Abgrall showed how to construct an EOS of the form (5) in order to deal with the mixture zone and which reduces to the usual perfect gas EOS in pure fluid areas.

$$\partial_t \rho + \operatorname{div}(\rho \vec{u}) = 0, \quad (1)$$

$$\partial_t \rho \vec{u} + \operatorname{div}(\rho \vec{u} \otimes \vec{u}) + \overrightarrow{\operatorname{grad}} p = \vec{F}, \quad (2)$$

$$\partial_t \rho e + \operatorname{div}[(\rho e + p) \vec{u}] = Q, \quad (3)$$

$$\partial_t \psi + \vec{u} \cdot \overrightarrow{\operatorname{grad}} \psi = 0, \quad (4)$$

$$p = p(\rho_1, \rho_2, \rho(e - |\vec{u}|^2/2)). \quad (5)$$

$\vec{F} = (F_x, F_y)$ and Q are source terms such as gravity, viscosity, surface tension, and thermal diffusion.

We briefly describe the numerical method used to implement the model. The Euler system is solved thanks to a Roe-type scheme (Roe P L, 1981), as done in (Abgrall R, 1988) for the variables $(\rho, \rho \vec{u}, \rho e)$. As mentioned by Karni in (Karni S, 1994), conservative schemes have difficulties to describe accurately the pressure near the interface generating spurious oscillations. To remedy this drawback, Abgrall proposes in (Abgrall R, 1988) to judiciously choose $\psi = 1/\kappa$ (κ being the Grünsein constant) as color function and derives a discretization of (4) that preserves numerical contact discontinuities. Gravity is treated as a centered source term, viscosity and thermal diffusion are discretized by standard finite differences while we use the continuous surface tension model of (Brackbill J, Kothe D and Zemach C, 1992)

for the interfacial tension. Second order accuracy in space is reached thanks to a MUSCL method with a minmod limiter. As for second order in time, we use a two-point Runge-Kutta method. Unfortunately even second order accuracy in space and time cannot help to decrease numerical diffusion for long. Indeed even if the numerical scheme succeeds in picturing the behaviour of the system, it may happen that an entire fluid component just disappears into the mixture zone. Let us emphasize that mixture zones do not have necessarily any real physical sense. Furthermore it may also become very difficult to describe jumps of variables across the interface which are of high interest for modeling physical process such as mass transfer due to phase changes. In the sequel we propose various methods to maintain the sharpness of contact discontinuities (material interfaces).

3. The transport equation model

We focus in this section on the specific problem of numerical diffusion of finite difference schemes near contact discontinuities. To begin with, let us underline that the structure which drives the contact discontinuity is a linearly degenerated field. Harten in (Harten A, 1978) enlightens the behavior of a discontinuity line advected by such field and approximated by a classical numerical scheme. The width of the numerical diffusion will inexorably grows as the number of time steps increases, on the contrary to shocks driven by genuinely non-linear fields which are enclosed in a viscous profile. The most simple equation that can mimic the critical behavior of such fields is a simple linear transport equation at constant speed. Let $u(t, x)$ be the solution of

$$\partial_t u + c \partial_x u = 0, \quad \forall x \in \mathbb{R}, \quad \forall t > 0 \quad (6)$$

with the initial condition $u(0, x) = u_0(x)$, c being a constant velocity. The exact solution is $u(t, x) = u_0(x - ct)$. We are interested in the case where u_0 is a step function, and study the numerical diffusion associated to a given numerical scheme. We first recall the influence of order accuracy upon numerical diffusion. All computations are done with an upwind scheme for $c = 1$ on a segment $I = [0, 1]$, meshed by 1000 regular cells, with periodic boundary conditions and $u_0 = \mathbb{1}_{[1/4, 3/4]}$. Second order in space and time are respectively implemented via a MUSCL method with minmod limiter and a two-points Runge-Kutta method. As expected, for the first order scheme the L^1 -norm of the error grows like \sqrt{n} , where n is the number of time steps. When switching to second order in space, the error L^1 -norm stops increasing after a few time steps. However, for both time and space second order, the numerical diffusion of the scheme grows again unbounded. Thus, second order in space with first order in time would be quite satisfactory, but in many cases second order in time is necessary to stabilize numerical

oscillations. For example, the simple computation of the hydrostatic pressure establishment upon the influence of gravity in a single fluid turns out to be impossible due to instabilities. Such simple examples motivate our study of procedures to bound this extra diffusion. The level set method provides a way to exactly control the thickness of the interface. It uses instead of the discontinuous color function a continuous function initialized as the signed distance to the interface as exposed in (Sethian J, 1996) or (Mulder S, Osher S and Sethian J, 1992). This function is frequently reinitialized during the computation by solving a suitable Hamilton-Jacobi equation as mentioned in (Sethian J, 1996) and (Sussman M, Smereka P and Osher S, 1994). Here, staying in the framework of the VOF method, we propose to add source terms in order to straighten up the front.

4. Sharpening source terms

To begin with, we introduce in (6) a source term $P(u) = \eta u(1-u)(u-1/2)$ where η is a real parameter. This source term does not modify the exact solution of this equation since it can only take the values 0 or 1. However, in the discrete problem it will act as a “repelling force” on the approximate solution. Values above 1/2 will be pushed towards 1, while those below will get closer to 0. Two numerical implementations of this source term are possible. First of all, it can simply be added into the discretized equation as a centered source term. Alternatively, a splitting-like method can be chosen: the classical upwind scheme resolution is stopped after N time steps then the approximated solution u^N at instant N is sharpened by solving $\partial_s v = P(v)$ with initial condition $v(0, x) = u^N$ until it reaches a steady state (s is an artificial time variable). Fortunately, there exists explicit solutions of this ODE, thus no new extra computational work is required. A second type of source term can be obtained by changing the constant η into a variable quantity $\eta \partial_x u$, which yields $Q(u, \partial_x u) = \eta u(1-u)(u-1/2) \partial_x u$. In this case, equation (6) can also be rewritten

$$\partial_t u + \partial_x [cu + (\eta/4)u^2(1-u)^2] = 0$$

Actually, this appears to be a flux modification of (6) which fits into the framework of the artificial compression method developed by Harten in (Harten A, 1978). Figure 1 displays a comparison of the growth for the L^1 -error between the different methods. Accuracy gain is obvious for the first order method as the L^1 -error stops growing after a few time steps. While for second order in space only the process doesn’t show real improvement, it succeeds in slowing down the error growth second order is applied to both space and time. The efficiency of such sharpening process can be quantified by studying the equivalent equation of the numerical scheme. Indeed, for

the case of an upwind first order scheme, a viscosity profile (having the shape of a tanh function) can be explicitly computed for the initial value problem (6).

5. Application to the bifluid model

The sharpening process described earlier is applied to the advection equation for $\psi = 1/\kappa$ in system (1)–(5) introducing the source term

$$Q(\psi) = -\eta(\psi - \psi_1)(\psi - \psi_2)(\psi - (\psi_1 + \psi_2)/2)$$

in the scheme as a centered source term. Figure 2 shows the effect of sharpening ($\eta = 0.2$) on the computing of a shock into an helium bubble surrounded by air (Abgrall R, 1996) on a $1\text{m} \times 1\text{m}$ mesh discretized in 100×100 regular cells. Notice that the other variables seem to be unaffected by the sharpening source term.

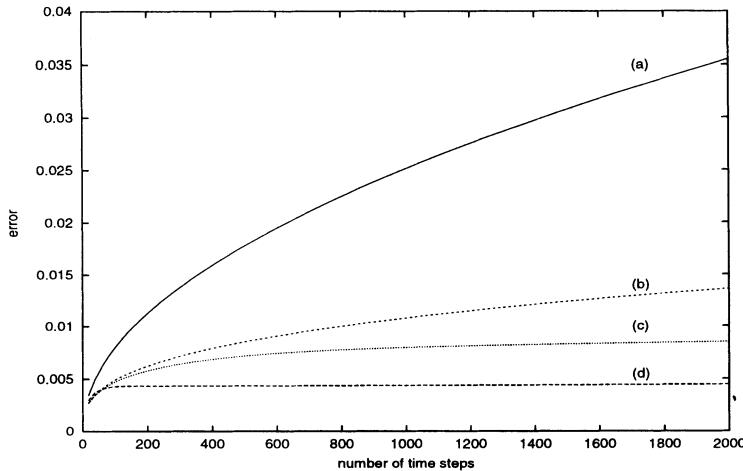


Figure 1. L^1 -error for an impulse advection after 2000 time steps, (a): order one, (b):order 2 scheme in both time and space, (c) order 2 scheme in both time and space with centered source term correction for $\eta = 5 \times 10^{-3}$, (d): order one with centered source term correction for $\eta = 0.1$.

6. Conclusion and perspectives

We have developed a simple method for sharpening the advection of discontinuities in finite differences numerical schemes. This method is easy to use and adds no extra complexity when dealing with 2-D and 3-D problems. The primary goal is to improve the localization of material interfaces in compressible two-phase flow simulation. Important variables, such as pressure, density and velocity, do not seem to be noticeably affected by this sharpening. We hope to be able to derive new further estimates for second order

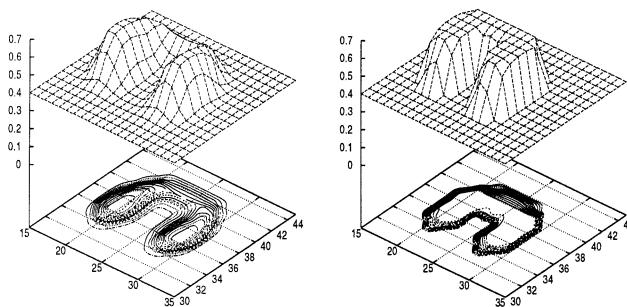


Figure 2. profile for κ , without sharpening on the left, with sharpening on the right, at instant $t = 0.35s$

schemes in both time and space. Concerning applications to two-phase flows with phase change at the interface, we have obtained preliminary results by adding a kinetic relation to determine the interface velocity as described in (Truskinowsky L, 1991). Alternate advection schemes for the color function such as level set and characteristic methods have been implemented, while the extension of the model to stiffened gas is in progress. This will be the topic of future reports.

References

- Abgrall R (1996). How to prevent pressure oscillations in multicomponent flow calculations: a quasi conservative approach. *J. Comp. Phys.*, **125**, pp 150–160.
- Abgrall R (1988). Généralisation du schéma de Roe pour le calcul d'écoulements de mélanges de gaz à concentrations variables. *La Recherche Aérospatiale*, **6**, pp 31–43.
- Brackbill J, Kothe D and Zemach C (1992). A continuum method for modelling surface tension. *J. Comp. Phys.*, **100**, pp 335–354.
- Godlewski E and Raviart P A (1996). Numerical Approximation of Hyperbolic Systems of Conservation Laws Springer, Applied Mathematical Sciences, vol. 118.
- Harten A (1978). The Artificial Compression Method for Computation os Shocks and Contact Discontinuities: III-Self-Adjusting Hybrid Schemes. *Math. Comp.*, **142**, pp 363–389.
- Karni S (1996). Hybrid multifluid algorithms. *SIAM J. Sci. Comput.*, **5**, pp 1019–1039.
- Karni S (1994). Multicomponent flow calculations by a consistent primitive algorithm. *J. Comp. Phys.*, **112**, pp 31–43.
- Mulder S, Osher S and Sethian J (1992). Computing interface motion in compressible gas dynamics. *J. Comp. Phys.*, **100**, pp 209–228.
- Roe P L (1981). Approximate Riemann solvers, parameter vectors, and difference schemes. *J. Comp. Phys.*, **43**, pp 357–372.
- Sethian J (1996). The level set method. Cambridge University Press.
- Sussman M, Smereka P and Osher S (1994). A level set method for computing solutions to incompressible two-phase flows. *J. Comp. Phys.*, **114**, pp 146–159.
- Truskinowsky L (1991). Kinks versus shocks. in *Shock induced transitions and phase structure in general media*, Fosdick R et al. eds., Springer Verlag, Berlin.

RELATIVISTIC MHD SIMULATIONS USING A GODUNOV-TYPE METHOD

SERGUEI KOMISSAROV

*Department of Applied Mathematics,
University of Leeds,
Leeds LS2 9JT,
U.K.*

Emails: serguei@amsta.leeds.ac.uk

Abstract. The Godunov approach is found to be very successful in relativistic gas dynamics where other more “traditional” methods fail. However, in many interesting astrophysical applications one cannot ignore the dynamical effects of magnetic field. This explains the growing interest in the development of Godunov-type numerical schemes for relativistic MHD. In this paper we outline our results in this area and describe our first astrophysical numerical simulations.

1. Relativistic MHD

We only consider the case of a space-time with a specified time-independent metric form. Once we have chosen the set of coordinates $\{t, x^i\}$, where $t = \text{const}$ describes a space-like hypersurface and ∂_t is time-like at infinity, the system of ideal MHD equations can be written as follows:

$$\text{Continuity: } \partial_t (\sqrt{-g} \rho u^t) + \partial_i (\sqrt{-g} \rho u^i) = 0; \quad (1)$$

$$\text{Energy-Momentum: } \partial_t (\sqrt{-g} T_\alpha^t) + \partial_i (\sqrt{-g} T_\alpha^i) = \frac{\sqrt{-g}}{2} \partial_\alpha (g_{\mu\nu}) T^{\mu\nu}; \quad (2)$$

$$\text{Maxwell: } \partial_t (\sqrt{-g} F^{*\nu t}) + \partial_i (\sqrt{-g} F^{*\nu i}) = 0. \quad (3)$$

Here ρ is the rest mass proper density, u^ν is the fluid 4-velocity, $T^{\mu\nu}$ is the energy-momentum tensor, $g_{\mu\nu}$ is the metric tensor, $F^{*\mu\nu}$ is the dual 2-vector of electromagnetic field:

$$F^{*\mu\nu} = \frac{1}{2} e^{\mu\nu\alpha\beta} F_{\alpha\beta}, \quad (4)$$

$F_{\alpha\beta}$ is the 2-form of electromagnetic field and $e^{\mu\nu\alpha\beta} = -\frac{1}{\sqrt{-g}}\epsilon^{\mu\nu\alpha\beta}$ is the Levi-Civita alternating tensor. In the limit of infinite conductivity one has

$$F_{\mu\nu} u^\nu = 0, \quad (5)$$

that allows us to introduce the 4-vector of magnetic field

$$b^\mu = -F^{*\mu\nu} u_\nu \quad (6)$$

and to write down F^* and T in the following concise form (Anile, 1989):

$$F^{*\mu\nu} = b^\mu u^\nu - u^\nu b^\mu \quad (7)$$

$$T^{\mu\nu} = (w + b^2)u^\mu u^\nu + (p + b^2/2)g^{\mu\nu} - b^\mu b^\nu. \quad (8)$$

(1)-(3) constitute a system of 8 evolution PDEs for 8 unknowns, e.g. ρ , p , u^t , and b^i , where $i = 1, 2, 3$.

Quite generally, equations (3,5) can be written in exactly the same form as the corresponding equations of classical MHD in cartesian coordinates:

$$\partial_t B^i + \epsilon^{ijk} \partial_j E_k = 0, \quad (9)$$

$$\partial_i B^i = 0, \quad (10)$$

$$E_k = -\epsilon_{kij} v^i B^j, \quad (11)$$

where the electric field, E_i , the magnetic field, B^i , and the 3-velocity, v^i are defined as follows

$$E_i = F_{it}, \quad B^i = \frac{1}{2} \epsilon^{ijk} F_{jk}, \quad v^i = u^i/u^t. \quad (12)$$

Here $\epsilon^{ijk} = \epsilon_{ijk}$ is the 3-dimensional Levi-Civita alternating symbol.

2. Numerical scheme

Like the equations of gas dynamics the evolution equations (1)-(3) of MHD constitute a hyperbolic system of quasi-linear conservation laws which can be written in the following integral form

$$\frac{d}{dt} \int_V \mathbf{Q} \sqrt{-g} dV + \int_{\delta V} \mathbf{F}^i \sqrt{-g} dS_i^* = \int_V \mathbf{S} \sqrt{-g} dV. \quad (13)$$

This similarity suggests that in MHD the Godunov method can be as successful as in gas dynamics. Although, the MHD Riemann problem is rather complicated and its exact solution is computationally expensive a much cheaper linear Riemann solver suffices for most problems. The 1D scheme of such type described in (Komissarov, 1999a) is quite robust and efficient.

Unfortunately, the multidimensional extension of Godunov-type schemes in MHD is not that straightforward as in gas dynamics. The standard prescription (Roe, 1998) produces a scheme which fails in regions of highly nonuniform magnetic field (e.g. at shocks!). The reason seems to be related to the fact that such a scheme does not preserve magnetic field divergence free (Brackbill and Barnes, 1980). Indeed, the normal component of magnetic field at the cell interfaces cannot be found via the Riemann problem but needs to be determined by other means. Obviously, magnetic Lorentz force of numerical origin will be introduced if the way in which the normal components are determined does not accommodate the condition of vanishing magnetic flux over the cell surface. The way we use to overcome this problem is to evolve the normal components of magnetic field at the cell interfaces using the integral form of equation 9,

$$\frac{d}{dt} \int_S B^i dS_i^* + \int_{\delta S} E_i dx^i = 0, \quad (14)$$

in the same manner as (13) is used to evolve the cell centred variables. When applied to the interfaces of computational cells this equation ensures the conservation of magnetic flux over the cell interfaces and ensures that the integral $\int_S B^i dS_i^*$ over the cell surface vanishes (Evans and Hawley, 1988). Obviously, this requires the use of a staggered grid. At the beginning of each time step the cell-centred magnetic field is computed using linear interpolation of the interface-defined field and then the evolution of conserved variables is computed following the standard prescription (Falle and Komissarov, 1996). Moreover, the solutions to the Riemann problems at the cell interfaces are used to interpolate v^i and B^i to the edges of the interfaces and, thus, to determine the fluxes of magnetic field lines through the edges. Since we have to recalculate the total energy and momentum of the cells the scheme is not quite conservative. However, it seems to retain all the useful properties of Godunov-type schemes and captures shocks quite accurately. Most of the details can be found in (Komissarov, 1999a).

3. Results of numerical simulations

Here we present the results of our recent numerical simulations of relativistic MHD flows of some astrophysical importance.

3.1. PROPAGATION OF RELATIVISTIC MAGNETISED JETS

Here we briefly describe the results of our study of the effects of purely toroidal magnetic field on the propagation of axisymmetric relativistic jets (Komissarov, 1999b).

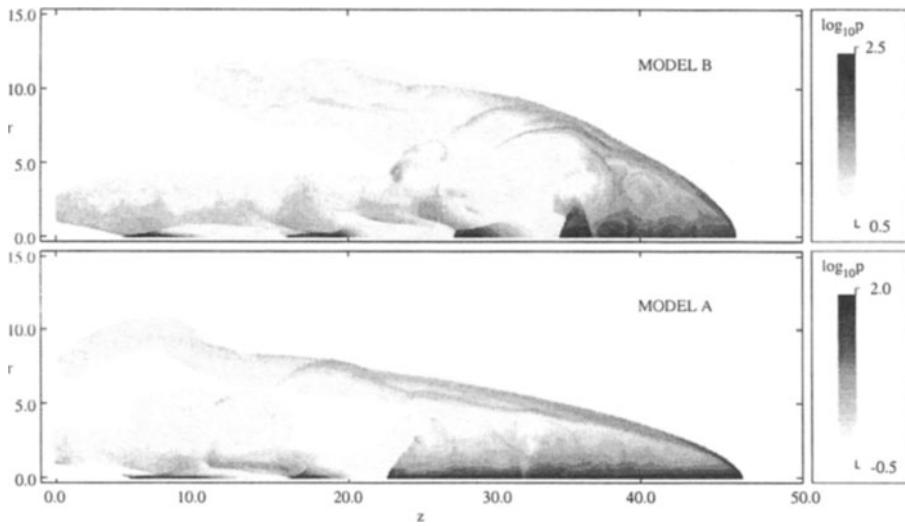


Figure 1. Propagation of relativistic jets with toroidal magnetic field. Gas pressure at the end of the simulations. Top panel: the kinetic-energy-dominated jet behaves almost like a non-magnetic jet. Bottom panel: the jet with comparable kinetic and Poynting fluxes develops conspicuous “nose cone” ($22 < z < 47$).

The computational domain is $0 < r < 15$, $0 < z < 50$. The computational grid is uniform with a mesh spacing $\Delta = 0.05$. The initial jet radius is $r_j = 1$ which corresponds to 20 cells. At $z = 0$ we use symmetry (or reflection) boundary conditions for $r > r_j$ and inflow boundary conditions for $r < r_j$. At the other boundaries we also use symmetry boundary conditions and make sure that the bow shock does not reach the outer boundaries during the simulations. Initially the domain is filled with a uniformly distributed unmagnetised gas. At the inlet the jet velocity is parallel to the z -axis and uniform. Its density is also uniform. In order to compare our simulations with the nonrelativistic simulations (Lind et al., 1989) we used the same pressure and magnetic field profiles at the inlet as they did. Throughout the whole domain we use the equation of state of a polytropic gas with $\Gamma = 4/3$.

We have computed two models, A and B, both with the same initial Lorentz factor, $\gamma_j = 10$, the same mean ratio of the gas pressure to the magnetic pressure, $\bar{\beta} = 1.0$, the same ratio of the mass densities of the jet and the external gas, $\rho_j \gamma_j^2 / \rho_e = 10^{-1}$, but the different ratios of the rest mass-energy and the magnetic energy densities, $\sigma = 2.9$ in model A and $\sigma = 29$ in model B. As a result, in model B the kinetic energy flux of the jet dominates all other kinds of energy fluxes and in model A it is comparable to

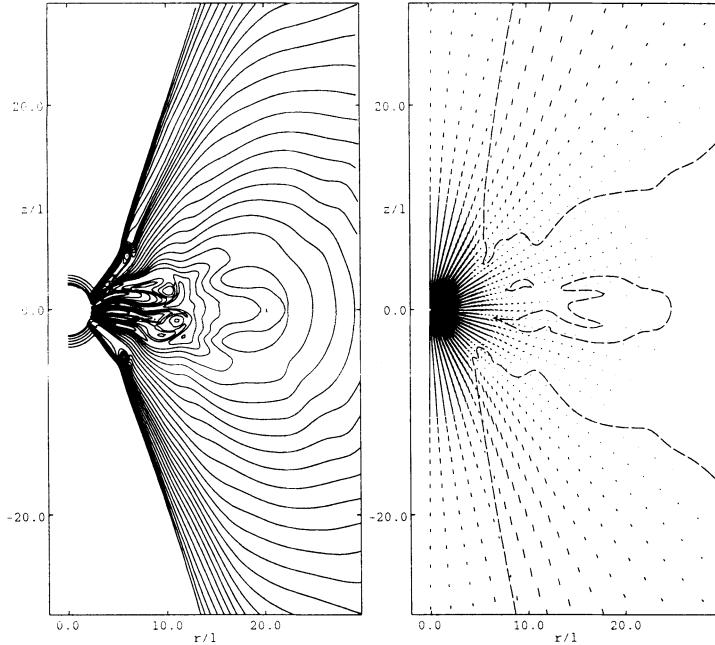


Figure 2. Magnetically driven accretion onto a non-rotating black hole. Left panel: rest mass density field. There are 20 logarithmic contours equally spaced within three orders of magnitude. Right panel: velocity field. The arrows show the direction of the 3-velocity vector as measured by the local FIDOs. Their length is proportional to the velocity magnitude. The dashed line shows the surfaces of vanishing radial component of velocity. Near the z -axis the flow is directed towards the black hole.

the Poynting flux. Figure 1 shows the pressure distributions in both models by the end of the simulations. While the results for model B are similar to those obtained for unmagnetised jets with similar parameters, the jet in model A develops a conspicuous ‘‘nose cone’’, the high pressure region of the shocked jet plasma accumulated in the jet head, similar to the one found earlier in the simulations of Newtonian magnetised jets (Lind et al., 1989; Clarke et al., 1986; Kössl et al., 1990; van Putten, 1996).

Thus, our results reveal that the development of a nose cone depends not on the initial magnetisation parameter β but rather on the ratio of the kinetic and electromagnetic energy fluxes. This conclusion is supported by the analysis of strong relativistic MHD shocks presented in (Komissarov, 1999b).

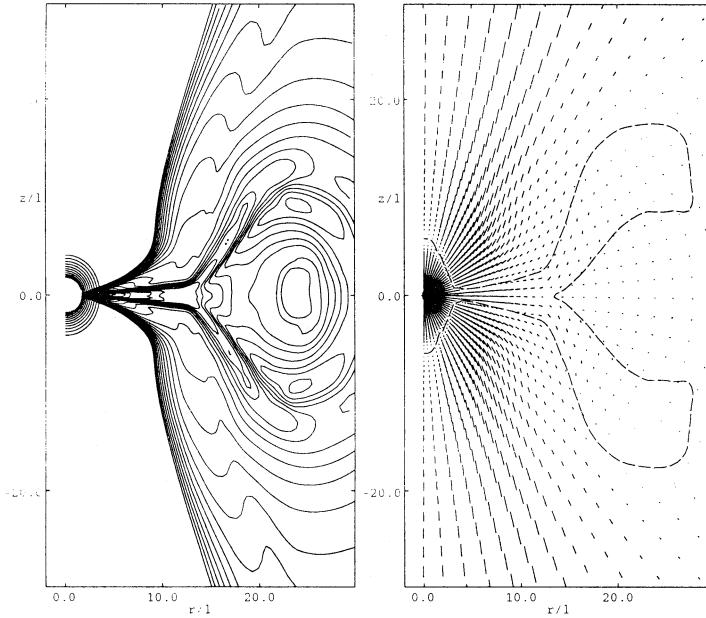


Figure 3. Magnetically driven accretion onto a rapidly rotating black hole ($a = 0.9$). Left panel: rest mass density field. Right panel: velocity field. The dashed line shows the surfaces of vanishing radial component of velocity. In contrast to figure 2 the flow near the z -axis is directed outwards everywhere except the very vicinity of the black hole.

3.2. MAGNETICALLY DRIVEN ACCRETION ONTO AND OUTFLOWS FROM BLACK HOLES

The relativistic jets of active galaxies are believed to be produced in the magnetospheres of black holes hidden in the centres of galactic nuclei. The fully general relativistic numerical study of this problem has just begun. Koide et al.(1999) have studied the dynamics of magnetised thin disks accreting onto a non-rotating black hole. In their simulations a relatively fast collimated outflow develops from the inner region of the disk.

In order to carry out the simulations of this sort we have modified our special relativistic MHD code (Komissarov, 1999a) in the way explored in (Pons et al., 1998; Ibáñez et al., this volume) for gas dynamics. In the process of testing it, we tried to reproduce the jet simulations of (Koide et al., 1999) and obtained rather different results. For example, we found no outflow in the case of a free falling corona both for similar and much higher a resolution. Moreover, we obtained different numerical solution to the problem of sub-Keplerian unmagnetised disk. This rather simple problem can be studied analytically and such analysis shows that the inner parts of the disk must fall onto the black hole without bounce that agrees with our

results and contradicts to the results of (Koide et al., 1999).

Here we are presenting the results of our simulations for the case of a thick accretion disc and a rotating black hole. In these axisymmetric 2D simulations we employ the Boyer-Lindquist coordinates $\{t, \phi, r, \Theta\}$. The computational grid is uniform only in the Θ -direction. The cell size in the r -direction gradually decreases towards the inner boundary in such a way that in both directions the physical cell size is approximately the same. The length scale is $l = r_s/2 = GM/c^2$ and the time scale is $\tau = l/c$. The inner outflow boundary is placed at $r = 1.1r_+$, where $r_+/l = 1 + \sqrt{1 - a^2}$ is the outer event horizon and a is the specific angular momentum of the black hole. The outer free flow boundary is placed at $r \approx 150l$.

The initial solution includes i) a marginally bound polytropic ($\gamma = 4/3$) fat disk (Abramowicz et al., 1978), ii) low density external gas at rest relative to fiducial observers (FIDOs) and iii) magnetic field given by a linear combinations of the divergence free solutions of the form

$$\begin{aligned} B^r &= B_0 r^\alpha \sin \Theta \cos \Theta, \\ B^\Theta &= -\frac{1}{2}\alpha B_0 r^{\alpha-1} \sin^2 \Theta, \\ B^\phi &= 0. \end{aligned}$$

This field is uniform at infinity. The interior of the disk is dominated by gas pressure (β up to 100) and its exterior is magnetically dominated.

Figure 2 shows the numerical solution at $t = 800\tau$ for the case of a non-rotating black hole. The equatorial region of the disk has transferred a fraction of its angular momentum to the outer regions via Maxwell stresses and developed a rather nonuniform inflow. The outflow generated in the outer regions of the disk reaches a speed of $0.4c$ (relative to the local DIDO) near the disc-external gas interface where the flow is magnetically dominated. In the funnel around the symmetry axis the external gas is accreting onto the black hole with the inflow velocity reaching $0.9c$ at the inner boundary.

Figure 3 shows the numerical solution at the similar time for the case of a rapidly rotating black hole ($a = 0.9$) with a corotating fat disk. One can see that at this stage the accretion proceeds only via the narrow cusp-like region near the equatorial plane while the rest of the disk is pushed out by the increased magnetic pressure in the black hole magnetosphere. The outflow is launched not only in the outer regions of the disk but also inside the funnel. In fact, the maximum outflow velocity, $\approx 0.8c$, is reached in the funnel near its interface with the disc. Here the plasma is highly magnetised, with $\beta \approx 2 \times 10^{-3}$ and, thus, the outflow is certainly magnetically driven. The physical mechanism driving the funnel outflow is most likely to be related to the one proposed in (Blandford and Znajek, 1977) where the outflow is powered by the rotational energy of the black hole. Further study is however needed to verify this conclu-

sion. In these simulations the Poynting flux at the inner boundary is, as required, directed towards the hole. However, the flow through the boundary is neither superfast nor superalfvenic and this makes the use of the outflow boundary conditions rather unjustified. Our attempts to place the inner boundary closer to the black hole horizon were unsuccessful. The reason is most likely related to the fact that in this region the flow is highly magnetically dominated and the matter contribution to the total energy-momentum density gradually drops below the level of computational errors. This suggests that we should abandon the energy-momentum conservation and use a numerical scheme where the magnetic field enters the dynamics in the form of the Lorentz force. Moreover, the high computational cost of simulations with the inner boundary placed very close to the horizon and the fact that one does not know beforehand how close to the horizon this boundary should be to insure a superfast inflow velocity suggests that we should abandon the Boyer-Lindquist foliation and use so-called horizon adapted coordinates as proposed in (Papadopoulos and Font, 1998; Font et al., this volume). This will allow us to put the inner boundary inside the horizon.

References

- Abramowicz M, Jaroszyński M and Sikora M (1978). *A&A* **63**, pp 221-224.
 Anile A M (1989). Relativistic Fluids and Magneto-Fluids. Cambridge University Press.
 Blandford R D and Znajek R L (1977). *MNRAS* **179**, pp 433.
 Brackbill J U and Barnes D C (1980). *J. Comp. Phys.* **35**, pp 426.
 Clarke D E, Norman M L and Burns J O (1986). *ApJ.* **311**, pp L63-L67.
 Evans C R and Hawley J F (1988). *ApJ.* **332**, pp 659-677.
 Falle S A E G and Komissarov S S (1996). *MNRAS* **278**, pp 586-602.
 Font J A, Ibáñez J M and Papadopoulos P, this volume.
 Ibáñez J M, Aloy M A, Font J A, Martí J M, MirallesJA and PonsJA, this volume.
 Koide S, Shibata K and Kudoh T (1999). *ApJ.* **522**, pp 727-752.
 Komissarov S S (1999a). *MNRAS* **303**, pp 343-366.
 Komissarov S S (1999b). *MNRAS* **308**, pp 1069-1076.
 Kössl D, Müller E and Hillebrandt W (1990). *A&A* **229**, pp 378-396.
 Lind K R, Payne D G, Meier D L and Blandford R D (1989). *ApJ.* **344**, pp 89-103.
 Papadopoulos P and Font A (1998). *Phys. Rev.D* **58(2)**/024005, pp 1-10.
 Pons J A, Font J A, Ibáñez J M, Martí J M, and Miralles J A (1998). *A&A* **339**, pp 638-642.
 Roe P L (1986). *Ann. Rev. Fluid Mech.* **18**, pp 337-365.
 van Putten M H P M (1996). *ApJ.* **467**, pp L57-L60.

GODUNOV TYPE METHODS ON UNSTRUCTURED GRIDS AND LOCAL MESH REFINEMENT

D. KRÖNER AND T. GESSNER

*Institut of Applied Mathematics, University of Freiburg,
Hermann–Herder–Str. 10, D–79104 Freiburg i. Br. , Germany
Email: dietmar|gessner@mathematik.uni-freiburg.de*

Abstract. One of the most effective method for improving the efficiency of numerical codes is the local mesh adaption according to the local accuracy of the numerical solution. Unfortunately for real applications (reactive Euler equations, MHD equations, compressible Navier Stokes equations with large Reynolds–number) based on the system of the compressible Euler equations there are no theoretical results which are able to control the adaptive process rigorously. Therefore we try to get these results at least for scalar conservation laws and to use them for the design of error indicators for the more complex systems. But before proving a–posteriori error estimates we have to know how to prove a–priori error estimates. Therefore in the first part of this paper we will present an overview of known results concerning convergence and a–priori error estimates for Godunov or Godunov type schemes. Then in the second part we will show some recent results for a–posteriori error estimates also for convection–diffusion–reaction equations. Finally in the third part, we will show how the grid indicators have to be constructed for the simulation of supersonic reacting gas flow.

1. Introduction

We will consider numerical schemes in conservation form in particular Godunov type schemes in a first order form or also as a MUSCL type higher order scheme. First, they have been developed for problems in 1–D. By dimensional splitting they can also be used in 2–D and 3–D on Cartesian grids. It turns out that they can be easily generalized also to unstructured grids in multi dimensions \mathbb{R}^d consisting of very general cells. The classical

explicite Godunov scheme (Godunov, 1959) for scalar conservation laws in 1-D

$$\begin{aligned} \partial_t u + \partial_x f(u) &= 0 && \text{in } \mathbb{R} \times \mathbb{R}^+, \\ u(., 0) &= u_0 && \text{on } \mathbb{R} \end{aligned} \quad (1)$$

can be written in the form

$$u_j^{n+1} := u_j^n - \frac{\Delta t}{\Delta x} \left(g(u_j^n, u_{j+1}^n) - g(u_{j-1}^n, u_j^n) \right)$$

where $x_j := j\Delta x$, $t_n := n\Delta t$ for a uniform grid, g is a Lipschitz continuous, consistent and monotone numerical flux, u_j^n will denote the approximation of the exact solution u on $]x_j - \frac{\Delta x}{2}, x_j + \frac{\Delta x}{2}[\times]t_n, t_{n+1}[$. The classical Godunov scheme can be written in this form for a suitable numerical flux g , satisfying the conditions above. Let

$$u_h(x, t) := u_j^n \quad \text{if } (x, t) \in]x_j - \frac{\Delta x}{2}, x_j + \frac{\Delta x}{2}[\times]t_n, t_{n+1}[.$$

Convergence of u_h to the unique entropy solution of (1) and error estimates of the form

$$\|u_h(., t) - u(., t)\|_{L^1(\mathbb{R})} \leq ch^{\frac{1}{2}}$$

are shown in (Kuznetsov, 1976). Generalizations by dimensional splitting on Cartesian grids in \mathbb{R}^d and corresponding error estimates with $h^{\frac{1}{2}}$ are also shown in (Kuznetsov, 1976).

Now let us discuss further generalizations to unstructured grids. Let $\mathcal{T} = \{T_j \mid j \in I \subset \mathbb{N}\}$ be a mesh of \mathbb{R}^d such that the interface of two neighboring cells T_j , T_l of \mathcal{T} is included in a hyperplane. Let T_l , $l \in N_j$ be the neighboring cells of T_j , $|T_j|$ the volume of T_j and S_{jl} be the joint edge of T_j and T_l . Additionally ν_{jl} denotes the scaled normal of the joint edge S_{jl} with $|S_{jl}| = |\nu_{jl}|$. We assume that there exists an $\alpha > 0$ such we have for $h := \max_j \operatorname{diam} T_j$

$$\begin{aligned} \alpha h^d &\leq \operatorname{meas}(T_j) \\ \alpha \operatorname{meas}(S_{jl}) &\leq h^{d-1} \end{aligned}$$

for all $j, l \in I$. For any $j, l \in I$ there is a numerical flux $g_{jl} : \mathbb{R}^2 \rightarrow \mathbb{R}$ which satisfies the following conditions for all $u, v, u', v' \in [A, B]$.

The numerical flux $g_{jl}(u, v)$ is monotone increasing
with respect to u and monotone decreasing with respect to v . (2)

Furthermore

$$g_{jl}(u, v) = -g_{lj}(v, u) \quad (3)$$

$$g_{jl}(u, u) = \nu_{jl} f(u) \quad (4)$$

$$|g_{jl}(u, v) - g_{jl}(u', v')| \leq L(A, B) h_{jl} (|u - u'| + |v - v'|) \quad (5)$$

where

$$h_{jl} := \max\{\text{diam } T_j, \text{diam } T_l\}.$$

Then a Godunov type scheme for solving

$$\begin{aligned} \partial_t u + \operatorname{div} f(u) &= 0 && \text{in } \mathbb{R}^d \times \mathbb{R}^+, \\ u(., 0) &= u_0 && \text{on } \mathbb{R}^d \end{aligned} \quad (6)$$

can be written in the following form.

DEFINITION (Finite volume scheme). Let

$$\begin{aligned} u_j^0 &:= \frac{1}{|T_j|} \int_{T_j} u_0 \\ u_j^{n+1} &:= u_j^n - \frac{\Delta t}{|T_j|} \sum_{l \in N_j} g_{jl}(u_j^n, u_l^n) \end{aligned} \quad (7)$$

for all $n \in \mathbb{N}$ and $j \in I$.

For the time step we assume the following CFL-condition

$$\Delta t \leq \frac{\alpha^2 h}{2L} \quad (8)$$

where L is the Lipschitz constant from (5). u_j^n is supposed to become an approximation of the exact solution u of (6) on $T_j \times [t_n, t_{n+1}[$. For $g_{jl}(u, v)$ we can use all exact and approximate Riemann solver, e.g. (Engquist and Osher, 1981)

$$g_{jl}(u, v) = c_{jl}(0) + \int_0^u (c'_{jl}(s))^+ ds - \int_0^v (c'_{jl}(s))^- ds \quad (9)$$

where $c_{jl}(s) := \nu_{jl} f(s)$. Similar as before let

$$u_h := u_j^n \text{ if } (x, t) \in T_j \times]t_n, t_{n+1}[.$$

The convergence of u_h for $h \rightarrow 0$ to the entropy solution of (6) has been shown in (Kröner and Rokyta, 1994), (Vila, 1994), (Cockburn et. al., 1994), (Cockburn and Gremaud, 1996). The convergence rate can be estimated as

$$\|u_h(., t) - u(., t)\|_{L^1(\mathbb{R}^d)} \leq ch^{\frac{1}{4}} \quad (10)$$

(see (Vila, 1994), (Cockburn et. al., 1999), and (Cockburn and Gremaud, 1996)). This is not optimal since also in this case we expect $h^{\frac{1}{2}}$. The proof for $h^{\frac{1}{2}}$ on unstructured grids is still an open problem.

Convergence as well as an error estimates of the form (10) can be also shown for the staggered Lax–Friedrichs schemes on unstructured grids (Küther, 2000b), (Haasdonk et. al., 2000). In the case of weakly coupled systems convergence to the entropy solution was proved in (Rohde, 1996).

Using MUSCL type reconstructions (Durlofsky et. al., 1992) the scheme (7) can be generalized to a higher order one. Assume that for all T_j there exists

$$L_j(x) := u_j + (x - w_j)s, \quad (11)$$

where w_j is the center of gravity of T_j and s is a reconstructed gradient, which has to satisfy the conditions in (Kröner et. al., 1995). In general the construction of s is nonlinear and leads to additional difficulties. Then the higher order scheme with respect to space is given by

$$u_j^{n+1} := u_j^n - \frac{\Delta t}{|T_j|} \sum_{l \in N_j} g_{jl}(L_j(x_{jl}), L_l(x_{jl})), \quad (12)$$

where x_{jl} is the midpoint of S_{jl} . Also for this scheme we get convergence to the entropy solution of (6) (Kröner et. al., 1995) and an error estimate (Noelle, 1995) with $h^{\frac{1}{4}}$ of the same type as in (10).

These results have been generalized to higher order schemes in space for the more general conservation law

$$\begin{aligned} \partial_t u + \operatorname{div} F(x, t, u) &= 0 && \text{in } \mathbb{R}^d \times \mathbb{R}^+ \\ u(x, 0) &= u_0(x) && \text{in } \mathbb{R}^d \end{aligned} \quad (13)$$

in (Chainais–Hillairet, 1995) and for higher order schemes in space and time in (Küther, 2000a). Unfortunately, up to now there are no results concerning a-priori error estimates with h^α , $\alpha > \frac{1}{4}$ for (1) in the case of higher order schemes of the form (12). There are some results for the streamline diffusion shock capturing method applied to the linear transport equation with $\alpha = \frac{3}{2}$, cf. (Johnson and Szepessy, 1987), and for the streamline diffusion finite element method applied to the quasi linear convection–diffusion equation with $\alpha = \frac{3}{2}$, cf. (Lube, 1992). Corresponding results for the discontinuous Galerkin method can be found in (Cockburn et. al., 1999) and the references therein. In an improved version of (Kröner and Rokyta, 1994), we have applied the convective part of the first and the higher order schemes (12) to the discretization of the boundary value problem for the convection–diffusion equation

$$\begin{aligned} Lv := -\varepsilon \Delta v + \operatorname{div}(bv) + cv &= f && \text{in } \Omega, \\ v &= 0 && \text{on } \partial\Omega \end{aligned}$$

on an regular grid consisting of equal sided triangles on an polygonal domain Ω . For this we could prove that the error in the ε -weighted H^1 -norm can be estimated for the first and for the second order scheme by

$$\sqrt{h}||u||_{H^2}, \quad h^{\frac{3}{2}}||u||_{H^2} \quad (14)$$

respectively. Although the PDE is linear, the reconstruction process, which is used for the definition of $L_j(x)$ (see (11)), is nonlinear and makes the proof of (14) nontrivial.

2. A-posteriori Error Estimates

The a-priori error estimates which we have presented in the preceding section are the basis for proving a-posteriori error estimates for upwind finite volume discretization of nonlinear conservation laws on unstructured grids in multi dimensions. In general a-posteriori error estimates are extremely useful for accelerating comprehensive computations in multi dimensions in particular in 3-D. They are used to adapt the grid size locally in order to minimize the computing time. The local grid size should be chosen such that the error $||u - u_h||$ between the exact solution u and the numerical solution u_h is less than a given tolerance and such that the total numerical costs are as small as possible. This can be obtained if an a-posteriori error estimate of the form

$$||u - u_h|| \leq c \sum_j \eta_j(u_h) + R_h \quad (15)$$

is given where $\eta_j(u_h)$ are local quantities related to each cell or edge such that $\eta_j(u_h)$ can be computed if u_h is known and where R_h is related to the approximation of the data. Also the constant c should be known.

For elliptic and parabolic problems the theory of a-posteriori error estimates of the form (15) is well developed (Eriksson et. al., 1994), (Dörfler, 1998). But for initial value problems for nonlinear conservation laws of the form (13) only just recently first results have been shown (Cockburn and Gau, 1999), (Katsoulakis et. al., 1999), (Kröner and Ohlberger, 2000).

Unfortunately, there are no rigorous error estimators for systems of conservation laws, which we need in many applications. Instead of them, error indicators, shock indicators, or grid indicators have been used in order to find those regions with steep gradients. Usually, these indicators are based on discrete gradients or higher order discrete derivatives of the discrete solution. They are used to control the local process of refining and coarsening the grid. But these indicators give no information about the true error $||u - u_h||$. Nevertheless, the results for scalar equations are useful for the construction of the error indicators for systems. For instance in case

of source terms the indicators mentioned above are not able to detect the critical regions and have to be modified. The necessary modification can be obtained from the results for scalar equations and will be shown for detonation waves in the following section.

In this section we will present the results from (Kröner and Ohlberger, 2000) and (Ohlberger, 2000), i.e. rigorous a-posteriori error estimate of the form (15) in the L^1 -norm for the initial value problem for the conservation law (13) and for the convection-diffusion-reaction problem (see (Ohlberger, 2000)). Former results related to this topic are published in (Tadmor, 1991) (Cockburn and Gau, 1999) for the nonlinear case and in (Houston et. al., 1999) for linear systems.

In addition to the assumption of the previous section we have to assume the following conditions for the data:

$$u_0 \in L^\infty(\mathbb{R}^d) \cap BV_{loc}(\mathbb{R}^d) \quad (16)$$

with constants A and B such that $A \leq u_0 \leq B$ a.e.

and

$$F \in C^1(\mathbb{R}^d \times \mathbb{R}^+ \times \mathbb{R}, \mathbb{R}^d), \quad (17)$$

$$\sum_j \frac{\partial F_j}{\partial x_j}(x, t, s) = 0 \quad \text{for all } (x, t, s) \in \mathbb{R}^d \times \mathbb{R}^+ \times \mathbb{R}. \quad (18)$$

For all compact sets $K \subset \mathbb{R}$ there exists a constant $c_0(K)$ such that

$$|\partial_s F(x, t, s)| \leq c_0(K) \quad (19)$$

$$|\partial_x \partial_s^\beta F(x, t, s)| + |\partial_t \partial_s^\beta F(x, t, s)| \leq c_0(K) \quad (20)$$

for almost all $(x, t, s) \in \mathbb{R}^d \times \mathbb{R}^+ \times K$ and $\beta = 0, 1$.

$$g_{jl}(u, u) = \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} \int_{S_{jl}} F(x, t, u) n_{jl} dx dt \quad (21)$$

where Δt is the time step and n_{jl} is the outer unit normal to S_{jl} .

Now it turns out that the same tools which have been used for proving a-priori error estimates for (13) in (Chainais-Hillairet, 1995) can be used to show an a-posteriori estimate. Let $R, \omega = c_0(K)$ (see (19)), T be given and

$$\begin{aligned} I_0 &:= \left\{ n \mid 0 \leq t^n \leq \min \left\{ \frac{R+1}{\omega}, T \right\} \right\} \\ A &:= \{(x, t) \mid |x - x_0| + \omega t < R + 1\} \\ M(t) &:= \{j \mid \text{there exists } x \in T_j \text{ such that } (x, t) \in A\}. \end{aligned} \quad (22)$$

THEOREM. Assume (2), ..., (5), (8), and (16), ..., (21). Let $K \subset \mathbb{R}^d \times \mathbb{R}^+$, $\omega = c_0(K)$ (see (19)) and choose T, R and x_0 such that $T \in]0, \frac{R}{\omega}[$ and

$$K \subset \cup_{0 \leq t \leq T} B_{R-\omega t}(x_0) \times \{t\}.$$

Let u be the entropy solution of (13) and u_h be the numerical solution of the first order finite volume scheme as described in (7). Then we have

$$\int_K |u - u_h| \leq CT \int_{|x-x_0| < R+1} |u_0(x) - u_h(x, 0)| dx + R + \sqrt{R}$$

where

$$\begin{aligned} R := & C \sum_{n \in I} \sum_{j \in M(t^n)} |u_j^{n+1} - u_j^n| \Delta t h_j^2 \\ & + 2C \Delta t \sum_n \sum_{edges} (\Delta t + h_{jl}) \max_{u_l^n \leq v \leq w \leq u_j^n} (g_{jl}(w, v) - g_{jl}(w, w)) \delta_{jl}^n \\ & + 2C \Delta t \sum_n \sum_{edges} (\Delta t + h_{jl}) \max_{u_l^n \leq v \leq w \leq u_j^n} (g_{jl}(w, v) - g_{jl}(v, v)) \delta_{jl}^n \\ & + 2CTM_1 \sum_n \sum_{edges} h_{jl} (\Delta t + h_{jl})^2 \delta_{jl}^n \end{aligned}$$

and

$$\begin{aligned} \delta_{jl}^n &:= 0 \quad \text{if } T_j \times [t^n, t^{n+1}] \cap A = \emptyset \quad \text{and} \quad T_l \times [t^n, t^{n+1}] \cap A = \emptyset \\ \delta_{jl}^n &:= 1 \quad \text{otherwise.} \end{aligned}$$

Here g_{jl} corresponds to the edge jl .

COROLLARY. Under the assumption of the previous theorem and if $F(x, t, v) = F(v)$ we have

$$\int_K |u - u_h| \leq TC \int_{|x-x_0| < R+1} |u_0(x) - u_h(x, 0)| dx + Q + \sqrt{Q}$$

where

$$Q := \sum_{n \in I} \sum_{j \in M(t^n)} \Delta t h_j^2 |u_j^{n+1} - u_j^n| + 2L \Delta t \sum_n \sum_{E(t_n)} (\Delta t + h_{jl}) h_{jl} |u_j^n - u_l^n|$$

and $E(t_n)$ is the set of all edges, which lay in $M(t^n)$ and u_j, u_l are the values on both sides of the edge jl .

In (Ohlberger, 2000) this result has been generalized to the convection-diffusion-reaction case.

$$\begin{aligned} \partial_t c + \operatorname{div}(uf(c) - D\nabla c) + \lambda c &= 0 & \text{in } \mathbb{R}^d \times]0, T[\\ c(., 0) &= c_0 & \text{in } \mathbb{R}^d \end{aligned} \tag{23}$$

where $u : \mathbb{R}^d \times \mathbb{R}^+ \rightarrow \mathbb{R}^d$ denotes a given velocity, f as before, $D = const \geq 0$ denotes the diffusion parameter and λ the reaction coefficient. Notice that $D(c)$ is allowed to degenerate and that the following a-posteriori error estimate will hold uniformly with respect to the lower bound of $D(c)$. The basic idea of the proof is based on the Kuznetzov technique (Kuznetsov, 1976), similar as for the a-priori estimates, and avoids energy methods. If energy methods are used to prove a-priori or a-posteriori error estimates for (23), the constants or the norms in the estimates will in general strongly depend on the lower bound of $D(c)$. For many applications (flow through porous media, flows with large Reynolds-number) it is necessary to control the case of small diffusion or degenerating diffusion and to avoid large constants in the error estimates.

In order to treat this case the concept of entropy solution has to be generalized to second order problems (Carrillo, 1999), (Ohlberger, 2000). For defining the numerical solution of (23) we follow the lines in (Ohlberger, 2000) and use a triangulation \mathcal{T} , satisfying the conditions of the previous sections. The nodes of this triangulation will be denoted by p_j . Then we define a dual mesh consisting of cells Ω_j around the vertex p_j by connecting the centers of gravity of the cells T_l , $l = 1, \dots, m_j$, surrounding p_j , with the centers of gravity of the neighboring edges $\Gamma_{j,l}$, connecting p_j with p_l . The joint edges of Ω_j and Ω_l consists of two parts and will be denoted by S_{jl}^+ and S_{jl}^- and the adjacent triangles to $\Gamma_{j,l}$ by T_{jl}^+ and T_{jl}^- . In addition to the condition of the previous sections we have to assume that there is no triangle T_l with an angle greater than $\pi/2$.

On each S_{jl}^+ and S_{jl}^- we have to define convective numerical fluxes g_{jl}^{n+} and g_{jl}^{n-} which have to satisfy the conditions (2), (3), (5) and instead of (4)

$$g_{jl}^{n+}(w, w) = \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} \int_{S_{jl}^+} u(x, t) n_{jl}^+ dx dt f(w)$$

where n_{jl}^+ denotes the outer unit normal to S_{jl}^+ with respect to Ω_j . The corresponding conditions should also hold for $g_{jl}^{n-}(w, w)$.

The viscous fluxes are defined by

$$d_{jl}^{n+}(w_h) := D_{jl}^{n+}(w_l - w_j) \quad \text{with} \quad D_{jl}^{n+} := |S_{jl}^+| \frac{D}{2} (\nabla N_j - \nabla N_l) n_{jl}^+$$

where N_j , N_l are the shape functions of the finite element space of the piecewise linear functions with respect to the endpoints p_j and p_l of Γ_{jl} in the triangle T_{jl}^+ . Analogously we define d_{jl}^{n-} . Now the numerical scheme can be defined as follows. For all $n \in \mathbb{N}$ and all $j \in I$ let

$$c_j^0 := \frac{1}{|\Omega_j|} \int_{\Omega_j} c_0$$

$$\begin{aligned}
c_j^{n+1} := & \quad c_j^n \\
& - \frac{\Delta t}{|\Omega_j|} \sum_{l \in N(j)} \sum_{* \in \{+, -\}} \left(g_{jl}^{n+1,*}(c_j^{n+1}, c_l^{n+1}) - d_{jl}^{n+1,*}(c_h^{n+1}) \right) \\
& + \Delta t \lambda_j^{n+1}
\end{aligned}$$

and λ_j^n denotes the mean value of λ on $\Omega_j \times]t^n, t^{n+1}[$. Define c_h^n now as the piecewise linear continuous function such that $c_h^n(p_j) = c_j^n$ for all $n \in \mathbb{N}$ and all $j \in I$ and

$$c_h(\cdot, t) := c_h^{n+1} \quad \text{for } t \in]t^n, t^{n+1}[. \quad (24)$$

THEOREM (A-posteriori estimate for degenerate convection-diffusion-reaction problems (Ohlberger, 2000)). Assume that the conditions mentioned above, $D(c) = D$, and the stability condition

$$\min_{j,n} (1 + \Delta t \lambda_j^{n+1}) \geq \gamma > 0$$

are satisfied and that the data are sufficiently smooth and $u = \text{const}$, $\lambda = \text{const}$, $d = 2$. Let c denote the exact solution of (23) and c_h the numerical solution as defined in (24). Then we have uniformly in D

$$\begin{aligned}
\|c - c_h\|_{L^1([0,T] \times \mathbb{R}^2)} \leq & \quad c \left(\int_{\mathbb{R}^2} |c_h(x, 0) - c_0(x)| dx \right. \\
& + \sum_{n,j} \|\lambda\|_{L^\infty} h \Delta t \int_{\Omega_j} |\nabla c_h(x, t^{n+1})| dx \\
& \left. + \sqrt{Q_0} + Q_1 + \sqrt{Q_1} \right)
\end{aligned} \quad (25)$$

where

$$\begin{aligned}
Q_0 : = & \quad \sum_{n,j} \|\lambda\|_{L^\infty} |c_j^{n+1}| h \Delta t \int_{\Omega_j} |\nabla c_h(x, t^{n+1})| dx \\
& + \sum_n \sum_{jl \in \text{edges}} D[\nabla c_h^{n+1} n_{jl}]_{\Gamma_{jl}} |c_j^{n+1} - c_l^{n+1}| \Delta t |\Gamma_{jl}| \\
Q_1 : = & \quad \sum_{n,j} |c_j^{n+1} - c_j^n| \Delta t |\Omega_j| \\
& + \sum_n \sum_{jl \in \text{edges}} (h + \Delta t) \Delta t h |c_j^{n+1} - c_l^{n+1}| \\
& + \sum_n \sum_{jl \in \text{edges}} D[\nabla c_h^{n+1} n_{jl}]_{\Gamma_{jl}} \Delta t |\Gamma_{jl}| (h + \Delta t) \\
& + \sum_{n,j} (1 + \|u\|_{L^\infty}) h \Delta t \int_{\Omega_j} |\nabla c_h(x, t^{n+1}) \Gamma_{jl}| dx.
\end{aligned}$$

Here $[\nabla c_h^{n+1} n_{jl}]_{\Gamma_{jl}}$ denotes the jump of $\nabla c_h^{n+1} n_{jl}$ across Γ_{jl} .

These estimates indicate that the source term λc as well as the velocity u has to be included into the error estimates, in particular if they are large. In the following section we will see that it is necessary to take into account these quantities for the simulation of detonation waves and detonation patterns.

3. Application: Grid Indicators for Supersonic Reactive Flow

In this section we will describe some results for the simulation of detonation cells (Geßner, 2000), which appear for instance in the well known experiment of Strehlow (Strehlow, 1969). In this experiment we consider a gas mixture in two different states separated by a membrane. The mixture consists of oxygen, hydrogen and argon at different physical states. Initially the gas is at rest and the pressure on the left side is much larger than on the right side. If the temperature on the left side is large enough the mixture will ignite, the hydrogen–oxygen detonation will start and an unstable reaction front moves to the right. A Schlieren picture of the experiment is shown in Figure 1 (right part) and exhibits the detonation cell pattern. The model of this numerical simulation includes a detailed reaction mechanism which consists of 48 elementary reactions of 9 species.

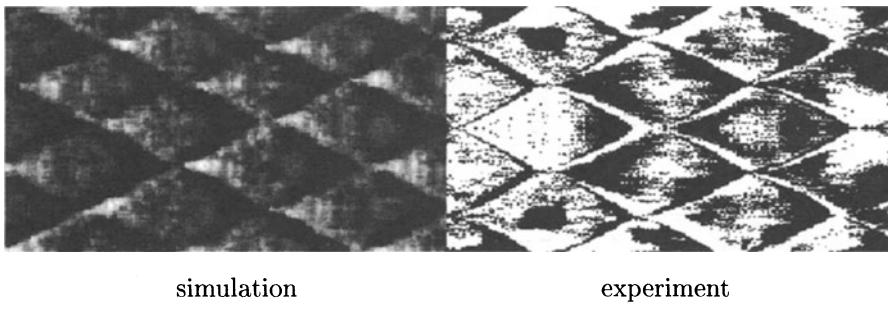


Figure 1. Regular detonation cells: Simulation – Experiment of Strehlow

3.1. THE MATHEMATICAL MODEL

The mathematical model for this problem is based on the compressible Euler equations of gas dynamics in two space dimensions describing the inviscid ideal gas flow. Additionally we have the conservation laws with source terms for an arbitrary number of different species. For details concerning the model beyond the following brief description, we refer e.g. to (Williams, 1985).

The basic quantities in the system of equations are the density ρ , the velocities $v_{1,2}$ in x - and y direction, the specific total energy E , the pressure p , and the mass fractions z_i of the different gas components $i \in \{1, 2, \dots, N\}$. Then the complete system of conservation laws can be written in the following form

$$\partial_t u + \partial_x f_1(u) + \partial_y f_2(u) = s(u), \quad (26)$$

where

$$u = (\rho, \rho v_1, \rho v_2, \rho E, \rho z_1, \dots, \rho z_N)^t \quad (27)$$

is the vector of the conservative variables. The nonlinear fluxes are

$$f_1(u) = \begin{pmatrix} \rho v_1 \\ \rho v_1^2 + p \\ \rho v_1 v_2 \\ (\rho E + p)v_1 \\ \rho v_1 z_1 \\ \vdots \\ \rho v_1 z_N \end{pmatrix}, \quad f_2(u) = \begin{pmatrix} \rho v_2 \\ \rho v_1 v_2 \\ \rho v_2^2 + p \\ (\rho E + p)v_2 \\ \rho v_2 z_1 \\ \vdots \\ \rho v_2 z_N \end{pmatrix} \quad (28)$$

and the right hand side (RHS) is

$$s(u) = (0, 0, 0, 0, \dot{m}_1, \dots, \dot{m}_N)^t, \quad (29)$$

the source term including the chemical production rates \dot{m}_i of the different gas components. The chemical production rate of species i for an arbitrary number R of chemical reactions is

$$\dot{m}_i = M_i \sum_{k=1}^R (\nu''_{ik} - \nu'_{ik}) \omega_k, \quad (30)$$

where M_i is the molecular weight of species i , ω_k is the reaction rate of elementary reaction $k \in \{1, 2, \dots, R\}$, and ν'_{ik} , ν''_{ik} are the stoichiometric coefficients. The reaction rates ω_k are expressed as follows

$$\omega_k := k_k \prod_{i=1}^N \left(\frac{\rho z_i}{M_i} \right)^{\nu'_{ik}} \quad \text{with} \quad k_k(T) := A_k T^{\beta_k} \exp \left(\frac{-E_{act_k}}{\mathcal{R}T} \right).$$

The rate coefficients $k_k(T)$ are modeled by a modified Arrhenius law, where A_k is the frequency factor, β_k is the pre-exponential temperature exponent, E_{act_k} is the activation or threshold energy of reaction k , and \mathcal{R} is the universal gas constant. The heats of reaction are modeled as follows: The specific enthalpy h for an ideal-gas mixture is

$$h := \sum_{i=1}^N z_i h_i(T) \quad \text{where} \quad h_i(T) := h_i^0 + \int_{T^0}^T c_{pi}(T') dT'$$

is the partial specific enthalpy of species i which depends only on T . Thereby c_{pi} is the specific heat capacity at constant pressure, and h_i^0 is the reference enthalpy at the reference temperature of species i . To close the system the equation of state (EOS) for a mixture of thermally ideal gases

$$p = \frac{\rho \mathcal{R}T}{\bar{M}}, \quad \text{with } \bar{M} := \left(\sum_{i=1}^N \frac{z_i}{M_i} \right)^{-1},$$

and the caloric EOS

$$p = \rho h - \rho E + \frac{1}{2} \rho |\vec{v}|^2 \quad (31)$$

are necessary. As a simplification of (31), the equation of state for a polytropic ideal gas with the specific heat ratio γ can be used

$$h := c_p T + h^0, \quad c_p = \frac{\gamma}{\gamma - 1} \frac{\mathcal{R}}{\bar{M}}. \quad (32)$$

3.2. THE NUMERICAL METHOD

The main problem for solving (26) with (28), (29) numerically is due to the stiffness of the problem. The stiffness of the reactive flow equations is a problem associated with the short time scales. Beyond this problem associated with the short time scales, there is a further numerical phenomenon associated with the spatial resolution used to solve the system of equations: In the stiff case, spurious unphysical solutions that only occur in numerics, such as discontinuities moving with wrong speed and bifurcations instead of a single reaction front can be observed. This problem of “spurious numerical solutions” is systematically analyzed by LeVeque and Yee in (LeVeque and Yee, 1990) and can be solved by effectively resolving the scales of the stiff source term via methods such as adaptive mesh refinement (cf. Chapter 3.3 for references), front tracking, or subcell resolution. Another approach, pursued in (Berkenbosch et. al., 1998), (Helzel et. al., 1999), is to prevent spurious numerical solutions without resolving the scales of the stiff source term.

The numerical scheme bases on an unstructured conforming triangular discretization \mathcal{T} (with elements T_j) of the considered domain. The convective terms in (26) are discretized using an upwind finite volume scheme (see e.g. (Kröner, 1997)) similar to (7). To cope with the stiffness of the system of equations, different methods to integrate the chemical source terms (30) have been implemented. Among them are the integrator for stiff systems of ordinary differential equations of Kaps and Rentrop (Kaps and Rentrop, 1979) (A-stable of fourth order) and Young and Boris (Young and Boris,

1977) (hybrid second order scheme called SAIM for stiff ODE's associated with chemical kinetics). We restrict ourselves to the description of the explicit time stepping. Furthermore an implicit time stepping can be used (see (Geßner, 2000)). The adaptive step size control of the explicit schemes takes the CFL condition and an a-posteriori error estimate of the source term integration into account. This step size control is combined with a time-scale splitting and a local time stepping for the source term integration. Then the explicit numerical scheme including these features is

$$\begin{aligned} u_j^0 &:= u_0(x_j), \quad x_j \in T_j \in \mathcal{T} \\ u_j^{n,k+1} &:= \mathcal{S}\left(u_j^n, u_j^{n,1}, \dots, u_j^{n,k}, \Delta t_{\text{ST}}^{n,k}, s\right), \quad u_j^{n,0} := u_j^n \\ u_j^{n+1} &:= u_j^{n,K_n} - \frac{\Delta t_{\text{FV}}^n}{|T_j|} \sum_{l=1}^3 g_{jl}(u_j^n, u_{jl}^n), \end{aligned}$$

for any source source term integration method

$$\mathcal{S} : \left(\mathbb{R}^K, \mathbb{R}^{K \cdot K_n}, \mathbb{R}, C^2(\Omega, \mathbb{R}^K)\right) \rightarrow \mathbb{R}^K$$

and numerical flux $g_{jl}(\cdot, \cdot)$. Thereby step sizes for the source term integration $\Delta t_{\text{ST}}^{n,k}$ and the finite volume scheme Δt_{FV}^n have to fulfill the condition

$$\Delta t_{\text{FV}}^n = \sum_{k=1}^{K_n} \Delta t_{\text{ST}}^{n,k}.$$

Concerning the efficiency of this numerical method it is very important to note that the source term integration can be performed in each element T_j of the triangulation \mathcal{T} independently with a local time step. A synchronization is only necessary before calculating the numerical fluxes in the new time step.

3.3. DYNAMIC MESH ADAPTION

Since this time integration has to be performed in each cell we can reduce the whole CPU-time considerably if we minimize the number of cells, i.e. to use time dependent locally adapted grids. The criterion we use to control this dynamic mesh adaption bases on heuristic considerations founded on experiences in non-reactive flow simulations (Geßner, 1994), theoretical results (see (25) of (Ohlberger, 2000)) and numerical experiments. It evaluates three different quantities and takes the space and time derivations of the flow and the chemical production rate into consideration. As quantity controlling the mesh adaption in pure flow simulations we define the

following improved (compared with (Geßner, 1994)) weighted space–time–gradient

DEFINITION (Local weighted flow quantity). Consider the triangle $T_j \in \mathcal{T}$ with the constant vector valued u_j^n at the time t^n . Let v_j^n be a scalar component of the conservative flow variables (see (27)) of u_j^n . The values in the neighbors T_{jl} of T_j ($l \in \{1, 2, 3\}$) and at time t^{n-1} have the corresponding indices. Then

$$\eta_j^n := \left(|T_j| \left(\left(\frac{v_j^n - v_j^{n-1}}{\Delta t^n} \right)^2 + \sum_{l=1}^3 \left(\frac{v_j^n - v_{jl}^n}{w_j - w_{jl}} \right)^2 \right) \right)^{\frac{1}{2}}$$

defines the local weighted flow quantity, where w_j is the center of gravity of T_j . (In boundary elements of \mathcal{T} the sum is calculated using the reduced set of neighboring triangles).

Numerical experiments in (Geßner, 2000) have shown that the indicator η_j^n is not able to control the adaption of the grid for reacting flow problems. It turned out that it is necessary to take the reaction rates into account. Also the rigorous error estimates in (Ohlberger, 2000) demonstrated, that the source term has to be taken into consideration (see (25)). To characterize the local intensity of the chemical reaction we define

DEFINITION (Local reaction quantity). In the triangle $T_j \in \mathcal{T}$ with the constant vector-valued u_j^n at the time t^n

$$\mu_j^n := \max_{i=1 \dots N} \{ \dot{m}_i(u_j^n) \}$$

defines the local reaction quantity, where \dot{m}_i as defined in (30).

Now our mesh adaption criterion is the worst case combination of the two local quantities defined above:

DEFINITION (Local reactive flow mesh adaption quantity). Let

$$\bar{\eta}^n := \frac{1}{\#I} \sum_{j \in I} \eta_j^n, \quad \bar{\mu}^n := \frac{1}{\#I} \sum_{j \in I} \mu_j^n$$

be the averages of η_j^n and μ_j^n on the triangulation \mathcal{T} . Then

$$\zeta_j^n := \max \left(\frac{\eta_j^n}{\bar{\eta}^n}, \frac{\mu_j^n}{\bar{\mu}^n} \right) \quad (33)$$

is the local relative reactive flow mesh adaption quantity on triangle T_j at time t^n .

The flow and the reaction quantities participate relatively to their mean values on \mathcal{T} in their common reactive flow quantity ζ_j^n to simplify the control of the adaptive mesh refinement (AMR) for different problems. In the

following criterion controlling the dynamic mesh adaption the parameters are almost independent of the reactive flow simulation performed.

DEFINITION (Reactive flow mesh adaption criterion). Consider the triangle $T_j \in \mathcal{T}$ at the time t^n with ζ_j^n according to (33). Then we mark T_j for refinement or coarsening as follows:

$$\begin{aligned} \text{if } \zeta_j^n > A_{\text{FINE}} &\rightarrow \text{mark } T_j \text{ for refinement} \\ \text{if } \zeta_j^n < A_{\text{COARSE}} &\rightarrow \text{mark } T_j \text{ for coarsening} \end{aligned}$$

with positive constants $A_{\text{FINE}} > A_{\text{COARSE}}$.

Good choices for $A_{\text{COARSE}}/A_{\text{FINE}}$ are e.g. 0.1/0.5 for a very sensitive mesh refinement up to 0.7/1.3 where only the most intense structures in the numerical solution force a mesh refinement. Altogether the controlling of the dynamic mesh adaption requires some additional parameters to increase the efficiency of the algorithm and to make it more comfortable to handle. For details concerning the comfortable user control of the dynamic mesh adaption we refer to (Geßner, 2000).

We use the conforming mesh adaption algorithm of Bänsch (Bänsch, 1989). His approach bases on local bisection and the inverse coarsening process.

3.4. NUMERICAL RESULTS

The numerical scheme described in the previous chapters is now applied to approximate the solutions of different reactive flow problems modeled analogously Section 3.1. All numerical experiments pursue the object to validate all parts of the numerical schemes and to show the flexibility (and limitations) of the numerical method concerning different problems and setups. We compare the results of the numerical simulations with ZND solutions and determine ignition time, velocity of reaction front, detonation cell size analytically, numerically, and experimentally. Furthermore qualitative comparisons with experimental data are performed. Additionally, we want to determine the best and fastest numerical method by comparison of the respective components, e.g. numerical fluxes, time discretization, or source term integration. The results of these comparisons can be found in (Geßner, 2000).

3.4.1. ZND Detonation Waves

At first, we consider the following simplified problem: The model described in Section 3.1 for the single irreversible reaction $R \rightarrow P$ from reactant to product (corresponding to unburnt and burnt gas) together with the equation of state for a polytropic gas (32). For this setup, the theory of Zeldovich, von Neumann, and Döring (ZND; see e.g. (Williams, 1985)) make an “exact” solution in one space dimension available. An exact solution

with a very small half reaction length (HRL, to classify the length scale of the reaction; see (Bourlioux et. al., 1991)) is chosen. Therefore spurious solutions are very likely in the absence of a fine spatial discretization.

Figure 2 shows the results for different numerical simulations with and without AMR. In each part of Figure 2, the thick line (with the small peak) shows the “exact” ZND solution. Without AMR,

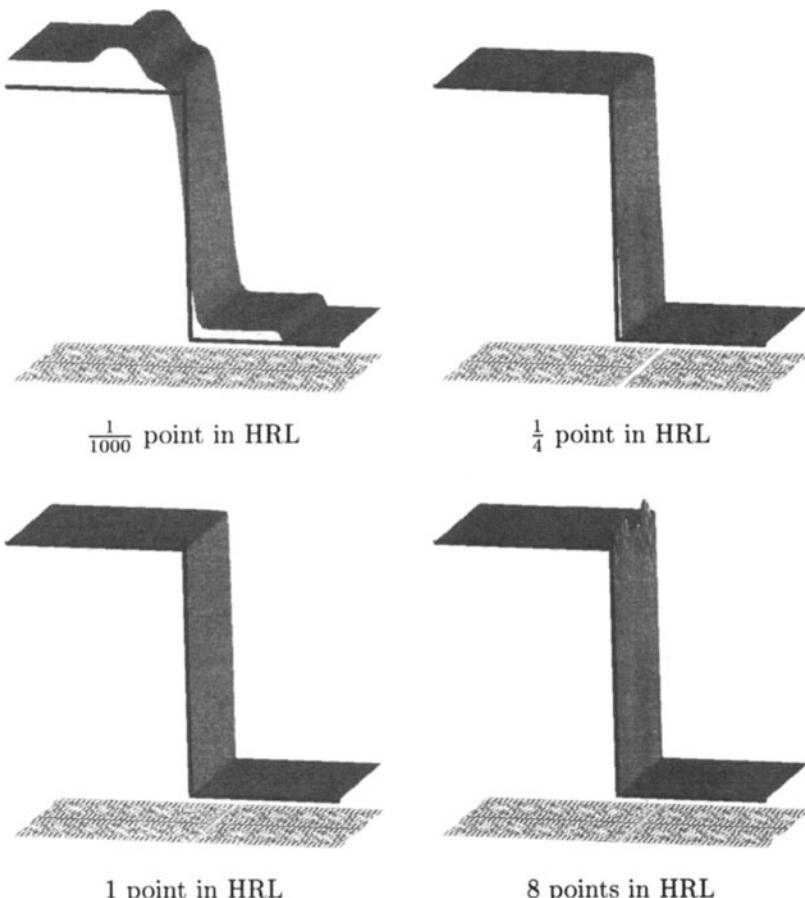


Figure 2. Numerical Simulation of a ZND detonation in 2D

there is only 1/1000 of a point in the HRL. A typical spurious solution traveling with wrong speed instead of a single reaction wave can be observed. With dynamic mesh adaption, the solution is quite right. In this example, 8 points in the HRL are necessary to resolve the reaction zone.

3.4.2. Unstable Detonation Waves in Two Space Dimensions

The unstable behavior of detonation waves in two space dimensions is extensively studied in (Bourlioux and Majda, 1992). Using the same simplified setup as in the previous simulation, we consider an unstable detonation in the shock frame (transformation in Galilean coordinates). Figure 3 compares the results of numerical simulations using dynamic mesh adaption and on a uniformly refined mesh at different times. The behavior of both

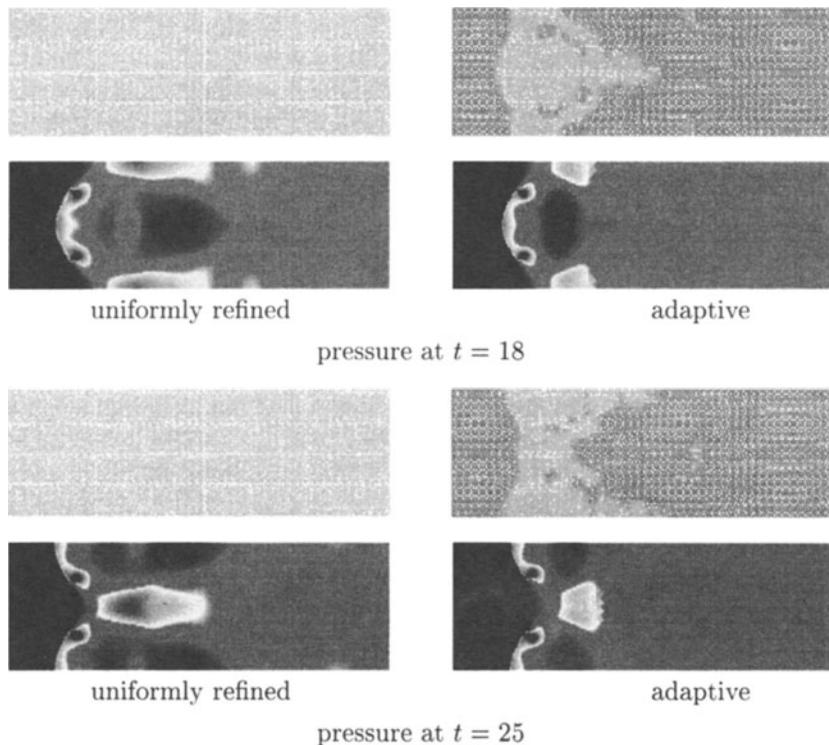


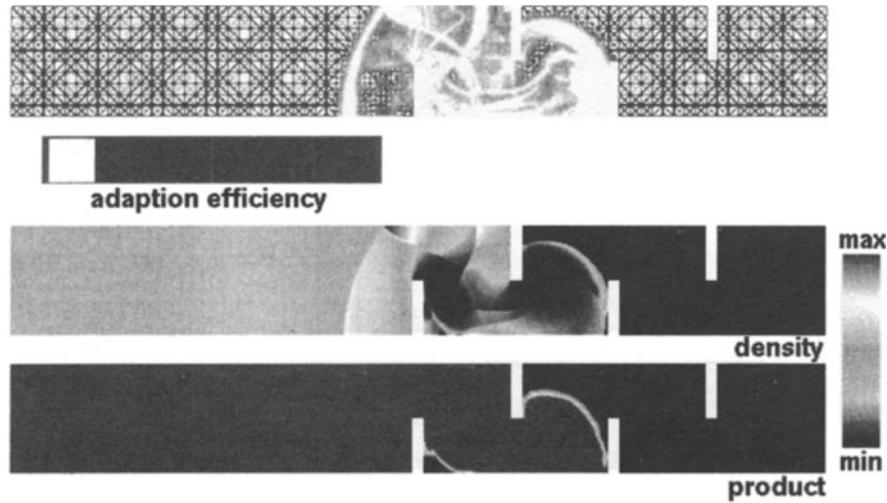
Figure 3. Unstable detonation in 2-D: with/without dynamic mesh adaption

numerical solutions in time is similar. The reaction front computed with AMR is as well resolved as on the uniformly refined mesh with only a fraction of computational cost. Solely the subsonic effects behind the reaction front are better resolved on the uniformly refined mesh.

3.4.3. Detonation in a Channel with Barriers

Everything else being the same as in the first simulation, the detonation wave now meets a cascade of barriers. The resulting complex reactive flow severely challenges the mesh adaptor. Behind the first barrier, the transi-

tion of the detonation to a deflagration can be observed. Figure 4 shows the adaptive mesh, the density, and reactant respectively product at subsequent times. During the whole simulation all structures in the solution are



adaptive mesh, density and mass fraction of the burnt gas
at subsequent times

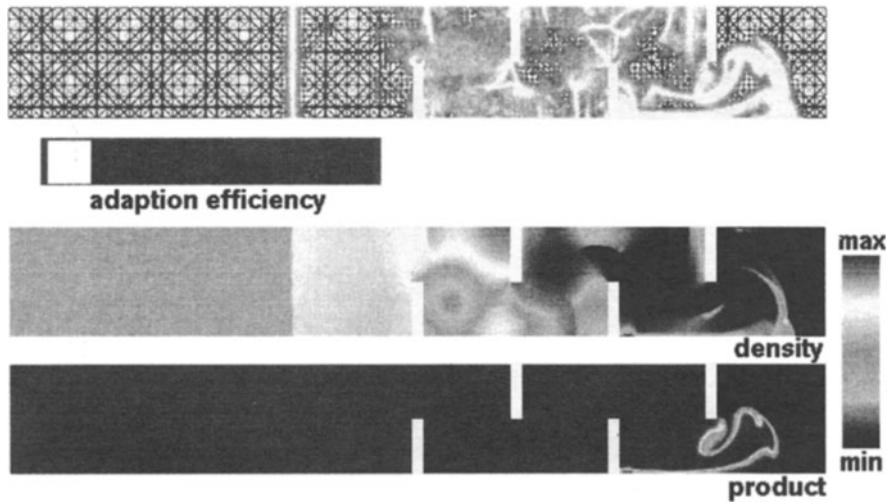


Figure 4. Dynamic mesh adaption in channel with cascade of barriers

resolved. For example the reaction front is captured by the dynamic mesh

adaption as well as the less intense discontinuities in the density. The bar chart (in the middle of each snapshot) compares the numerical cost of the displayed simulation with the cost of a fictitious simulation based on a uniformly refined mesh. In this complex geometry this ratio is less then 15%, which is an indication for the efficiency of the adaptive mesh refinement.

3.4.4. Two Dimensional Unstable Hydrogen–Oxygen Detonation

Now we consider an unstable hydrogen–oxygen detonation with high diluent argon. A detailed reaction mechanism with 9 species and 48 elementary reactions (see (Oran and Boris, 1981)) is used in the model. These low-pressure detonations have been extensively studied (e.g. (Oran et. al., 1998)) because such mixtures are known to produce extremely regular detonation structures and cellular detonation patterns. In Figure 1, we qualitatively compare the computed released chemical energy on the left with the experimental result of Strehlow (Strehlow, 1969) on the right. The size of the detonation cells in the numerical simulation (displayed in Figure 1) is approximately $5.7 \times 3.3\text{ cm}$, and the reaction front is propagating with $1591 \frac{\text{m}}{\text{s}}$. Oran et. al. (Oran et. al., 1998) simulated the same experiment with a totally different numerical method on a powerful parallel computer. Their computed cell size is $5.5 \times 3.0\text{ cm}$ with a reaction front velocity of $1625 \frac{\text{m}}{\text{s}}$. The detonation cell width measured in the corresponding experiment is 5.8 cm . The reason for this larger cell size is probably the energy loss at the walls, which cause the detonation to propagate less rapidly. (see (Oran et. al., 1998)).

The AMR is very efficient for the simulation in Figure 1. Less than 2% of computing time is necessary comparing the displayed simulation with a fictitious one based on a uniformly refined mesh.

3.5. CONCLUSIONS

The dynamic mesh adaption makes the resolution of the disparate physical scales of reactive flow problems possible. Also the other components of the numerical method we did not describe in detail are very important to increase its efficiency. Among these ingredients are the local time stepping, and the integrator for stiff systems of ODE. In summary, we obtain the possibility to simulate challenging reactive flow problems with detailed reaction mechanisms on a workstation with moderate cost.

Acknowledgment. This work has been partially supported by the German Research Association (DFG; Deutsche Forschungsgemeinschaft; Grant DFG Kr 795/5-1).

References

- Bänsch E (1989). Local Mesh Refinement in 2 and 3 Dimensions. *Report SFB256, Universität Bonn* **6**.
- Berkenbosch A C, Kaasschieter E F and Klein R (1998). Detonation Capturing for Stiff Combustion Chemistry. *Combust. Theory Modeling* **2**, pp 313-348.
- Bourlioux A and Majda A (1992). Theoretical and Numerical Structures for Unstable two-dimensional Detonations. *Comb. and Flame* **90**, pp 211-229.
- Bourlioux A, Majda A and Royd burd V (1991). Theoretical and Numerical Structures for Unstable one-dimensional Detonations. *SIAM J. Appl. Math.* **51**, pp 303-343.
- Carrillo J (1999). Entropy Solutions for Nonlinear Degenerate Problems. *Arch. Rational Mech. Anal.* **147**, pp 269-361.
- Chainais-Hillairet C (1995). A Posteriori Error Estimates for General Numerical Methods for Scalar Conservation Laws. *Comput. Appl. Math.* **114**, pp 37-47.
- Cockburn B, Coquel F and LeFloch P (1994). An Error Estimate for Finite Volume Methods for Multidimensional Conservation Laws. *Math. Comp.* **63**, pp 77-103.
- Cockburn B, Karniadakis G E and Shu C W (1999). The Development of Discontinuous Galerkin Methods. *Preprint*.
- Cockburn B and Gau H (1999). A Finite Volume Schemes for a Nonlinear Hyperbolic Equation. Convergence Towards the Entropy Solution and Error Estimates. *Mathematical Modelling and Numerical Analysis* **33**, pp 129-156.
- Cockburn B and Gremaud P A (1996). A Priori Error Estimates for Numerical Methods for Scalar Conservation Laws. Part I: The General Approach. *Math. Comp.* **65**, pp 533-573.
- Dörfler W (1998). Uniformly Convergent Finite-Element Methods for Singularly Perturbed Convection-Diffusion Equations. *Habilitationsschrift, Mathematische Fakultät Freiburg*.
- Durlofsky L J, Engquist B and Osher S (1992). Triangle Based Adaptive Stencils for the Solution of Hyperbolic Conservation Laws. *J. Comput. Phys.* **98**, pp 64-73.
- Eriksson K, Estep D, Hansbo P and Johnson C (1994). Numerical Solution of Partial Differential Equations by the Finite Element Method. Cambridge University Press.
- Engquist B and Osher S (1981). One-Sided Difference Approximations for Nonlinear Conservation Laws. *Math. Comp.* **36**, pp 321-351.
- Geßner T (2000). Zeitabhängige Adaption für Finite Volumen Verfahren höherer Ordnung. *Diplomarbeit, Institut für Angewandte Mathematik, Universität Bonn*.
- Geßner T (2000). Timedependent Adaption for Supersonic Combustion Waves Modeled with Detailed Reaction Mechanisms. *PhD Thesis, Mathematische Fakultät Freiburg*.
- Godunov S K (1959). Finite Difference Method for Numerical Computations of Discontinuous Solutions of the Equations of Fluid Dynamics. *Math. Sbornik* **47**, pp 271-306.
- Haasdonk D, Kröner D and Rohde C (2000). Convergence of a Staggered Lax-Friedrichs Scheme for Nonlinear Conservation Laws on Unstructured Two-Dimensional Grids. *Preprint Mathematische Fakultät Freiburg 00-07*.
- Helzel C, LeVeque R J and Warnecke G (1999). A Modified Fractional Step Method for the Accurate Approximation of Detonation Waves. *Technical Report, Department of Applied Mathematics, University of Washington* **99-04**.
- Houston P, Mackenzie J A, Süli E and Warnecke G (1999). A Posteriori Error Analysis for Numerical Approximations of Friedrichs Systems. *Numer. Math.* **82**, pp 433-470.
- Johnson C and Szepessy A (1987). Convergence of a Finite Element Method for a Nonlinear Hyperbolic Conservation Law. *Math. Comp.* **49**, pp 427-444.
- Kaps P and Rentrop P (1979). Generalized Runge-Kutta Methods of Order Four with Stepsize Control for Stiff Ordinary Equations. *Numer. Math.* **33**, pp 55-68.
- Katsoulakis M A, Kossioris G and Makridakis C (1999). Convergence and Error Estimates of Relaxation Schemes for Multidimensional Conservation Laws. *Commun. Partial Differ. Equations* **24**, pp 395-424.

- Kröner D (1997). Numerical Schemes for Conservation Laws. Wiley Teubner.
- Kröner D, Noelle S and Rokyta M (1995). Convergence of Higher Order Upwind Finite Volume Schemes on Unstructured Grids for Scalar Conservation Laws in Two Space Dimensions. *Num. Math.* **71**, pp 527-560.
- Kröner D and Ohlberger M (2000). A Posteriori Error Estimates for Upwind Finite Volume Schemes for Nonlinear Conservation Laws in Multi Dimensions. *Math. Comput.* **69**, pp 25-39.
- Kröner D and Rokyta M (1994). Convergence of Upwind Finite Volume Schemes for Scalar Conservation Laws in Two Dimensions. *SIAM J. Numer. Anal.* **31**, pp 324-343.
- Küther M (2000). Error Estimates for Second Order Finite Volume Schemes Using TVD-Runge-Kutta Time Discretization for a Nonlinear Scalar Hyperbolic Conservation Law. *Preprint Mathematische Fakultät Freiburg* **00-05**.
- Küther M (2000). A Priori and A Posteriori Error Estimates for the Staggered Lax-Friedrichs Scheme in Multi Dimensions for Scalar Nonlinear Conservation Laws *Preprint Mathematische Fakultät Freiburg*.
- Kuznetsov N N (1976). Accuracy of Some Approximate Methods for Computing the Weak Solutions of a First-Order Quasi-Linear Equation. *Comput. Math. and Math. Phys.* **16**, pp 105-119.
- LeVeque R J and Yee H C (1990). A Study of Numerical Methods for Hyperbolic Conservation Laws with Stiff Source Terms. *J. Comput. Phys.* **86**, pp 187-210.
- Lube G (1992). Streamline Diffusion Finite Element Method for Quasilinear Elliptic Problems. *Num. Math.* **61**, pp 335-357.
- Noelle S (1995). Convergence of Higher Order Finite Volume Schemes on Irregular Grids. *Adv. Comput. Math.* **3**, pp 197-218.
- Ohlberger M (2000). A Posteriori Error Estimates for Vertex Centered Finite Volume Approximations of Convection-Diffusion-Reaction Equations. *Preprint Mathematische Fakultät Freiburg* **00-12**.
- Oran E S and Boris J P (1981). Theoretical and Computational Approach to Modeling Flame Ignition. *Prog. Aeronaut. Astronautics* **76**, pp 154-171.
- Oran E S, Weber J W Jr, Stefaniw E I, Lefebvre M H and Anderson J D Jr (1998). A Numerical Study of a Two-Dimensional H₂-O₂-Ar Detonation using a Detailed Chemical Reaction Model. *Comb. and Flame* **113**, pp 147-163.
- Rohde C (1996). Weakly Coupled Hyperbolic Systems. *PhD Thesis, Mathematische Fakultät Freiburg*.
- Strehlow R A (1969). The Nature of Traverse Waves in Detonations. *Astronautica Acta* **14**, pp 539-548.
- Tadmor E (1991). A Local Error Estimates for Discontinuous Solutions of Nonlinear Hyperbolic Equations. *SIAM J. Numer. Anal.* **28**, pp 891-906.
- Vila J P (1994). Convergence and Error Estimates in Finite Volume Schemes for General Multi-Dimensional Scalar Conservation Laws. I: Explicit Monotone Schemes *M²AN* **28**, pp 267-295.
- Williams F A (1985). Combustion Theory. The Benjamin/Cummings Publishing Company, Inc.
- Young T R and Boris J P (1977). A Numerical Technique for Solving Stiff Ordinary Differential Equations Associated with the Chemical Kinetics in Reactive-Flow Problems. *J. Phys. Chem.* **81**, pp 2424-2427.

3D VISUALIZATION OF SHOCK WAVES USING VOLUME RENDERING

J. O. LANGSETH

Norwegian Defence Research Establishment,

P.O.Box 25, NO-2027 Kjeller, Norway

Email: jan-olav.langseth@ffi.no

Abstract. Interactive performance is of vital importance in data visualization. Huge 3D simulations put special demands on visualization tools in order to keep this quality. Volume rendering is a powerful tool for visualizing 3D data, and can utilize graphics hardware with 2D or 3D textures. This rendering technique is demonstrated on data from a simulation of a shock/vorticity problem modeled by the Euler equations, and solved using a wave propagation scheme from the software package CLAWPACK. Emphasis is placed on the visualization of shock waves and discontinuities.

1. Introduction

Interactivity is essential for visualization to be an effective tool for analyzing three-dimensional data. As 3D computations produce more data, the wish for interactive visualization seems harder to fulfill. This paper will focus on the use of (direct) volume rendering, an approach that utilizes graphics systems with texture hardware. The visualization software **Viz**, developed at the Norwegian Defence Research Establishment, implements a fairly simple voxel rendering scheme, but has the advantage that most of the algorithm is implemented in hardware via textures.

Viz has proved to be an effective tool for studying computational results involving shocks and other discontinuities. Here, this will be illustrated through the use of a shock/vorticity problem modeled by the three-dimensional Euler equations. The problem is solved using a wave propagation scheme from the software package CLAWPACK. This scheme is a Godunov-type scheme, in the sense that Riemann problems are fundamental building blocks.

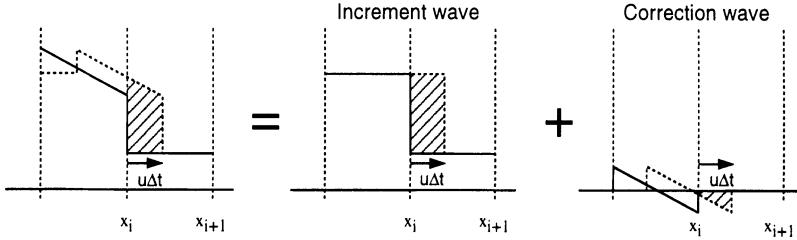


Figure 1. The flux updates in the CLAWPACK algorithms are split into the contribution from a constant state wave (the increment wave) and a linear wave (the correction wave). As usual in shock calculations, the slope of the latter is chosen to obtain both non-oscillatory and second order approximations.

2. CLAWPACK

The software package CLAWPACK (LeVeque, 1997), contains a collection of routines for solving a wide class of hyperbolic problems, both in conservative and in quasi-linear form. These unsplit schemes may be defined on both uniform rectangular and curvilinear grids.

The schemes are based on solving Riemann problems normal to the cell interfaces. The emanating waves are further split in the transverse direction by solving additional Riemann problems. Flux limiters are applied to suppress spurious oscillations arising from second order terms.

The solution to the Riemann problem is approximated by piecewise constant waves, e.g. those obtained in using a Roe solver. These waves, named increment waves, give rise to piecewise linear waves, named correction waves used for obtaining second order accuracy. In Figure 1, these waves are shown for the simple one-dimensional scalar advection equation $q_t + uq_x = 0$. The increment and correction waves in 3D are simple extensions of those obtained in 1D.

The dimensionality of the propagation of these waves defines a family of schemes with different accuracy and stability. For details on the 3D scheme cf. (Langseth and LeVeque, 1999).

There also exists a version of CLAWPACK using adaptive mesh refinement AMRCLAW (Berger and Leveque, 1998). This code is currently being rewritten using Fortran 90, and will eventually also include boundary embedding and parallel programming using MPI.

The CLAWPACK software is available on the web at
<http://www.amath.washington.edu/~rjl/clawpack.html>.

3. Problem and simulation

The three-dimensional Euler equations are used to model the interaction between shock waves and variable density regions. One practical application

of such problems is the study of how vorticity produced by shock waves mix two different gases. The mixing between two different gases in the 2D case has been studied in several papers, cf. (Wai Sun Don and Quillen, 1995), (Quirk and Karni, 1996). Here we study a one component gas only.

The initial condition consists of two cylindrically shaped regions, perpendicular to each other, cf. Figure 2a. Both regions contain constant state gas, initially at rest, and both cylinders have radius $r = 0.2$. The cylinder along the z -axis has pressure $p = 10$, while $p = 1$ elsewhere. Cylindrically shaped shock waves result from this region due to the overpressure. The other cylinder, with symmetry axis $x = 0.4$ and $z = 0$ contains a low density gas, with density $\rho = 0.1$. Elsewhere the density equals unity.

This configuration is chosen in order to produce vorticity, as easily seen from the vorticity equation,

$$\frac{\partial \vec{\omega}}{\partial t} = \nabla \times (\vec{u} \times \vec{\omega}) + \frac{\nabla \rho \times \nabla p}{\rho^2},$$

where $\vec{\omega} = \nabla \times \vec{u}$ is the vorticity.

As the shock wave propagates through the low density region, the latter winds up into two rotating regions (or rolls). The contact discontinuity resulting from the high pressure region will also be rolled up in this vortical motion. The upper roll rotates counter-clockwise, while the one below rotates in the clockwise direction. These two regions will after some time interact. Another vortical feature is the formation of vortex tubes on the envelope of the rolls. This is a pure three-dimensional feature, and creates a repeated pattern in the length direction of the low density region. These vortex tubes will get stretched due to the motion of the rolls and eventually burst, resulting in a turbulent looking region. In the early state of the shock interaction, the part of the wave that penetrates the low density region, speeds up due to the increased sound speed. This results in a splitting of the shock wave.

Symmetry is assumed across the planes $x = 0$, $y = 0$, and $z = 0$, and the computational domain is $[0, 1.5] \times [0, 1] \times [0, 0.5]$. Due to the turbulent-like behavior, these symmetries are not correct, but is selected to get a manageable grid size. The computation is performed on a $300 \times 200 \times 100$ grid, and the monotonized centered (MC) limiter is used. As expected, the details in the vortex dynamics depend on the limiters used, and the results in this region should only be taken qualitatively. The Roe approximate Riemann solver is used together with an entropy-fix, i.e. the standard Riemann solver in CLAWPACK for the Euler equations.

In Figure 2, the solution is shown at different times. Details on the quantities depicted and techniques used are given in the next section. Figure 2b shows an early stage ($t = 0.1$) of the interaction between the two cylinders.

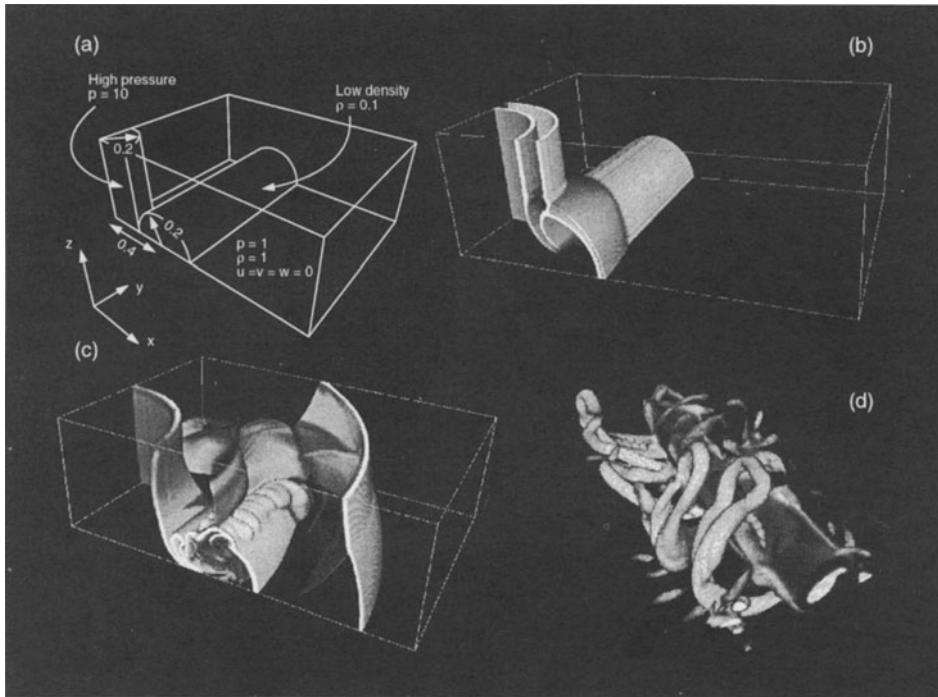


Figure 2. The solution to the shock/vorticity problem at different times: (a) The initial condition, (b) $t = 0.1$, (c) $t = 0.5$, and (d) $t = 0.8$. In (b) and (c), the quantity $|\nabla\rho|$ is depicted, while in (d) both the enstrophy $|\nabla \times \vec{u}|^2$ and the low-pressure is shown.

The roll-up of the contact discontinuities has started. Also note the wave that has started to encapsulate the low density region. As mentioned above, this is a result of the increased sound speed in the low density cylinder. The vertical cylinder with the largest radius is the shock wave, while the one inside is the contact discontinuity.

In Figure 2c ($t = 0.5$), the repeated pattern on the envelope of the roll is seen. Note that the shock wave has passed, and that another shock wave is visible on the top of the main vortex. This is the implosion shock resulting from the cylindrically shaped overpressure in the initial condition. In Figure 2d, details of the vortex tubes are shown at $t = 0.8$. The larger horizontal cylinder is the dominating low pressure region.

4. CFD visualization and volume graphics

In volume graphics, it is customary to classify rendering techniques into two groups. The first, and most common, is geometric rendering, where geometric objects are constructed from the 3D data and then rendered. A

typical example is the use of isosurfaces for scalar quantities, where the surface is constructed using simple polygons. Different lighting models can be applied to enhance the three-dimensionality. Rendering isosurfaces is a fairly slow process, since there do not exist hardware support for the construction of the surface. Another drawback with isosurfaces is that they are not very useful applied to complex flows.

The other main group of rendering techniques contains methods that render volumetric data without the use of geometries. This group is called volume rendering, and often *direct* volume rendering to indicate that there is no intermediate step where geometries are built. Volume rendering is most common in connection with visualization of scalar data.

Dealing with a volume of data, the concept of opacity (and it's complement transparency) is essential. The building block in volume rendering is the voxel, a cell with faces perpendicular to the scene's coordinate axes, and with information of the solution value (or associated color) and opacity. Typically one voxel is associated one data point. If the rendering is done the right way, it is possible to utilize special graphics hardware so that a high degree of interactivity is obtained even on huge data sets.

In many cases it may be valuable to combine both geometric and volume rendering. A typical example is the flow field around an obstacle. Volume rendering is perfect for visualizing the flow field, while the geometric approach is ideal for rendering the obstacle, say with coloring of the surface according to the pressure.

4.1. SHOCK VISUALIZATION

Ideally, the shock wave and other discontinuities should be handed over from the simulation as a mathematical surface. This will require a shock tracking approach that is not easy to realize, especially in 3D. Below, we focus on data obtained from a shock capturing method, where discontinuities are smeared over several cells.

Though there exist lighting models in volume graphics (Schroeder et. al., 1998), they are seldom used, much due to the complexity and the fact that the rendering is software-based. While geometric objects like isosurfaces have reflectant properties so that the use of light sources enhance the three-dimensionality, a volume set should be thought of as emitting light. One then could assume that volume rendering would result in "flat" images, with little information on the depth in the volume. In fact, using the right tools, obtaining images with a natural feeling of three-dimensionality is easy, cf. the images in this paper. Since features like shock waves or vortex tubes have a smeared approximation due to the discretization, we can assign darker values and decrease the opacity in the value range where the quantity

is falling off. The simple technique used is named limb-darkening, since the effect obtained is similar to what is well-known in solar physics.

Motivated by schlieren imaging, we have found the quantity $\mathcal{S} = |\nabla \rho|$ very useful. This quantity is large across both shock waves and contact discontinuities. Using volume graphics for visualizing shock structures, very little need to be done as long as this quantity is used. Typically, the opacity map is monotonic, starting with zero opacity in a neighbourhood of $\mathcal{S} = 0$. Then it increases to full opacity in another relatively small region. In the region where the opacity increases, the value V in the HSV color model (Schroeder et. al., 1998) should also change, typically from darker to brighter, to add shades or contrast to the image. (In the HSV model, a color is specified by its hue (H), its purity or saturation (S), and its lightness or value (V)). This simple, and very easy to use approach, is used for making Figures 2b-c. An isosurface-like image is shown, but there is no need to specify discrete values for \mathcal{S} . All discontinuities are visible, except for those very weak ones that may hide in the \mathcal{S} range, where zero opacity is specified. Note that due to the change in value V and opacity, weaker discontinuities appear to be darker, and more transparent. Discontinuities are rendered having a thickness depending on the resolution. For coarse grids, this may look a little bit odd. In Figure 2, the “internal” regions of the shock representation are filled with white. Note that the use of the HSV color model is essential in this context. An effect like limb-darkening is much harder to control using RGB.

Volume rendering is very useful when studying multiple scalar fields at the same time, and there exists several options for doing this. An obvious approach is to equip every field with its own color and opacity mapping. A less obvious possibility is to let one field determine the opacity and let the rest of the fields control the color.

As an example on the first approach, let \mathcal{S} and the enstrophy ($|\nabla \times u|^2$) have separate opacity and color mapping. Since \mathcal{S} is used only for visualizing the shock structure, only a single hue should be used. In Figure 3a-b, the solution is shown at $t = 0.2$. The schlieren quantity is assigned brighter grays, while the enstrophy is represented using darker tones. These images indicate that the vortex tubes originate from a region where the density cylinder collapses (marked with arrows).

Note that Figure 2d also is a two-field rendering of this type and that *limb brightening* is applied to the pressure field.

The second approach for visualizing two scalar fields at the same time is especially useful when visualizing the shock structure. Let \mathcal{S} define the opacity, so that only discontinuities will be visible. Next, let the pressure define the color. Then it will be easy to distinguish shock waves from contact discontinuities, since only the shock waves will be rendered with different

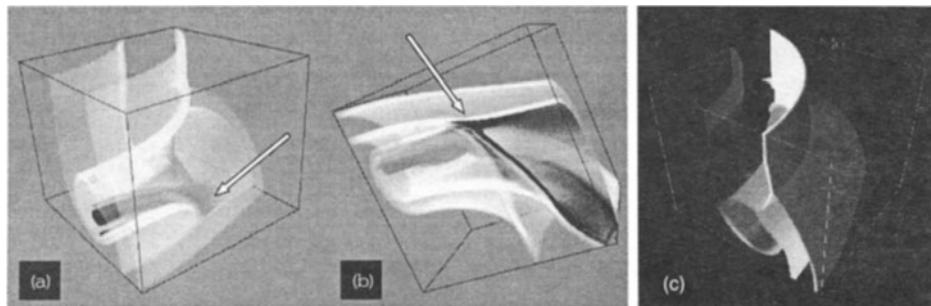


Figure 3. Visualizaton of two scalar fields simultaneously. In (a-b), the schlieren quantity \mathcal{S} (bright grays) and the enstrophy (dark grays) have separate color and opacity mappings. In (c), the opacity is defined by \mathcal{S} only, and the grays are defined by the pressure.

colors on the two sides of the surface. This is shown in Figure 3c, where a grayscale is used for the pressure, i.e. the higher pressure, the brighter grays. According to this, it is easily seen that the leading discontinuity is a shock wave, while the one in the back is a contact. (The use of grays is not very appropriate for these two-field images.)

5. Final remarks

All images shown are made using the volume visualization software **Viz**, developed at the Norwegian Defence Research Establishment. It is a total redesign of BoB (Bricks of Bytes) developed at the Army High Performance Computing Research Center at the University of Minnesota. While BoB was a typical software renderer, **Viz** utilizes 2D and 3D texture hardware, resulting in a considerable speed gain.

Texture mapping is a technique for adding images (textures) to a geometric object (Schroeder et. al., 1998). Dedicated hardware for performing texture mapping has become an important feature in all graphics computers, from the computer game machines to the super-computers making lifelike digital effects in movie productions. **Viz** is obtainable on the web at <ftp://ftp.ffi.no/spub/stsk/viz/index.html>.

The use of volume rendering is superior to the more usual isosurface approach when visualizing scalar fields. This has both to do with the fact that volume rendering allows a range of values to be studied, and that the use of texture hardware results in a speed gain allowing interactive visualization.

The popularity of volume rendering will increase with the general availability of appropriate hardware. Today, there exist promising graphics hardware for PCs with a performance that makes it possible to run **Viz** on fairly

large datasets.

Above, focus has been on visualizing scalar fields. Existing methods for visualizing vector fields would benefit from using texture based volume rendering, e.g. Line Integral Convolution (LIC) (Cabral and Leedom, 1993). Currently, a LIC algorithm for visualizing 3D vector fields is under development.

Color versions of the figures shown in this paper, in addition to other images from the shock/vorticity problem are available at
<ftp://ftp.ffi.no/spub/stsk/jol/web/godunov70/index.html>.

References

- Berger M J and LeVeque R J (1998). Adaptive mesh refinement using wave propagation algorithms for hyperbolic problems. *SIAM J. Numer. Anal.* **35**, pp 2298-2316.
- Cabral B and Leedom L (1993). Imaging vector fields using Line Integral Convolution. *Computer Graphics (SIGGRAPH '93 Proceeding)*, pp 263-272.
- Wai Sun Don, Quillen C B (1995). Numerical simulation of shock-cylinder interactions: I. Resolution. *J. Comput. Phys.* **122**, pp 244-265.
- Quirk J J and Karni S (1996). On the dynamics of a shock-bubble interaction. *J. Fluid Mech.* **318**, pp 129-163.
- Langseth J O and LeVeque R J (1999). A wave propagation method for three-dimensional hyperbolic conservation laws. *Submitted*.
- LeVeque R J (1997). Wave propagation algorithms for multi-dimensional hyperbolic problems. *J. Comput. Phys.* **131**, pp 327-353.
- Schroeder W, Martin K, and Lorensen B (1998). *The Visualization Toolkit: An Object-Oriented Approach to 3D Graphics*, 2nd ed. Prentice Hall PTR.

GAS FLOWS GENERATED BY PROPELLANT BURNING

C. A. LOWE

*Department of Applied Mathematics and Theoretical Physics,
Cambridge University,
Silver St., Cambridge, CB3 9EW, U.K.
Emails: c.a.lowe@damtp.cam.ac.uk
cazalowe@hotmail.com*

AND

J.F. CLARKE

*College of Aeronautics,
Cranfield University, Cranfield, Bedfordshire MK43 0AL, U.K.
Email: john.clarke@eidosnet.com*

Abstract. Exact, analytical, solutions to the set of Euler equations augmented by some rather particular forms of source functions are available, and play a useful role in helping to validate numerical solutions of some complex physical problems. As well as providing support for numerical algorithms, such exact solutions can sometimes also help with an understanding of the physics of a real problem. Here, an investigation of a combustion problem illustrates this fact.

1. Introduction

Some engineering problems in combustion and multiphase flow can be modelled by a system of partial differential equations, which express the familiar conservation laws for the flow of a fluid, but augmented by a number of source or interaction terms. In such situations it is the sources that provide the essential driving mechanisms for the physical problem that is to be studied. The work below extends a current topic of research, which describes a general class of exact solutions for the Euler equations with source terms (Lowe and Clarke, 2000). In this latter work a combination of analysis and numerics has been used to investigate a system in which

positive contributions to the mass, momentum and energy in the system are added, in a spatially non-uniform manner, via the source mechanisms; the primary purpose behind this work is to create test-problems for the validation of the numerical algorithms which must usually be employed to evaluate the complex nonlinear fields in such cases, although some light is also shed on physical behaviour within the system. In the present paper these ideas will be extended to explore a combustion scenario: The equations that model combustion do not fit into the class of problems for which exact, analytical, mathematical solutions are possible. The main objective of the present paper is to examine a particular feature, that is typical of combustion in a compressible medium by using numerical methods and corresponding mathematical analysis, which immediately reveals the origins of this feature in the production of entropy. The content of the work naturally follows an earlier study (Clarke and Toro, 1985) which presented numerical results that illustrated, but did not explain, the feature.

2. Conservation and Other Equations

The treatment of a system, in which motion of its gaseous components is driven by the burning of pieces of solid propellant immersed in the gas phase, is described in §§ 1 & 2 of (Clarke and Lowe, 1996). It is explained in that paper how the addition of mass, momentum and energy to the gas-phase, via the processes of burning of the solid propellant, can be modelled as a collection of sources, distributed in space and time, which augment the set of conservation equations for one-dimensional unsteady geometries, and we shall use these ideas in the present exercise. We shall write specific total energy of the gas-phase as E , defined by

$$E \equiv e + \frac{1}{2}u^2, \quad \text{where} \quad e = \sum c_i e_i, \quad (1)$$

u is gas velocity, e is the specific internal or intrinsic energy of the gaseous material and e_i is the internal energy of each chemical species i in the mixture; \sum implies summation over all relevant i . Assume that the specific heats at constant volume C_v are individually constant and the same for all species; in addition all the gases are assumed to be ideal. If the energy of formation for a species is written as Q_i , the internal energies for the gaseous components can be written as:

$$e_i = e^{th} + Q_i; \quad \text{where} \quad e_i^{th} = C_v T \equiv e^{th}. \quad (2)$$

Evidently e^{th} is the thermal energy per unit mass of each species in the mixture, and of the mixture as a whole; T is absolute temperature and

Q_i is the energy of formation of species i , as just defined. It follows that $e = \sum c_i e_i = e^{th} + \sum c_i Q_i$ and the total energy E is therefore given by

$$E = e^{th} + \frac{1}{2}u^2 + \sum c_i Q_i . \quad (3)$$

The ideal gas assumption implies that

$$p = \rho RT; \quad a^2 = \gamma RT = \gamma p v; \quad e^{th} = p v / (\gamma - 1), \quad (4)$$

where ρ is density, $v (= \rho^{-1})$ is specific volume, p is pressure, γ is the ratio of specific heats and $R = \mathcal{R}/\mathcal{W}$ is a constant where \mathcal{R} is the Universal gas constant, \mathcal{W} is the molecular weight of each gas (which again for simplicity will be assumed to be the same for all constituents of the gas-phase) and a is the local sound-speed. In this case the compressible inviscid Euler equations are given in differential form as

$$\mathbf{U}_t + \mathbf{F}_x(\mathbf{U}) = \mathbf{S}(\mathbf{U}) . \quad (5)$$

where

$$\mathbf{U} = \begin{pmatrix} \rho \\ \rho c_i \\ \rho u \\ \rho E \end{pmatrix}, \quad \mathbf{F}(\mathbf{U}) = \begin{pmatrix} \rho u \\ \rho u c_i \\ \rho u^2 + p \\ u \rho (E + p v) \end{pmatrix}, \quad \mathbf{S}(\mathbf{U}) = \begin{pmatrix} G \\ K_i + \phi_i G \\ F \\ H \end{pmatrix} . \quad (6)$$

G is the rate of mass addition to the system, K_i is the rate of chemical reaction of species i , ϕ_i is the fraction of total mass input G that consists of species i ; F is the rate of momentum addition and H is the rate of energy addition.

2.1. THE SYSTEM IN CHARACTERISTIC FORM

The flow equations in shock-free flow can be written in the following characteristic form:

$$-a^2 \frac{\partial \rho}{\partial t} \Big|_{\psi} + \frac{\partial p}{\partial t} \Big|_{\psi} = (\gamma - 1)(\hat{H} - \sum Q_i \hat{K}_i), \quad \frac{\partial x}{\partial t} \Big|_{\psi} = u; \quad (7)$$

$$\frac{\partial c_i}{\partial t} \Big|_{\psi} = v \hat{K}_i, \quad \frac{\partial x}{\partial t} \Big|_{\psi} = u; \quad (8)$$

$$\rho a \frac{\partial u}{\partial t} \Big|_{\beta} + \frac{\partial p}{\partial t} \Big|_{\beta} = a(F - uG) + (\gamma - 1)(\hat{H} - \sum Q_i \hat{K}_i) + a^2 G,$$

$$\frac{\partial x}{\partial t} \Big|_{\beta} = u + a; \quad (9)$$

$$\begin{aligned} -\rho a \frac{\partial u}{\partial t} \Big|_{\alpha} + \frac{\partial p}{\partial t} \Big|_{\alpha} &= -a(F - uG) + (\gamma - 1)(\hat{H} - \sum Q_i \hat{K}_i)) + a^2 G, \\ \frac{\partial x}{\partial t} \Big|_{\alpha} &= u - a; \quad (10) \end{aligned}$$

where $\hat{H} = H - u(F - uG) - G(E + pv)$ and $\hat{K}_i = K_i + (\phi_i - c_i)G$. Here, constant ψ defines a particle path and constant α and β correspond to acoustic wavelets propagating at speeds $u - a$ and $u + a$, respectively. The thermodynamic (or Gibbs) equation relates increments of entropy ds with increments in internal energy de , etc, via the relation

$$Tds = de + pdv - \sum \mu_i dc_i, \quad (11)$$

where μ_i are specific Gibbs potentials.

When the increments in s , e , etc, apply to continuous temporal changes in gaseous material following a particle path (11) can be rewritten as

$$\begin{aligned} T s_t \Big|_{\psi} &= e_t \Big|_{\psi} + p v_t \Big|_{\psi} - \sum \mu_i c_i t \Big|_{\psi}, \\ &= v \hat{H} - v \sum \mu_i \hat{K}_i, \end{aligned} \quad (12)$$

by using (7) and the fact (from (5) and (6)) that $e_t \Big|_{\psi} + p v_t \Big|_{\psi} = v \hat{H}$, provided, of course, that shock waves do not intervene.

We note that equations (9) and (10) can be rewritten in terms of dependent variables, chemically-frozen sound-speed a , gas velocity u , entropy s and mass fractions c_i to give:

$$\frac{2}{\gamma - 1} a_{\alpha} + u_{\alpha} = \frac{1}{\rho a} S_+ t_{\alpha} + \frac{a}{(\gamma - 1) C_p} s_{\alpha} - \frac{1}{\rho a} \sum [(p c_i)_{a,s,c_j \neq i}] c_i \alpha, \quad (13)$$

$$\frac{2}{\gamma - 1} a_{\beta} - u_{\beta} = \frac{1}{\rho a} S_- t_{\beta} + \frac{a}{(\gamma - 1) C_p} s_{\beta} - \frac{1}{\rho a} \sum [(p c_i)_{a,s,c_j \neq i}] c_i \beta, \quad (14)$$

where

$$S_{\pm} \equiv \pm a(F - uG) + (\gamma - 1)(\hat{H} - \sum Q_i \hat{K}_i) + a^2 G. \quad (15)$$

Remember that \sum indicates the sum over all indices i that denote a chemical species in the mixture.

Equations (13) and (14) demonstrate how changes of composition and entropy can have a direct effect on the local values of gas velocity and acoustic wave speed and hence, of course, on the propagation of waves of compression and expansion through the system. However, if the flow is homentropic and of fixed composition, then analytical solutions do exist; the system defined in (Clarke, 1987) fits into this category of problem. For such a homentropic problem equations (13) and (14) reduce to two coupled equations for u and a that can be, and have been, solved analytically.

3. A Model Combustion Problem

A simple combustion problem, previously investigated in (Clarke and Lowe, 1996) in a different context, will now be considered. Reactant gases, oxidant X and fuel F, are added to an enclosed system by the burning of solid propellant particles. It will be assumed that these reactants burn irreversibly to produce an inert product P, as in the chemical equation



Let the fraction of reactant gas added into the closed system be such that:

$$\phi_X = \phi_F = \frac{1}{2}. \quad (17)$$

In this case it can be shown (via (7) for $i = X, F$) that the system described in detail in §2 reduces so that a single reactant equation will describe the evolution of either oxidant mass-fraction or fuel mass-fraction, that is:

$$c_X = c_F \equiv c. \quad (18)$$

A combustion process can then be fully described by using the simple chemical scheme given by equation (16), the system presented in (13), (14), (15) and, most important, the following choice of sources:

$$\begin{aligned} G &= \rho \bar{G}/a, \quad K_X = K_F \equiv K = -\rho \Omega c^2, \\ F &= u \rho \bar{G}/a, \quad H = (e^{th} + \frac{1}{2} u^2 + p v + Q) \rho \bar{G}/a; \end{aligned} \quad (19)$$

where \bar{G} is a step function such that:

$$\bar{G} = \begin{cases} 1294301m^3/s; & x < 0 \\ 0; & x > 0. \end{cases} \quad (20)$$

Note that this choice of relation for G in terms of \bar{G} follows from the work described in (Clarke, 1987) in pursuit of simple analytical solutions; it has similar beneficial effects in the present case. Take particular note of the single chemical source term associated with chemical reaction, namely K , which implies that both oxidant and fuel react at the same rate. The chemical reaction is of second-order and the chemical frequency Ω is assumed to be given by the following Arrhenius-type expression:

$$\Omega = Ap \exp(-E_A/RT). \quad (21)$$

A is a constant pre-exponential factor and E_A is an activation energy. For the choice of source terms given in (19) it can be shown, from the definitions of \hat{H} and \hat{K}_i in §2.1, immediately below (10), that:

$$\hat{H} = GQ(1 - 2c) \quad \text{and} \quad \hat{K} = -\rho\Omega c^2 + (\frac{1}{2} - c)G. \quad (22)$$

3.1. CHEMICALLY INERT MOTION

Let the pre-exponential factor $A \rightarrow 0$, so that $\Omega \rightarrow 0$ too, and the reaction rate $K (= -\rho\Omega c^2)$ therefore also tends to zero; this implies that the chemical reaction is frozen. It follows that \hat{K} (see equation (22)) reduces to $(\frac{1}{2} - c)G$. If we now assume that the values of c at time zero are all equal to $\frac{1}{2}$ it means that

$$\hat{K} = 0 \quad (23)$$

and c will remain equal to $\frac{1}{2}$ from then on (*cf* (8)). In other words the gas-phase will consist of equal parts of X and F, which are now chemically-inert substances. Equation (22) indicates that, when $c = \frac{1}{2}$, $\hat{H} = 0$, and equations (13) and (14) reduce to two equations for just u and a alone; in fact these equations are identical to the system explored by Clarke (1987), for which a complete analytical solution can be explicitly derived.

3.2. VERY FAST CHEMISTRY; LOCAL EQUILIBRIUM

In contrast to the situation examined in §3.1 we now assume that the chemical reaction in the gas-phase is very fast, which means that we must examine what happens when $\Omega \rightarrow \infty$. The rate of consumption of X and F must remain bounded, for physical reasons, which means that, as $\Omega \rightarrow \infty$, c must approach zero and, as a consequence, so must $c_t|_\psi$. Then (8) shows that $\hat{K} \rightarrow 0$ and the bounded reaction rate $K (= -\rho\Omega c^2) \rightarrow -\frac{1}{2}G$, as is clear from (22) when $c \rightarrow 0$. Note that the relation between K and $-\frac{1}{2}G$ can be rewritten in the form

$$\mathcal{D}c^2 \rightarrow \frac{1}{2} \quad \text{where} \quad \mathcal{D} \equiv \frac{\rho\Omega}{G}. \quad (24)$$

\mathcal{D} is a local Damköhler number that defines a ratio of the rate of reactant consumption to the rate of supply of reactant. A graph of the quantity $\mathcal{D}c^2$ versus distance for three times is included in figure 3 below, where it will be seen that $\mathcal{D}c^2$ does indeed tend to the value one-half as c diminishes towards zero in appropriate, near-equilibrium, circumstances. No species equation is required for the conditions examined in this subsection as, throughout,

the gas will be composed of fully-reacted inert product alone and the mass, momentum and energy relations of equation (5), and the sources given in (19), define the fast-reaction, or equilibrium model, combustion problem. In this case, (22) shows that equation (12) for entropy reduces to

$$Ts_t|_{\psi} = GQ. \quad (25)$$

The entropy of the system is seen to increase along particle paths at a rate that is directly proportional to the energy of formation Q . It should be pointed out that the second term on the right-hand side of the entropy equation (12), namely $-v \sum \mu_i \dot{K}_i$, accounts for the creation of entropy by chemical reactions (*cf.* §1.14 in the book by Clarke & McChesney (1976)). The fact that this particular term vanishes in present circumstances is wholly consistent with the current condition of an arbitrarily-fast chemical reaction and its corollary of local states of chemical equilibrium, as described in §1.14 of the book just referred to above. The increases in entropy that do occur in the present exercise arise from the instantaneous liberation of chemical energy Q as reactants enter the gas-phase, via the sources of mass G , when $\Omega \rightarrow \infty$ as is clear from (25). The increase of entropy in this particular system, that is in chemical equilibrium, is significant. All analytical solutions that have been presented so far by (Clarke, 1987) and (Lowe and Clarke, 2000) have not involved the presence of entropy waves as a consequence of their *a priori* limitation to homentropic conditions. For combustion problems, (25) demonstrates that entropy waves will propagate and that their amplitude will bear a direct relationship to the magnitude of Q . The acoustic wavelets governed by (13) and (14) will also respond to the presence of entropy changes; in present circumstances $\hat{H} - \sum \mu_i \dot{K}_i = GQ$, so that S_{\pm} in (15) retains a dependence on Q . In this case there are three simultaneous equations, namely (25), (13) and (14), for the dependent variables u , a and s , and these cannot be reduced to a simple analytical form, such as that seen earlier in previous work, even though the system under discussion in this section is in local chemical equilibrium. More generally, the foregoing results demonstrate that, for an Euler system that includes combustion source terms, entropy increases will be present in the gas motion when limitingly fast chemical reactions occur, and local chemical equilibrium exists, as well as for the more realistic situation where non-equilibrium chemistry takes place. In the latter case, entropy variations will generally be stronger since the term $-v \sum \mu_i \dot{K}_i$, that is part of the last term in (12), must be positive under nonequilibrium conditions (§1.14 in Clarke & McChesney (1976)). Only in the case for which chemical reaction rates are zero, and therefore equivalent to the situation for which the flow is chemically inert (see §3.1) can entropy remain constant in the absence of shock waves. The following numerical solutions will confirm the above

theoretical deductions and will include some data for the case of realistic finite reaction rates that connects the two limiting cases that have been dealt with in some detail in §§3.1 and 3.2.

4. Numerical and Analytical Solutions

The numerical scheme uses time-operator splitting to produce two sub-problems; a homogeneous hyperbolic problem and a system of ordinary-differential-equations (ODE's). Shock-capturing finite volume methods are employed to solve the former, a second-order Godunov-type scheme, and a second-order Runge-Kutta method is used for the system of ODE's - details of the numerical method can be found in (Lowe, Toro and Clarke, 1995).

The input variables and initial conditions that have been used in the following simulations are: $\gamma = 1.4$; $p_0 = 101400 \text{ Pa}$; $a_0 = 330.34 \text{ ms}^{-1}$ and $u_0 = 0 \text{ ms}^{-1}$. The spatial domain is $[-0.5m, 0.5m]$ and a thousand (spatial) computational cells have been used.

Consider the problem discussed in section 3.1: Figure 1 describes the numerical solution for the case in which¹ $Q = 1 \text{ MJ/kg}$ and $A = 0$ so that the system corresponds with the one investigated in section 3.1 where the reactant mass-fraction is fixed at the value $c = 1/2$ throughout. Figure 1 compares the numerical solution of this particular form of the combustion problem with the analytical solution of the problem, described in (Clarke, 1987), which makes the *a priori* assumption that the system is chemically inert. The density, velocity and pressure are displayed at times 0.2 ms , 0.4 ms and 0.6 ms . It is evident that the numerical solution is identical with the analytical solution, as indeed it should be since, with $A = 0$, no combustion energy is being released; this system is identical with the one defined by making $Q = 0$ and $A \neq 0$, provided $c(x, 0) = 0$.

Comparison of figure 1 (inert flow; §3.1) with figure 2 (local equilibrium flow, §3.2) reveals some distinct differences between the two situations. The most immediately noticeable difference is apparent in the density profiles, which diminish monotonically as distance x increases when the system is inert, but which have sharp local maxima in $x > 0$ when chemical equilibrium prevails. However, *one must take careful note of the very different ordinate scales in the two figures*.

In broad terms, the excursions in density are smaller under equilibrium conditions (figure 2) than they are under inert, or frozen chemistry, conditions. The reverse is true for pressures. The simple inference (*cf.* 4) is that the local absolute temperatures T are higher when chemical equilibrium prevails than when no chemical energy is added to the flow, as one would expect!

¹A typical propellant has energy of combustion that is of this magnitude.

When T is higher, local sound speeds are higher (*cf.* equation (4)), as is apparent via comparison of the positions of expansion-wave heads, in $x < 0$, and of compression-wave heads, in $x > 0$, at times 0.2 ms, 0.4 ms and 0.6 ms. Note the strong shock wave² that has been created at the right-hand end of the regions of compression in $x > 0$ by the time t is equal to 0.4 ms. The implication is that entropy has increased markedly in regions behind the shock³, which prompts us to look closely at the effects of entropy production that arise when chemical reaction rates are limitingly fast, as they are for the conditions illustrated in figure 2.

The relation between the rate of production of entropy per unit mass following the motion of a fluid particle, namely $s_t|_\psi$, the local rate of mass-addition G , and the quotient Q/T , is given in (12) for the fast-chemistry, or local equilibrium, limit. Recalling the choice of sources that has been made in (19) and (20), G is only greater than zero in $x < 0$ and it can then be seen from (12) that entropy will only be created in fluid particles when these are themselves in $x < 0$. All non-zero gas velocities are positive, and there will therefore be a path in $x > 0$ for $t > 0$ (say, $\psi = \psi_c$), which passes through $x = 0$ when $t = 0$, which divides gas originally in $x > 0$ from the remainder of that material. In the circumstances, we should expect to see some change in the character of flow fields across such a particle path, which resembles a contact discontinuity. The results displayed in figure 2 suggest that pressure variations are smooth in those regions of x -space that contain sharp peaks in the density-profiles at particular values of t . Since we can write ρ as a function of p and s in the present case, it follows that any jumps in density gradients ρ_x will imply jumps in gradients of entropy s_x , which is a plausible explanation for the behaviour seen in figure 2 in light of the arguments that have been put forward earlier in this paragraph. The sharp peaks in density will lie on the path ψ_c .

The numerical results in figures 3 and 4 are derived for $Q = 1\text{ MJ/kg}$ and $A = 1\text{ Pa/s}$ and the addition of initial-value information about the concentration field $c(x, t)$, namely that $c(x, 0) = 0$. The gasdynamical motion is therefore accompanied by strictly non-equilibrium chemical effects (last term on the right-hand side of (12) with values of \hat{H} and \hat{K} for the model combustion problem given in (22)) superimposed on the influences of the first term on the right-hand side of (12). That features from both figure 1 and figure 2 can be seen in figures 3 and 4 is therefore not a surprise and neither is the more complicated character of the resulting flow fields in view of the additional mechanisms for the creation of entropy.

²The shock Mach number is roughly equal to 2 at $t = 0.4\text{ ms}$.

³Note the discussion in (Clarke, 1987) of the geometric extent of the entropy domain behind such a shock.

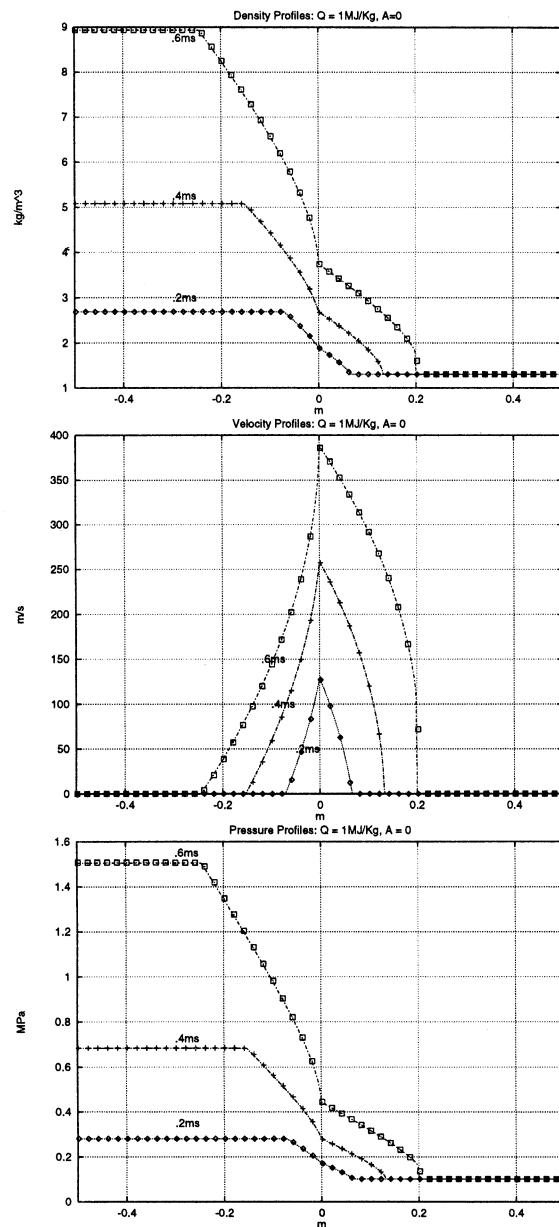


Figure 1. Numerical solution for non-equilibrium problem: $Q = 1\text{MJ/Kg}$, $A = 0$, and analytical solution for equilibrium case where $Q = 0$ at times 0.2ms, 0.4ms and 0.6ms: line is exact solution and symbol is numerical.

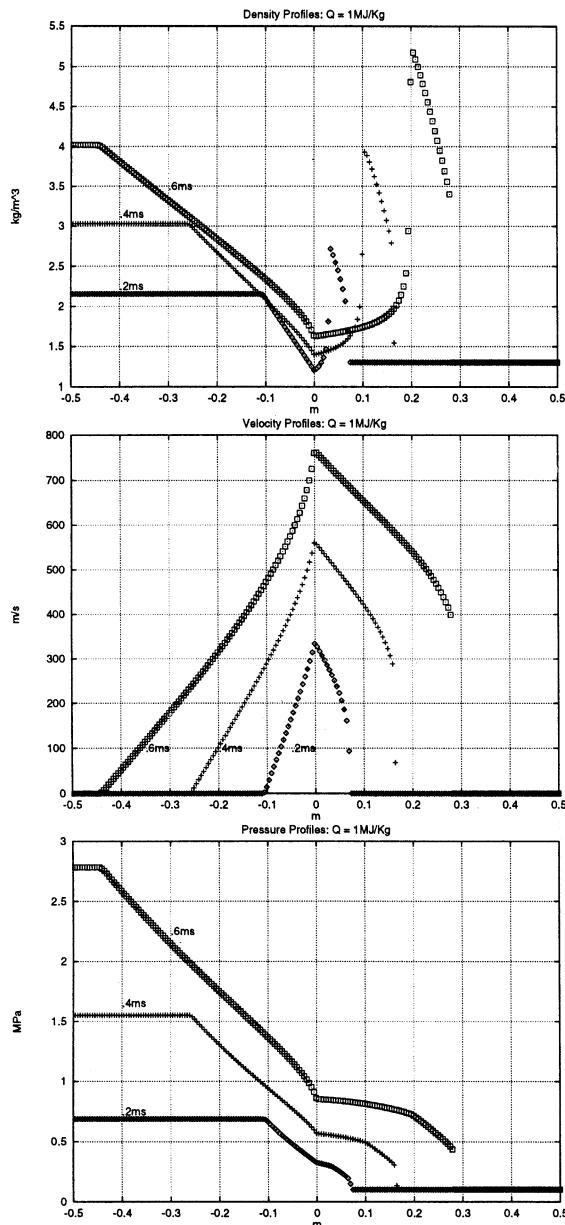


Figure 2. Numerical solution for equilibrium problem: $Q = 1 \text{ MJ/kg}$ at times 0.2ms , 0.4ms and 0.6ms .

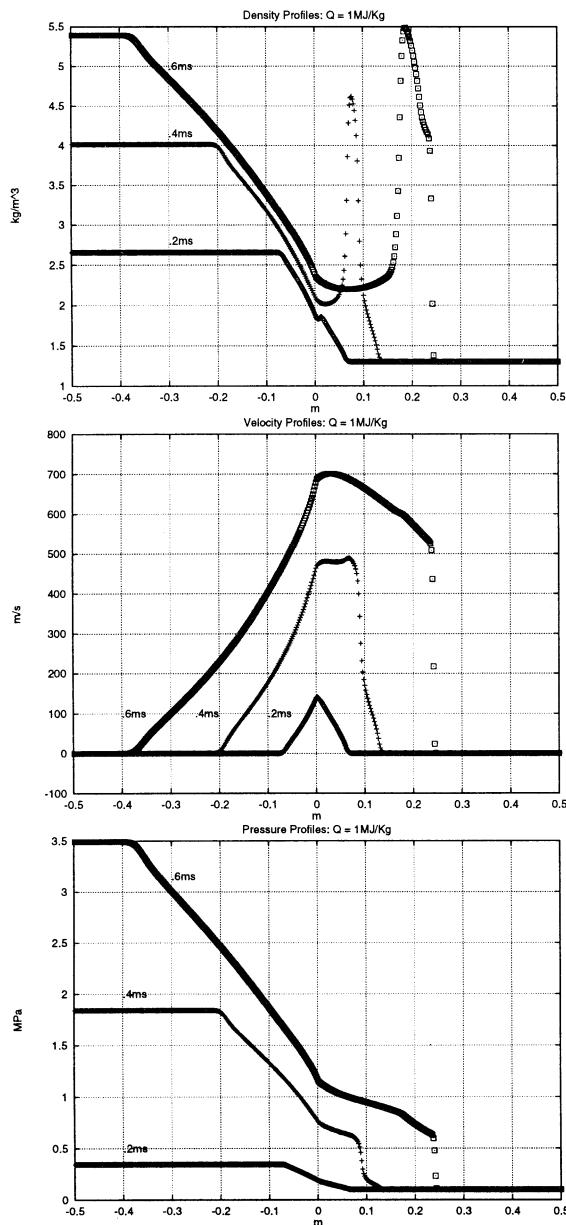


Figure 3. Numerical solution of density, velocity and pressure for non-equilibrium problem: $Q = 1 \text{ MJ/kg}$, $\frac{E_A}{R} = 1,000 \text{ K}$, $A = 1 \text{ Pas}^{-1}$ at times 0.2 ms , 0.4 ms and 0.6 ms . Observe how the discontinuous pressure, which is the movement of a propagating shock, travels faster than the contact wave, corresponding to the discontinuous drop in mass-fraction seen in figure 4 below.

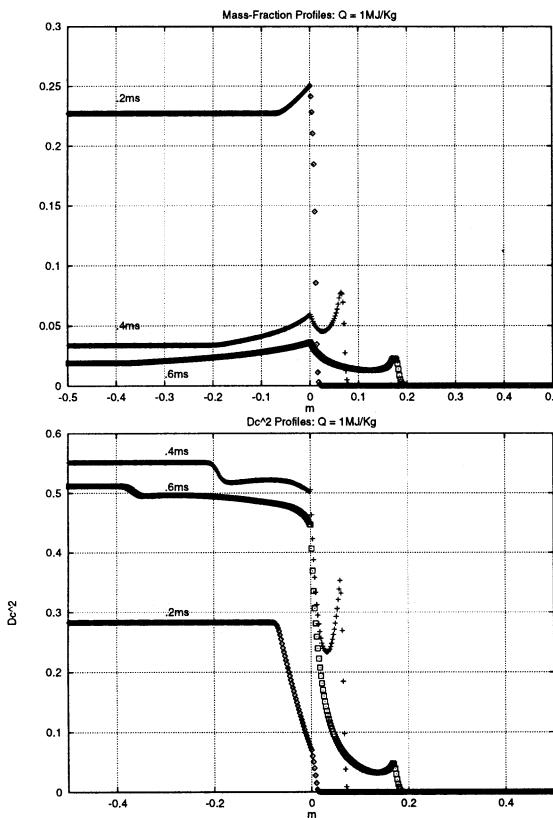


Figure 4. Numerical solution of mass-fraction and Dc^2 for non-equilibrium problem: $Q = 1 MJ/kg$, $\frac{E_A}{R} = 1,000 K$, $A = 1 Pas^{-1}$ at times $0.2ms$, $0.4ms$ and $0.6ms$.

5. Conclusion

This study illustrates how the inclusion of source terms associated with material combustion will initiate the production of large disturbances to local entropy values. The amplitude of these disturbances is directly proportional to the material exothermicity. For non-equilibrium simulations, significant local variations of entropy will begin to emerge once significant amounts of exothermic energy start to be released. For both non-equilibrium *and* equilibrium problems with combustion source terms, entropy increases will play an important role in the flow behaviour, as can be seen from the figures.

Acknowledgements

Many thanks to Clive Woodley of the Defence Evaluation and Research Agency (DERA) at Fort Halstead, U.K. who sponsored this work.⁴

References

- Clarke J F and Lowe C A (1996). Combustion with Source Flows. *Mathematical and Computer Modelling* **24** 8, pp 95-104.
- Lowe C A and Clarke J F (1999). Aspects of solid propellant combustion. *Philosophical Transactions of the Royal Society* **357** 1764, pp 3639-3653.
- Lowe C A and Clarke J F(2000) A class of analytical solutions to the Euler equations with source terms: Part I. Submitted to *Mathematical and Computer Modelling*.
- Clarke J F and Toro E F (1985). Gas Flows Generated by Solid-Propellant Burning. Numerical simulation of combustion phenomena, pp 192-205. Glowinski R, Larroutourou B and Teman R (Editors). Springer-Verlag Publishers.
- Clarke J F (1987). Compressible flow produced by distributed source of mass: An Exact solution. College of Aeronautics Report: COA -8710, Cranfield University, Cranfield, Beds. MK43 0AL, U.K.
- Clarke J F and McChesney M (1976). Dynamics of Relaxing Gases. Butterworths, London.
- Lowe C A, Toro E F and Clarke J F (1995). Numerical Methods for propellant systems. The First Asian Computational Fluid Dynamics Conference pp 541-549 at Hong Kong University of Science and Technology, Clearwater Bay, Hong Kong.

FINITE VOLUME EVOLUTION GALERKIN METHODS FOR MULTIDIMENSIONAL HYPERBOLIC SYSTEMS

M. LUKÁČOVÁ-MEDVIĐOVÁ

*Institute of Mathematics, Technical University Brno,
Technická 2, 616 39 Brno, Czech Republic,
(also Otto-von-Guericke Universität Magdeburg)
Emails: Lukacova@fme.vutbr.cz,
Maria.Lukacova@mathematik.uni-magdeburg.de*

K.W. MORTON

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

AND

G. WARNECKE

*Otto-von-Guericke-Universität Magdeburg,
PSF 4120, 39 106 Magdeburg, Germany
Email: Gerald.Warnecke@mathematik.uni-magdeburg.de*

Abstract. This contribution describes high-resolution genuinely multidimensional finite volume evolution Galerkin schemes. These methods couple a finite volume formulation with approximate evolution Galerkin operators. They are constructed using the bicharacteristics of the multidimensional hyperbolic system, such that all of the infinitely many directions of wave propagation are taken into account. We have derived approximate evolution operators for the linear wave equation system as well as for the nonlinear Euler equations. Second order resolution is obtained with a conservative piecewise bilinear recovery and the second order midpoint rule for the time integration.

1. Introduction

Consider a general hyperbolic conservation law in d space dimensions

$$\underline{U}_t + \sum_{k=1}^d (\underline{F}_k(\underline{U}))_{x_k} = 0, \quad \underline{x} = (x_1, \dots, x_d)^T \in \mathbb{R}^d, \quad (1)$$

where $\underline{F}_k = \underline{F}_k(\underline{U})$, $k = 1, \dots, d$ represent given physical flux functions and the conservative variables are $\underline{U} = (u_1, \dots, u_m)^T \in \mathbb{R}^m$. Let us denote by $E(s) : (H^k(\mathbb{R}^d))^m \rightarrow (H^k(\mathbb{R}^d))^m$ the exact evolution operator associated with a time step s acting on Sobolev spaces for the system (1), i.e.

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

We suppose that S_h^p is a finite element space consisting of piecewise polynomials of order $p \geq 0$. Let \underline{U}^n be an approximation in the space S_h^p to the exact solution $\underline{U}(\cdot, t_n)$ at a time $t_n > 0$ and take $E_\tau : S_h^r \rightarrow (H^k(\mathbb{R}^d))^m$ to be a suitable approximation to the exact evolution operator $E(\tau)$, $r > 0$. We denote by $R_h : S_h^p \rightarrow S_h^r$ a recovery operator, $r > p \geq 0$. In the present paper we shall limit our considerations to cases of constant time step Δt , i.e. $t_n = n\Delta t$, and of a uniform mesh consisting of d -dimensional cubes with a uniform mesh size h .

Definition 1.1 Starting from some initial value \underline{U}^0 at time $t = 0$, the finite volume evolution Galerkin method (FVEG) is recursively defined by means of

$$\underline{U}^{n+1} = \underline{U}^n - \frac{1}{h} \int_0^{\Delta t} \sum_{k=1}^d \delta_{x_k} \underline{F}_k(\underline{U}^{n+\tau/\Delta t}) d\tau, \quad (2)$$

where the central difference $v(x + h/2) - v(x - h/2)$ is denoted by $\delta_x v(x)$ and $\delta_{x_k} \underline{F}_k(\underline{U}^{n+\tau/\Delta t})$ represents an approximation to the edge flux difference at intermediate time levels $t_n + \tau$, $\tau \in (0, \Delta t)$. The cell boundary flux $F_k(\underline{U}^{n+\tau/\Delta t})$ is evolved using the approximate evolution operator E_τ to $t_n + \tau$ and averaged along the cell boundary, i.e. e.g. on vertical edges for \underline{U} itself

$$\underline{U}^{n+\tau/\Delta t} = \frac{1}{h} \int_0^h E_\tau R_h \underline{U}^n dS_y d\tau. \quad (3)$$

Analogous formula holds for horizontal edges.

For the computation of fluxes on cell interfaces the value of \underline{U} has to be determined by means of an approximate evolution operator. The most important advantage of this formulation is that even a first order accurate approximation E_τ to the evolution operator $E(\tau)$ yields an overall second order update from \underline{U}^n to \underline{U}^{n+1} . The second order scheme is obtained by a conservative discontinuous bilinear recovery using the vertex values and by

the midpoint rule for time integration. Thus, the finite volume evolution Galerkin scheme (2) gives

$$\underline{U}^{n+1} = \underline{U}^n - \frac{\Delta t}{h} \sum_{k=1}^d \delta_{x_k} \underline{F}_k(\underline{U}^{n+*}), \quad (4)$$

where

$$\underline{F}_k(\underline{U}^{n+*}) = \frac{1}{h} \int_0^h F_k(E_{\Delta t/2} R_h \underline{U}^n) dS. \quad (5)$$

In the next sections we illustrate this procedure for the wave equation system and for the nonlinear Euler equation system in two space dimensions and derive approximate evolution operators for both systems.

2. Wave equation system

Denote by ϕ, u, v the unknown functions of the wave equation system

$$\begin{aligned} \phi_t + c(u_x + v_y) &= 0, \\ u_t + c\phi_x &= 0, \\ v_t + c\phi_y &= 0. \end{aligned} \quad (6)$$

Consider a cone with the apex $P = (\underline{x}, t + \Delta t)$ and the base points $Q = Q(\theta) = (\underline{x} + c\Delta t \cos \theta, \underline{y} + c\Delta t \sin \theta, t)$ parametrized by the angle $\theta \in [0, 2\pi]$. Denote by $P' = (\underline{x}, t)$ the center of the base of the cone. The lines from $Q(\theta)$ to P generating the mantle of the so-called bicharacteristic cone are called bicharacteristics, see, e.g., (Lukáčová, Morton and Warnecke, 2000) for more details.

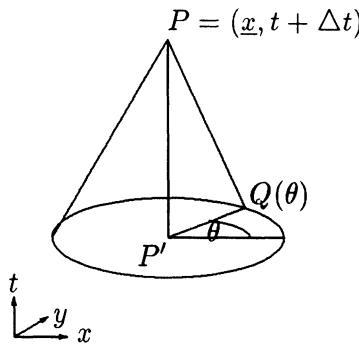


Figure 1. Bicharacteristic along the Mach cone through P and $Q(\theta)$

Using the theory of bicharacteristics it can be shown that the solution (ϕ, u, v) at the point P is determined by its values on the base as well as

on the mantle of the characteristic cone and the exact evolution formulae can be derived. In our recent papers (Lukáčová, Morton and Warnecke, 1998), (Lukáčová, Morton and Warnecke, 1999), (Lukáčová, Morton and Warnecke, 2000) several approximate evolution operator E_Δ for the wave equation system were derived and analysed. The following evolution operator, which was derived using a general theory of the hyperbolic systems, gives the smallest amount of numerical diffusion, see (Lukáčová, Morton and Warnecke, 2000). In what follows we omit the formulae for v_P because they are analogous to the approximate evolutions of u_P .

Approximate Evolution Operator E_Δ for the wave equation

$$\phi(P) = \frac{1}{2\pi} \int_0^{2\pi} \phi(Q) - 2u(Q) \cos \theta - 2v(Q) \sin \theta d\theta + O(\Delta t^2), \quad (7)$$

$$\begin{aligned} u(P) = & \frac{1}{2} u(P') + \frac{1}{2\pi} \int_0^{2\pi} -2\phi_Q \cos \theta + u(Q)(3 \cos^2 \theta - 1) \\ & + 3v(Q) \sin \theta \cos \theta d\theta + O(\Delta t^2). \end{aligned} \quad (8)$$

3. Euler equations

In (Lukáčová, Morton and Warnecke, 1998) some results of the evolution Galerkin methods for the Euler equations have already been shown. Here we will derive a new scheme using the finite volume formulation. In order to compute fluxes the cell interface value \underline{U}^{n+*} using the bicharacteristics has to be determined. In fact, any finite volume evolution Galerkin scheme can be considered as a generalization of the Osher-Solomon flux (Osher and Solomon, 1982) in a multidimensional manner.

In order to evolve fluxes at the cell interfaces let us consider the two-dimensional system of the Euler equations in primitive variables

$$\underline{W}_t + \underline{\underline{A}}_1(\underline{W})\underline{W}_x + \underline{\underline{A}}_2(\underline{W})\underline{W}_y = 0, \quad (9)$$

where

$$\underline{W} := \begin{pmatrix} \rho \\ u \\ v \\ p \end{pmatrix}, \quad \underline{\underline{A}}_1 := \begin{pmatrix} u & \rho & 0 & 0 \\ 0 & u & 0 & 1/\rho \\ 0 & 0 & u & 0 \\ 0 & \gamma p & 0 & u \end{pmatrix}, \quad \underline{\underline{A}}_2 := \begin{pmatrix} v & 0 & \rho & 0 \\ 0 & v & 0 & 0 \\ 0 & 0 & v & 1/\rho \\ 0 & 0 & \gamma p & v \end{pmatrix}.$$

Here ρ , u , v , p denote density, components of velocities and pressure, respectively. The isentropic exponent γ is taken to be 1.4 for dry air. To

derive the evolution operator the system (9) is linearized locally at a point $P' := (x, y, t)$. Denote by c' the local speed of sound at the point P' , i.e. $c' := \sqrt{\gamma p/\rho(P')}$, and by u' , v' the local flow velocities at the point P' . Due to the advection of flow given by the nonzero diagonal entries of the Jacobians \underline{A}_1 and \underline{A}_2 the Mach cone is slanted. The apex of the Mach cone is a point $P := (x, y, t + \Delta t)$ and the center of the base is $Q_0 := (x - u'\Delta t, y - v'\Delta t, t)$. The mantle of the Mach cone is generated by bicharacteristics, which are slanted lines connecting the apex P and the base points $Q := (x - u'\Delta t + c'\Delta t \cos \theta, y - v'\Delta t + c'\Delta t \sin \theta, t)$. Thus, according to subsonic or supersonic character of the local flow the point P' lies inside or outside the base of the Mach cone, respectively. Integrating the characteristic equations along the bicharacteristics yields the exact integral representation. Applying the rectangle rule to time integration and integration by parts for simplification of the source terms, cf. Lemma 2.1 in (Lukáčová, Morton and Warnecke, 2000) the following approximate evolution operator can be derived.

Approximate evolution operator E_Δ for the Euler equations

$$\begin{aligned} \rho(P) &= \rho(Q_0) - \frac{p(Q_0)}{c'^2} + \frac{1}{2\pi} \int_0^{2\pi} \frac{p(Q)}{c'^2} - 2\frac{\rho'}{c'} u(Q) \cos \theta \\ &\quad - 2\frac{\rho'}{c'} v(Q) \sin \theta d\theta + O(\Delta t^2), \end{aligned} \quad (10)$$

$$\begin{aligned} u(P) &= \frac{1}{2} u(Q_0) + \frac{1}{2\pi} \int_0^{2\pi} -\frac{2}{\rho' c'} p(Q) \cos \theta + u(Q)(3 \cos^2 \theta - 1) \\ &\quad + 3v(Q) \sin \theta \cos \theta d\theta + O(\Delta t^2), \end{aligned} \quad (11)$$

$$\begin{aligned} p(P) &= \frac{1}{2\pi} \int_0^{2\pi} p(Q) - 2\rho' c' u(Q) \cos \theta - 2\rho' c' v(Q) \sin \theta d\theta \\ &\quad + O(\Delta t^2), \end{aligned} \quad (12)$$

where $\rho' = \rho(P')$. Now the approximate evolution operator works on the primitive variables \underline{W} instead of using the conservative variables $\underline{U} := (\rho, \rho u, \rho v, e)$, cf. (4). Afterwards \underline{U}^{n+*} at cell interfaces will be recomputed from \underline{W}^{n+*} and fed into the FV formulation (5). This approach gives a reasonable approximation for continuous nonlinear problems, but if the solutions are discontinuous and shocks occur a sophisticated approximation of the flow speeds c' , u' , v' has to be made, e.g. by using some multidimensional upwinding.

4. Conclusions

In the present paper we have described a new finite volume evolution Galerkin scheme for the linear wave equation system and the nonlinear Euler equations. In our recent papers (Lukáčová, Morton and Warnecke, 1998), (Lukáčová, Morton and Warnecke, 1999), (Lukáčová, Morton and Warnecke, 2000) we have shown that the proposed schemes work successfully for linear problems. We are currently carrying out numerical experiments for nonlinear problems. The finite volume evolution Galerkin methods are a genuine generalisation of the original idea of Godunov (Godunov, 1959) using an evolution operator for a system in multidimensions. They combine the usually conflicting design objectives of using the conservation form and following the characteristics, or bicharacteristics. Instead of solving one dimensional Riemann problems in normal directions to cell interfaces by some approximate Riemann solvers, we use the genuinely multidimensional approach. The approximate solution at cell interfaces is computed by means of an approximate evolution operator using bicharacteristics. This is a novel feature of our method.

Acknowledgements

The present research has been supported under the DFG Grant No. Wa 633/6-2 of Deutsche Forschungsgemeinschaft, and partially by the Grant No. CZ 39001/2201 of the Technical University Brno as well as the German-Israeli-Foundation (GIF) Grant I-318-195 06/93.

References

- Godunov S K (1959). Finite Difference Methods for Numerical Computation of Discontinuous Solutions of the Equations of Fluid Dynamics. *Matem. Sbornik* **47(89)**, pp. 271 - 306. (in Russian).
- Lukáčová - Medviđová M, Morton K W, Warnecke G (1998). On the Evolution Galerkin Method for Solving Multidimensional Hyperbolic Systems. *Proceedings of ENUMATH'98*, World Scientific Publishing Company, Singapore, pp. 445-452.
- Lukáčová - Medvidová M, Morton K W, Warnecke G (1999). High-Resolution Finite Volume Evolution Galerkin Schemes for Multidimensional Conservation Laws. *Proceedings of ENUMATH'99*, World Scientific Publishing Company, Singapore.
- Lukáčová - Medvidová M, Morton K W, Warnecke G (2000). Evolution Galerkin Methods for Hyperbolic Systems in Two Space Dimensions. appear in *MathComp*.
- Osher S, Solomon F (1982). Upwind Difference Schemes for Hyperbolic Systems of Conservation Laws. *Math. Comp.* **38**, pp. 339 - 374.

THE NUMERICAL SIMULATION OF RELATIVISTIC FLUID FLOW WITH STRONG SHOCKS

ANTONIO MARQUINA

*Departamento de Matemática Aplicada,
University of València,
46100 Burjassot, Spain.
Emails: marquina@uv.es*

Abstract. In this review we present and analyze the performance of a Godunov type method applied to relativistic fluid flow. Our model equations are the corresponding Euler equations for special relativistic hydrodynamics. By choosing an appropriate vector of unknowns, the equations of special relativistic fluid dynamics (RFD) can be written as a hyperbolic system of conservation laws. We give a complete description of the spectral decomposition of the Jacobian matrices associated to the fluxes in each spatial direction, (see (Donat et al., 1998), for details), which is the essential ingredient of the Godunov-type numerical method we propose in this paper. We also review a numerical flux formula that avoids/reduces numerical difficulties appearing in the ultrarelativistic regime, (i.e., high Lorentz factors). Using the spectral decompositions in a fundamental way, we construct high order versions of the basic first order scheme described by Donat and Marquina in ((Donat, Marquina, 1996)). We study, as a sample, a particular shock tube test where special difficulties arise. We show two dimensional simulations where strong shocks are present, including a supersonic jet stream in a strongly ultra-relativistic scenario.

1. Introduction

Astrophysical scenarios involving relativistic flows have drawn the attention and efforts of many researchers since the pioneering studies of May and White (May and White, 1967) and Wilson (Wilson, 1971), (Wilson, 1972). Relativistic jets, accretion onto compact objects (in X-ray binaries or in the inner regions of active galactic nuclei), stellar core collapse, coalescing

compact binaries (neutron star and/or black holes) and recent models of Gamma-ray bursts (GRBs) are examples of systems in which the evolution of matter is described within the frame of the Theory of Relativity (Special or General).

The term Relativistic Fluid Dynamics, (RFD), applies to both those flows in which the velocities (of individual particles or of the fluid as a whole) approach c , the velocity of light in vacuum, or those where the effects of the background gravitational field -or that generated by the matter itself- are so important that a description in terms of Einstein theory of gravity becomes necessary.

In the most simple model which considers the matter as a *perfect fluid* (free of energy transport processes, no magnetic fields,...), its evolution in those situations of astrophysical interest is governed by the *hyperbolic system of conservation laws of the relativistic hydrodynamics*. Simulations based on the numerical integration of the hydrodynamical equations provide a valuable tool to confront the theoretical models with the observations (as in Astrophysics) or the experimental results (as in Nuclear Physics, (see (Clare and Strottman, 1986) and (Strottman, 1989))), which explains the rapid progress, during the last few years towards the development of reliable RFD-codes that work accurately under the extreme conditions of interest. The first Eulerian code in RFD was developed by Wilson (Wilson, 1972), on the basis of explicit finite-differencing techniques and monotonic transport. The code incorporated artificial viscosity techniques based on earlier work of Richtmyer and Morton for the non-relativistic flow equations. Wilson's code and its sequels have been widely used in numerical RFD simulations; however, despite its popularity (almost all codes in numerical relativistic hydrodynamics in the eighties were based in Wilson's procedure) it turned out to be unable to resolve the extremely strong shock structures that appear in the ultrarelativistic regime.

Norman and Winkler analyzed in depth the artificial viscosity approach to RFD in (Norman and Winkler, 1986). Their research led them to the conclusion that a fully implicit treatment of the relativistic equations was the only way to increase the accuracy of artificial viscosity formulations in the ultrarelativistic regime.

By the mid-eighties, and fueled by an increasing awareness that the artificial viscosity approach was of limited use in the ultrarelativistic regime, part of the numerical RFD community started to look at other shock capturing techniques that had been successfully applied in classical gas dynamics to obtain accurate numerical approximations in the presence of strong shocks. Although shock capturing techniques for hyperbolic systems of conservation laws were designed with the Euler equations in mind, great care was devoted to the task of formulating them within a systematic mathemat-

ical framework that could turn them into general purpose methods for *any* hyperbolic system of conservation laws.

The first explicit shock capturing codes in RFD without artificial viscosity appear in the early nineties (Martí, 1991; Marquina, 1992; Eulderink, 1993). These codes follow the so-called "Godunov approach", and their design is based on two main points: 1) The ability to write the RFD equations as a system of hyperbolic conservation laws, identifying a suitable vector of unknowns. 2) An *approximate Riemann solver* built using the spectral decomposition of the Jacobian matrices of the system. Nowadays, several research groups in Astrophysics and Nuclear Physics have successfully implemented many of the most successful Godunov-type techniques and their high resolution (higher order) extensions in the RFD context. As a result, accurate numerical simulations in the ultrarelativistic regime have started to appear (Balsara, 1994; Dai, Woodward, 1997; Dolezal et al., 1995; Duncan et al., 1994; Eulderink and Mellema, 1995; Falle et al., 1996; Martí and Mueller, 1995; Martí et al., 1995; Schneider et al., 1993; Wen et al., 1997). Excellent recent reviews can be found in (LeVeque, 1998), (Martí and Mueller, 1999).

High resolution shock capturing, (HRSC for short), methods are now routinely used in classical gas dynamics to discretize the convective derivatives of a general system of convection-diffusion-reaction equations in any number of spatial dimensions. It is well known, although not particularly well understood, that many of these shock capturing techniques can, on occasions, fail quite spectacularly. An excellent review on the numerical pathologies that can be encountered in gas dynamics simulations is given by Quirk (Quirk, 1994). Usually, the pathological behavior is local, and does not cause the code to crash, but, in complicated situations, we have observed disastrous effects on the numerical approximations.

As in Newtonian hydrodynamics, HRSC methods are starting to become part of numerical codes designed to model more complicated situations in RFD. In the references given above, many of the local pathologies observed in Newtonian hydrodynamics can also be observed in relativistic tests.

In (Donat et al., 1998), an explicit, ready-to-use, formulation of the *full spectral decomposition* of the Jacobian matrices associated with the fluxes in three dimensions was given. This was the essential ingredient in many of the more sophisticated HRSC techniques. The explicit description of the spectral decomposition of the Jacobian matrices makes possible to use a new numerical method described in (Donat, Marquina, 1996). The numerical experimentation shown in (Donat, Marquina, 1996) indicates that this shock capturing technique is less prone to developing numerical pathologies than some well-used methods, (such as Roe-type methods or Shu-Osher characteristic-based ENO schemes).

The scheme described in (Donat, Marquina, 1996), which will be referred to as *Marquina's flux split scheme* henceforth, and its ability to elude some of the local pathological behavior encountered when using some of the better known HRSC alternatives has been put to work to obtain accurate numerical simulation in highly ultrarelativistic regimes (Donat et al., 1998).

In this paper, we shall present a review of this method in the context of Special Relativistic Euler equations, showing its robustness when it is applied to some model problems where strong shocks are present, showing its advantages and disadvantages with respect to other standard shock-capturing methods.

The review is organized as follows: In section 2 we describe Marquina's flux-split algorithm in the context of characteristic-based schemes. In section 3 we review our model equations, i.e., the special relativistic Euler equations, (SREE), showing the complete spectral decomposition of the Jacobian. In section 4 we study, as a sample, a particular shock tube test where special difficulties arise. In section 5 we explore a local pathology appearing in the relativistic extension of a benchmark experiment in Newtonian hydrodynamics: Emery's step test. We also present an astrophysical application, the evolution of some relativistic jets moving at different supersonic speeds.

2. Characteristic-based Methods

To ensure that discontinuities are captured by the scheme, i.e. they move at the right speed even if they are unresolved, we must write the discrete equations in *conservation form*. That is a form in which the rate of change of conserved quantities is equal to a difference of fluxes. This form guarantees that we conserve the total amount of the states U present, in analogy with the integral form of the system of conservation laws.

A fully discrete *conservative method* has the form

$$U_i^{n+1} - U_i^n + \frac{\Delta t}{\Delta x} [F_{i+\frac{1}{2}} - F_{i-\frac{1}{2}}] = 0$$

where $F_{i+\frac{1}{2}} = F(U_{i-p}, \dots, U_{i+q})$ and F is the numerical flux function of the scheme.

In the Riemann solver approach, the numerical flux function is computed by solving a Riemann problem at each cell boundary. The main purpose served by introducing a Riemann solver (either exact or approximate) into a finite difference scheme is that of providing physical realism by correctly discriminating between information which should propagate with different speeds.

This is a recurrent theme when solving hyperbolic equations, since the direction in which information propagates is determinant to construct sta-

ble upwind finite difference schemes capable of approximating their exact solutions.

The local characteristic structure, and thus the local upwind directions, can also be obtained by diagonalizing the Jacobian matrix rather than by solving directly a Riemann problem. This approach has been used in flux-vector splitting schemes (e.g. (Steger and Warming, 1981)) and it is the general technique used in characteristic-based methods.

Let us consider a system of N convective conservation laws in one spatial dimension,

$$U_t + [F(U)]_x = 0.$$

The basic idea of characteristic numerical schemes is to transform this non-linear system to a set of (nearly) independent scalar equations of the form

$$u_t + vu_x = 0$$

discretize each scalar equation independently in a v -upwind biased fashion, and then transform the discretized system back into the original variables.

In a smooth region of the flow, we can get a better understanding of the structure of the system by expanding out the derivative as

$$U_t + JU_x = 0$$

where $J = \frac{\partial F}{\partial U}$ is the Jacobian matrix of the system. In a hyperbolic system this matrix is diagonalized by the matrices of left-multiplying and right-multiplying eigenvectors of J . If L is the matrix whose rows are the left eigenvectors of J and R is the matrix whose columns are the right eigenvectors of J we have

$$LJR = \text{Diag}(\lambda^p)$$

and the eigenvectors λ^p are all real.

Suppose we want to discretize our equation at the point x_0 , where L and R have values L_0 and R_0 . To get a locally diagonalized form, we multiply our system of equations by the *constant* matrix L_0 which nearly diagonalizes J over the region near x_0 (we require a constant matrix so that we can move inside all derivatives):

$$[L_0 U]_t + L_0 J R_0 [L_0 U]_x = 0$$

We have inserted $I = R_0 L_0$ to put the equation in a more recognizable form. The spatially varying matrix $L_0 J R_0$ is exactly diagonalized at the point x_0 , with eigenvalues λ_0^p , and it should be nearly diagonal at nearby points. The equations should thus be sufficiently decoupled for us to apply upwind-biased discretizations independently to each component, with λ_0^p determining the wind direction for the p -th component equation. Once this

system is fully discretized, we multiply the entire system by $L_0^{-1} = R_0$ to return to the original variables.

The Jacobian matrix J is quite important to any characteristic based scheme, as it defines the local linearization of the non-linear problem. It determines the transformation to the local characteristic fields, and thus what the upwind directions are, as well as what quantities are to be upwind differenced.

Recall that in a conservative scheme we require values of the numerical flux at the cell boundaries, i.e. the midpoints between nodes. Thus, in order to transform to characteristic fields to evaluate numerical fluxes, we need the spectral decomposition (eigenvalues and eigenvectors) of the Jacobian at each cell wall. Since only the values of U at grid points are known, the evaluation of the Jacobian at each cell boundary requires some form of interpolation.

The characteristic based approach has been extensively used in the design of Essentially Non Oscillatory (ENO) schemes. In standard ENO schemes it was thought that the precise form of this interpolation was not so important, but recent developments show that in fact it can make a great deal of difference in causing or eliminating certain numerical pathologies.

The standard ENO method uses a single Jacobian evaluated at the linear average of the states at nodes adjacent to the midpoint,

$$J_{i+\frac{1}{2}} = J \left(\frac{U_i + U_{i+1}}{2} \right)$$

In smooth regions, this centered linear approximation is second order accurate; however, in a smooth region it makes little difference whether the derivatives are computed in an upwind biased fashion or in some combination of upwind and downwind. The precise determination of the Jacobian (and the transformation to characteristic fields) is not so important there. It is between nodes in an unresolved steep gradient that the centrally averaged Jacobian (or even the Jacobian evaluated at some other average, like the Roe mean, for example) might cause problems. In this case, an artificially constructed averaged Jacobian can differ significantly from the left and right Jacobian matrices interpolated from left and right nodal state values, and there is no clear reason why any averaged Jacobian should be the right choice for a proper transformation to characteristic variables at a cell boundary.

Near an unresolved steep gradient in the flow, in which the states vary by a large amount from one node to the next, the unambiguous values of the two Jacobian matrices obtained by extrapolation from nodal data at each side of the cell boundary might differ substantially. It, thus, makes sense to try to use these two Jacobian matrices separately, in an upwind fashion,

rather than attempt to define a single representative midpoint Jacobian. This is the driving principle of the flux-split formula described in (Donat, Marquina, 1996) and (Donat et al., 1998).

For the sake of completeness, we shall include here the analytical formulation of the numerical flux separating the interfaces U_L and U_R in Marquina's flux-split scheme.

Compute first the *sided* local characteristic variables and fluxes:

$$\begin{aligned}\omega_L^p &= L^p(U_L) \cdot U_L & \phi_L^p &= L^p(U_L) \cdot F(U_L) \\ \omega_R^p &= L^p(U_R) \cdot U_R & \phi_R^p &= L^p(U_R) \cdot F(U_R)\end{aligned}$$

Here $L^p(U_L)$, $L^p(U_R)$, for $p = 1, 2, \dots, m$, are the (normalized) left eigenvectors of the Jacobian matrices $J(U_L), J(U_R)$. Let $\lambda_p(U_L), \lambda_p(U_R)$, $p = 1, 2, \dots, m$, be their corresponding eigenvalues.

Then proceed as follows:

For $k = 1, \dots, m$

If $\lambda_k(U)$ does not change sign in $[U_L, U_R]$, then

If $\lambda_k(U_L) > 0$ then

$$\begin{aligned}\phi_+^k &= \phi_L^k \\ \phi_-^k &= 0\end{aligned}$$

else

$$\begin{aligned}\phi_+^k &= 0 \\ \phi_-^k &= \phi_R^k\end{aligned}$$

endif

else

$$\begin{aligned}\alpha_k &= \max_{U \in \Gamma(U_L, U_R)} |\lambda_k(U)| \\ \phi_+^k &= .5 \cdot (\phi_L^k + \alpha_k \omega_L^k) \\ \phi_-^k &= .5 \cdot (\phi_R^k - \alpha_k \omega_R^k)\end{aligned}$$

endif

$\Gamma(U_L, U_R)$ is a curve in the space of states of the system connecting U_L and U_R . For any hyperbolic system where the fields are either genuinely nonlinear or linearly degenerate, we can test the possible sign changes of $\lambda_k(U)$ by checking the sign of $\lambda_k(U_L) \cdot \lambda_k(U_R)$. Also, α_k can be determined as

$$\alpha_k = \max\{|\lambda_k(U_L)|, |\lambda_k(U_R)|\}.$$

The numerical flux that corresponds to the cell-interface separating the states U_L and U_R is then

$$F^M(U_L, U_R) = \sum_{p=1}^m (\phi_+^p R^p(U_L) + \phi_-^p R^p(U_R)) \quad (1)$$

Marquina's scheme can thus be interpreted as a characteristic-based scheme that avoids the use of an *averaged* intermediate state to perform the transformation to the local characteristic fields. The ambiguity in choosing this average is avoided by using directly the unambiguous data on the left and right sides of each cell wall.

As it was pointed out in (Donat, Marquina, 1996) the first order Marquina's scheme suffers from a built-in heat conduction mechanism, that is substantially reduced for high order formulations. To construct higher order versions of the scheme, we follow the *method of lines* approach of (Shu and Osher, 1989). The discretization process is carried out in two steps: First we use an ENO spatial reconstruction for the numerical flux functions. Then we discretize in time using the TVD Runge-Kutta ODE solvers developed in (Shu and Osher, 1989). Our preferred ENO reconstruction is the Piecewise Hyperbolic Method (PHM) developed in (Marquina, 1994).

The extension to higher dimensions is accomplished, as in (Shu and Osher, 1989), in a dimension by dimension fashion, so that the one dimensional method applies unchanged to higher dimensional problems.

3. Special Relativistic Euler Equations

In this section we will review the Euler equations in special relativity, we used as model equations in (Donat et al., 1998). The choice of this model is justified since the equations for general relativity can be locally reduced, through a local change of coordinates, to the special relativistic case, and, therefore, any solver for the model equations may apply to a more complex problem, (see (Pons et al., 1998) for details).

The evolution of a relativistic fluid is described by a system of equations which are the expression of *local conservation laws*: the local conservation of baryon number density, and the local conservation of energy-momentum

$$\nabla_\mu(\rho u^\mu) = 0, \quad \nabla_\mu T^{\mu\nu} = 0 \quad (2)$$

(throughout the paper, Greek indices run from 0 to 3 and Latin indices from 1 to 3 and units in which the speed of light is equal to one are used). Here, ρ is the rest-mass density, u^μ the 4-velocity vector and ∇_μ stands for the covariant derivative. The energy-momentum tensor, $T^{\mu\nu}$, describes the physical properties of matter. For example, for a perfect fluid

$$T_{\mu\nu} = \rho h u_\mu u_\nu + p g_{\mu\nu} \quad (3)$$

where p is the pressure and h is the specific enthalpy, defined as

$$h = 1 + \varepsilon + p/\rho, \quad (4)$$

with ε being the specific internal energy. The tensor $g_{\mu\nu}$ defines the metric of the space-time \mathcal{M} where the fluid evolves. Here I will focus on Minkowski space-time, (readers interested in the general-relativistic formulation can address to Ibáñez talk in this volume).

Here $x^\mu = (t, x, y, z)$; ρ , h and p are as defined in section 1, and $v^i = u^i/W$ where W , the Lorentz factor ($W \equiv u^0$), satisfies $W = (1 - v^2)^{-1/2}$, with $v^2 = \delta_{ij}v^i v^j$.

We consider the following variables

$$\begin{aligned} D &= \rho W \\ S^j &= \rho h W^2 v^j \\ \tau &= \rho h W^2 - p - \rho W, \end{aligned} \tag{5}$$

which are, respectively, the rest-mass, momentum and total energy densities, measured in the laboratory frame. Thus, defining the vector of conserved quantities as

$$\mathbf{u} = (D, S^j, \tau) \tag{6}$$

then the system 2 takes the required *conservative form*:

$$\frac{\partial \mathbf{u}}{\partial t} + \sum_i \frac{\partial \mathbf{f}^i(\mathbf{u})}{\partial x_i} = 0. \tag{7}$$

The system of partial differential equations is closed, as usual, with an equation of state $p = p(\rho, \varepsilon)$. Anile (Anile, 1989) has shown that system 7 is hyperbolic for causal equations of state, i.e., those satisfying $c_s < 1$, where c_s , defined as

$$hc_s^2 = \frac{\partial p}{\partial \rho} + (p/\rho^2) \frac{\partial p}{\partial \epsilon}, \tag{8}$$

is the local sound velocity.

When applying a shock capturing technique to a conservative formulation of system 2, the code evolves the conserved quantities, in time. The local rest-frame variables $\{\rho, \varepsilon, p\}$ and the three-velocity v^j have to be computed at least once per time step in each computational cell. This computation requires a non-linear root-finding routine (Martí and Müller, 1995).

Following (Donat et al., 1998), we will derive the analytical expressions for the spectral decomposition of the three 5×5 Jacobian matrices \mathcal{B}^i associated to the fluxes $\mathbf{f}^i(\mathbf{u})$ of system 7

$$\mathcal{B}^i = \frac{\partial \mathbf{f}^i(\mathbf{u})}{\partial \mathbf{u}} \tag{9}$$

The eigenvalues of matrix $\mathcal{B}^x(\mathbf{u})$ are (the $i = y, z$ cases are easily obtained by symmetry):

$$\lambda_{\pm} = \frac{1}{1 - v^2 c_s^2} \left\{ v^x (1 - c_s^2) \sqrt{(1 - v^2)[1 - v^x v^x - (v^2 - v^x v^x)c_s^2]} \right\} \quad (10)$$

$$\lambda_0 = v^x \quad (\text{triple}) \quad (11)$$

Let us note however, that the characteristic wave speeds in the relativistic case not only depend on the fluid velocity components in the wave propagation direction, but also on the normal velocity components. This coupling adds new numerical difficulties which are specific to RFD.

To give the expression of the right and left eigenvectors, we define the following auxiliary quantities:

$$\mathcal{K} \equiv \frac{\tilde{\kappa}}{\tilde{\kappa} - c_s^2} \quad (12)$$

$$\mathcal{A}_{\pm} \equiv \frac{1 - v^x v^x}{1 - v^x \lambda_{\pm}} \quad (13)$$

A complete set of *right-eigenvectors* is,

$$\mathbf{r}_{0,1} = \left(\frac{\mathcal{K}}{hW}, v^x, v^y, v^z, 1 - \frac{\mathcal{K}}{hW} \right) \quad (14)$$

$$\mathbf{r}_{0,2} = \left(Wv^y, 2hW^2v^xv^y, h(1 + 2W^2v^yv^y), 2hW^2v^yv^z, 2hW^2v^y - Wv^y \right) \quad (15)$$

$$\mathbf{r}_{0,3} = \left(Wv^z, 2hW^2v^xv^z, 2hW^2v^yv^z, h(1 + 2W^2v^zv^z), 2hW^2v^z - Wv^z \right) \quad (16)$$

$$\mathbf{r}_{\pm} = (1, hW\mathcal{A}_{\pm}\lambda_{\pm}, hWv^y, hWv^z, hW\mathcal{A}_{\pm} - 1) \quad (17)$$

The corresponding complete set of *left-eigenvectors* is

$$\mathbf{l}_{0,1} = \frac{W}{\mathcal{K} - 1}(h - W, Wv^x, Wv^y, Wv^z, -W)$$

$$\mathbf{l}_{0,2} = \frac{1}{h(1 - v^x v^x)}(-v^y, v^x v^y, 1 - v^x v^x, 0, -v^y)$$

$$\mathbf{l}_{0,3} = \frac{1}{h(1-v^x v^x)} (-v^z, v^x v^z, 0, 1 - v^x v^x, -v^z)$$

$$\left[\begin{array}{c} hW\mathcal{A}_\pm(v^x - \lambda_\pm) - v^x - W^2(v^2 - v^x v^x)(2\mathcal{K} - 1)(v^x - \mathcal{A}_\pm \lambda_\pm) + \mathcal{K}\mathcal{A}_\pm \lambda_\pm \\ 1 + W^2(v^2 - v^x v^x)(2\mathcal{K} - 1)(1 - \mathcal{A}_\pm) - \mathcal{K}\mathcal{A}_\pm \\ W^2 v^y (2\mathcal{K} - 1) \mathcal{A}_\pm (v^x - \lambda_\pm) \\ W^2 v^z (2\mathcal{K} - 1) \mathcal{A}_\pm (v^x - \lambda_\pm) \\ -v^x - W^2(v^2 - v^x v^x)(2\mathcal{K} - 1)(v^x - \mathcal{A}_\pm \lambda_\pm) + \mathcal{K}\mathcal{A}_\pm \lambda_\pm \end{array} \right]$$

where Δ is the determinant of the matrix of right-eigenvectors.

$$\Delta = h^3 W (\mathcal{K} - 1) (1 - v^x v^x) (\mathcal{A}_+ \lambda_+ - \mathcal{A}_- \lambda_-) \quad (18)$$

For an ideal gas equation of state $\mathcal{K} = h$, thus $\mathcal{K} > 1$, and Δ is different from zero ($|v^x| < 1$).

4. A 1D numerical simulation

There are several 1D benchmark relativistic Riemann problems used to test the performance of a HRSC method. We can list the following 1D relativistic Riemann problems:

- Relativistic blast waves, (see (Norman and Winkler, 1986), (Donat et al., 1998), (Martí and Mueller, 1995), (Martí and Mueller, 1999)).
- Relativistic colliding slabs and relativistic shock heating, (see (Aloy et al., 1999), (Donat et al., 1998), (Dolezal et al., 1995), (Schneider et al., 1993), (Martí and Mueller, 1995), (Martí and Mueller, 1999)).
- Collision of relativistic blast waves, (see (Martí and Mueller, 1995), (Martí and Mueller, 1999)).
- Low density Riemann problems, (see (Aloy, Martí, Marquina, 1999)).

We concentrate here on the “relativistic colliding slabs” Riemann problem. The initial set-up is that of a relativistic shock tube test (see (Donat et al., 1998) for details) for an ideal gas, $p = (\Gamma - 1)\rho\epsilon$, with $\Gamma = 5/3$, and the following conditions at $t = 0$:

$$\{\epsilon_L = 10^{-4}, \rho_L = 1, v_L = .9995\} \quad \{\epsilon_R = 10^{-4}, \rho_R = 1, v_R = -.9995\}$$

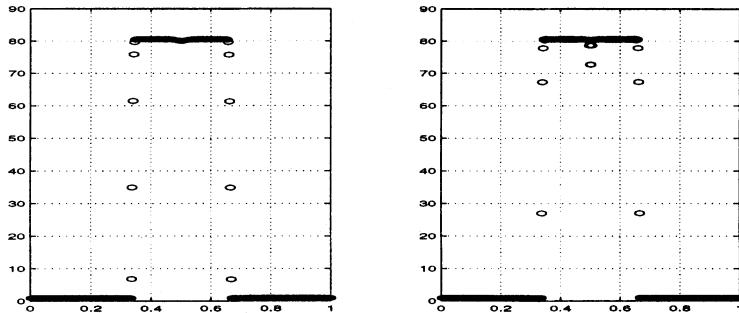


Figure 1. density plots: Marquina-ENO3 left, Shu-Osher-ENO3 right.

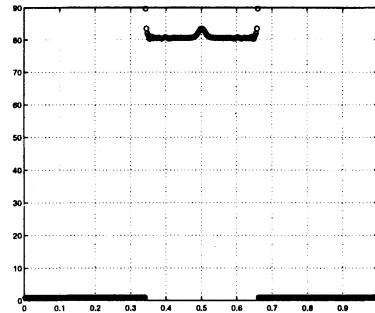


Figure 2. density plot: HLLE-ENO3

In the rigid-wall collision of the two slabs of cold gas, a numerical pathology known as ‘overheating’ ((Noh, 1987), (Donat et al., 1998; Donat, Marquina, 1996) and references therein) occurs on most shock capturing schemes without a heat conduction mechanism. These numerical experiments were taken from (Donat et al., 1999).

For our numerical simulations we use a grid of 400 equally spaced points in $[0, 1]$ and $\Delta t / \Delta x = .2$. In figures 1 and 2 we show numerical approximations to the density obtained with a Shu-Osher code, Marquina’s scheme and the relativistic extension of the HLLE scheme developed in (Schneider et al., 1993). The spatial reconstruction procedure is ENO third order polynomial (see (Shu and Osher, 1989)), and the time discretization procedure is the third order TVD Runge-Kutta method of (Shu and Osher, 1989). We see that Marquina’s scheme is able to dissipate adequately the initial overheating. It is worth mentioning the spurious overshoots obtained by the HLLE scheme, implemented with the procedure explained by Schneider et al., (see (Schneider et al., 1993)). Observe that the velocities of the gas slabs are extremely close to the speed of light. The Lorentz factor is $W \approx 100$, well into the ultrarelativistic regime, and the jumps of the state variables at the

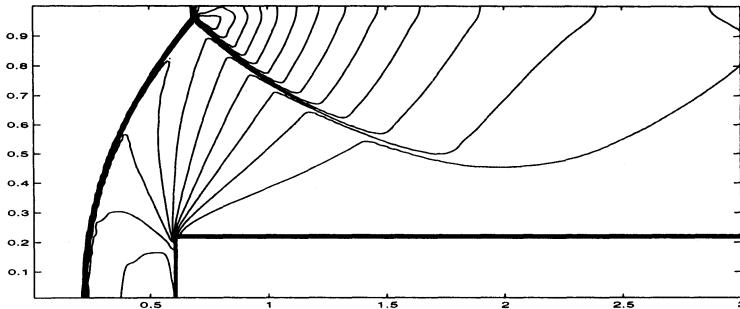


Figure 3. density plot: Marquina's scheme (PHM), $t=3.0$

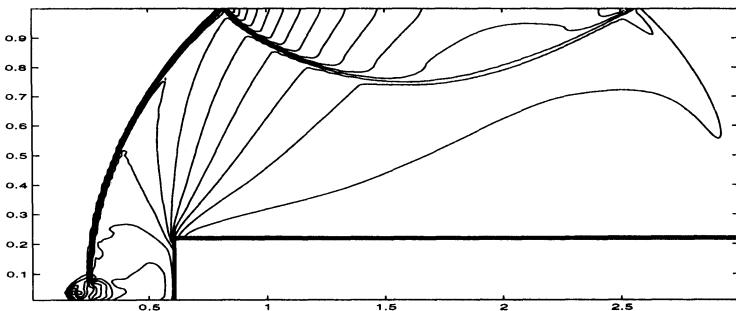


Figure 4. density plot: Roe-Fix Shu-Osher (PHM), $t=2.4$

shocks are very large; in particular, the pressure jump (not shown) is of 10 orders of magnitude and the density jump is a factor 20 greater than its classical value.

A work in progress to reduce or eliminate the overheating problem, is to apply the isobaric-fix boundary condition, introduced in (Fedkiw, Marquina, Merriman, 1999). We are also working on the extension of the ghost fluid method, introduced in (Fedkiw et al., 1999) for gas dynamics, to relativistic flow, (see (Aloy, Martí, Marquina, 1999)).

5. 2D numerical simulations

In (Donat et al., 1998) we show a numerical simulation of a relativistic extension of Emery's step test, a benchmark test in Newtonian hydrodynamics (Woodward and Colella, 1984). We refer to (Donat et al., 1998) for details on the initial set-up of the test. Here we shall concentrate on an interesting phenomenon observed when the grid is refined. When using the standard Roe-fix Shu-Osher ENO schemes, and a resolution of 240×80 cells, a small protuberance starts to form at the base of the bow shock. The protuberance grows and causes the code to crash. The numerical pathol-

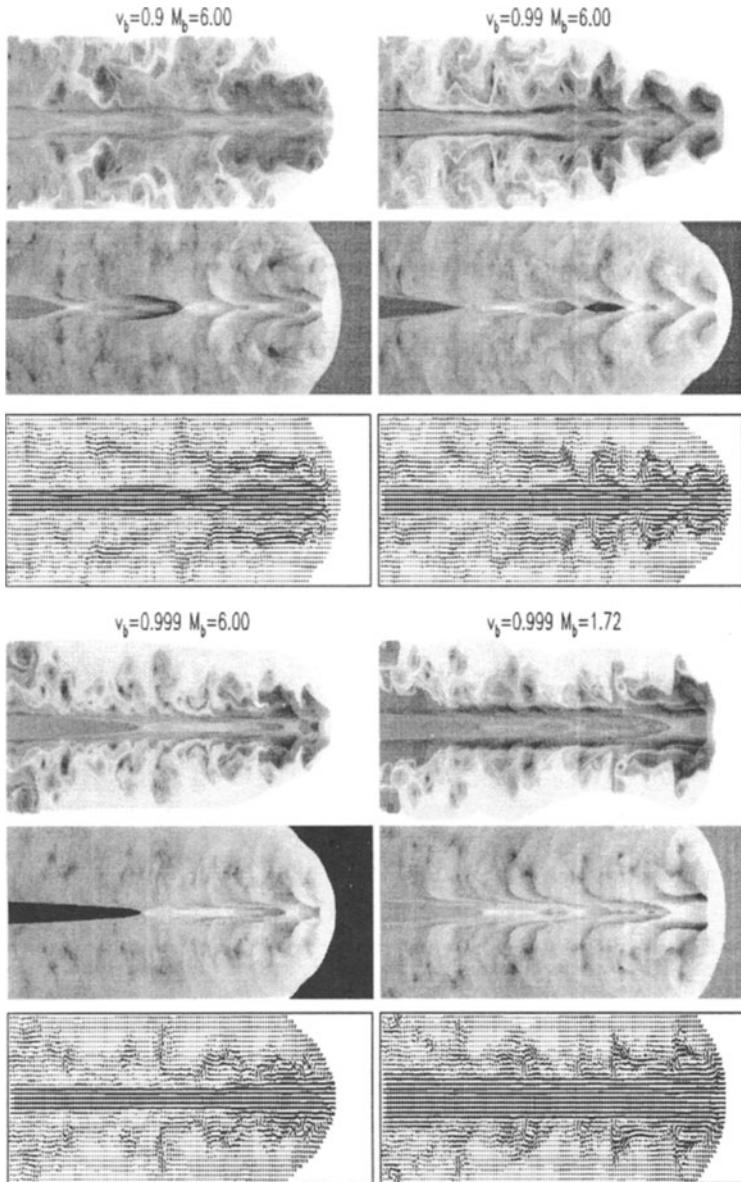
ogy can be observed in the density plot of figure 4. On the other hand, when using Marquina's flux-split algorithm, the behavior of the numerical approximation is consistent with the physics of the problem and we observe no numerical pathologies. In figure 3, we show the numerical approximation obtained with Marquina's scheme and the PHM reconstruction procedure, at time $t=3.0$; at this time the equivalent Roe-fix Shu-Osher code of figure 4 has already crashed. The initial velocity is $v_0^x = .9995$, that is $W \approx 100$.

As a 2D application we have simulated the evolution of several axisymmetric relativistic fluids injected supersonically into the computational domain through a small nozzle. This simple initial setup allows for the study of the morphology and dynamics of *relativistic jets* encountered in some astrophysical scenarios. A comprehensive study can be found in (Martí *et al.*, 1997), where Marquina's solver was used together with the piecewise-parabolic reconstruction procedure of Woodward and Colella, adapted to the relativistic equations by Martí and Müller in (Martí and Mueller, 1995).

We use cylindrical coordinates (r, z) to discretize the numerical domain, which is 50 units long in the z -direction and 7 units wide in the r -direction. The domain is covered by a uniform numerical grid consisting of 1000×140 zones. The beam fluid is injected into the grid parallel to the symmetry axis (the z axis) through a nozzle at the bottom ($r = 0$) of the left boundary of the grid ($z = 0$), which is 20 zones wide (i.e. of length unity). Outflow boundary conditions are used at all boundaries except at the symmetry axis ($r = 0$ boundary) where reflection conditions are imposed, and at the nozzle, where fixed inflow beam conditions are used. The initial model that we consider for the injected beam fluid corresponds to an ideal gas equation of state with $\Gamma = 5/3$. The density ratio between the beam gas over the external medium is equal to 0.01 and the component v_z of the velocity of the fluid at the nozzle is a free parameter, ($v^r = 0$). In Figure 5 the logarithm of rest mass density, the logarithm of the pressure and the velocity field for different inflow velocities, ($v_z = 0.9, v_z = 0.99$ and $v_z = 0.999$), are represented.

6. Computational Remarks

The flux-split formula presented here is not merely a Roe-type scheme with Local-Lax Friedrichs entropy fix, but it is a nonlinear way to dosify viscosities, determined by the sonic points encountered in sided local characteristic variables. This makes the method formally less viscous than HLLE, although it has a higher computational cost. The computational cost of Marquina's flux-split method is nearly double than any standard Roe-type method, but it is more parallelizable than any other, as it is shown in the GENESIS 3D code developed by Aloy *et al.* in (Aloy *et al.*, 1999). An efficient imple-



mentation of Marquina's flux formula for relativistic flow can be found in (Aloy, Pons, Ibáñez, 1999). On the other hand, multiresolution techniques are being developed in order to save time in the computation of numerical fluxes, (see (Chiavassa, Donat, 1999)).

Acknowledgements This work has been supported by the Spanish DG-ICYT (grant PB97-1402) and ARPA URI Grant ONR-N00014-92-J-1890. I thank to Prof. Martí for provide me the runnings of the jet simulation appearing in this paper and for many helpful and intriguing discussions about the morphology and dynamics of astrophysical jets. I also thank to Prof. Rosa Donat and Prof. J.M. Ibáñez for many helpful remarks and criticism. Finally, I would like to thank to Ron Fedkiw for many interesting discussions when visiting UCLA Department of Mathematics.

References

- Aloy M A, Ibáñez J M^a, Martí J M^a and Mueller E, GENESIS: A High-Resolution Code for 3D Relativistic Hydrodynamics, *ApJ*, **122**, pp 151-166 (1999).
- Aloy M A, Martí J M^a and Marquina A, The Ghost Fluid Method for Relativistic Flow, (1999) Preprint
- Aloy M A, Pons J A, Ibáñez J M^a, An efficient implementation of flux formulae in multi-dimensional relativistic hydrodynamical codes, *Computer Physics Communications*, **120**, pp. 115-121, (1999).
- Anile A M, Relativistic Fluids and Magnetofluids, (Cambridge University Press, Cambridge, UK, 1989).
- D. Balsara, Riemann Solver for Relativistic Hydrodynamics *J. Comput. Phys.*, **114**, 284 (1994).
- Chiavassa G and Donat R, Numerical Experiments with Multilevel Schemes for Conservation Laws, Preprint, (1999).
- Dai W and Woodward P R, An Iterative Riemann Solver for Relativistic Hydrodynamics, *SIAM J. Sci. Stat. Comput.*, **18**, pp 982-995 (1997)
- Dolezal A and Wong S S M Relativistic Hydrodynamics and Essentially Non-Oscillatory Shock Capturing Schemes *J. Comput. Phys.*, **120**, 266-277 (1995).
- Donat R, Font J A, Ibáñez J M, Marquina A, A Flux-Split Algorithm for Relativistic Flows *J. Comput. Phys.*, **146**, 58-81, (1998)
- Donat R and Marquina A, Capturing Shock Reflections: An Improved Flux Formula *J. Comput. Phys.*, **125**, 42 (1996).
- Donat R and Marquina A, Computing Strong Shocks in Ultrarelativistic Flows: A Robust Alternative, Hyperbolic Problems, Theory, Numerics, Applications, Seventh International Conference in Zuerich, Vol I. International Series of Numerical Mathematics, Birkhauser Verlag (Basel) **129**, 243-251, (1999).
- Duncan G C and Hughes P A, Simulations of Relativistic Extragalactic Jets, *Astrophys. J.*, **436**, L119 (1994).
- Eulderink F, Numerical Relativistic Hydrodynamics, Ph.D. Thesis, University of Leiden (1993).
- Eulderink F and Mellema G, General Relativistic Hydrodynamics with a Roe Solver, *Astron. Astrophys. Suppl.* **110**, pp 587-623 (1995).
- Falle, S A E G, Komissarov, S S An Upwind Numerical Scheme for Relativistic Hydrodynamics with a General Equation of State, *Mon. Not. Roy. Astronom. Soc.* **278**, 586-602 (1996)
- Fedkiw R, Merriman B, Donat R, Osher S J, UCLA CAM Report, vol. 96-18,(1996).
- Fedkiw R, Aslam T, Merriman B, Osher S J, UCLA CAM Report, vol. 98-17,(1998).
- Fedkiw R, Marquina A, Merriman B, An Isobaric Fix for the Overheating Problem in Multimaterial Compressible Flows, *J. Comput. Phys.*, **148**, 545-578, (1999).
- Font J A, Ibáñez J M^a, Marquina A and Martí J M^a, Multidimensional Relativistic Hydrodynamics: Characteristic Fields and Modern High-Resolution Shock-Capturing Schemes, *Astron. Astrophys.*, **282**, 304-314, (1994).

- Font J A, Miller M, Suen W and Tobias M, Three Dimensional Numerical General Relativistic Hydrodynamics I: Formulations, Methods, and Code Tests. *Phys. Rev. D*, **61**, pp 044011.1 - 044011.26., (2000).
- Ibáñez J M^a, Font J A, Martí J M^a and Miralles J A, Proceedings from the 18th. Texas Sym. on Relativistic Astrophysics, (World Scientific Press, 1997).
- Font J A, Ibáñez J M^a and Papadopoulos P, *Astrophys. J.* **507**, L67,(1998).
- Font J A, Ibáñez J M^a and Papadopoulos P, *Mon. Not. Roy. Astronom. Soc.*in press(1999); astro-ph/9810344.
- LeVeque R J, Numerical Methods for Conservation Laws, Birkhäuser (1991).
- LeVeque R J, in *Computational Methods for Astrophysical fluid flow*, Saas-Fee Advanced Course, **27** Eds. O. Steiner and A. Gautshy, pp. 1-159, Springer-Verlag, (Berlin),(1998).
- Marquina A, Martí J M^a, Ibáñez J M^a, Miralles J A and Donat R, *Astron. Astrophys.*, **258**, 566 (1992).
- Marquina A, Local Piecewise Hyperbolic Reconstruction of Numerical Fluxes for Non-linear Scalar Conservation Laws, *SIAM J. Scient. Comp.*, **15**, pp 892-915, (1994).
- Martí J M^a, Ibáñez J M^a and Miralles J A, Numerical Relativistic Hydrodynamics: Local Characteristic Approach, *Phys. Rev.*, **D43**, 3794 (1991).
- Martí J M^a and Mueller E, Extension of the Piecewise Parabolic Method to One-dimensional Relativistic Hydrodynamics, *J. Comput. Phys.*, **123**, p 1-14, (1996).
- Martí J M^a and Mueller E, Numerical Hydrodynamics in Special Relativity, *Living Reviews in Relativity*, Vol. **2**, (1999), (to appear).
- Martí J M^a, Mueller E, Font J A and Ibáñez J M^a, *Astrophys. J.*, **448**, L105 (1995).
- Martí J M^a, Mueller E, Font J A, Ibáñez J M^a and Marquina A, Morphology and Dynamics of Relativistic Jets, *Astrophys. J.*, **479** 151-163 (1997)
- May M A , and White R H (1967). Stellar Dynamics and Gravitational Collapse. *Math. Comp. Phys.*, **7**, pp 219 - 258.
- Noh, W F, Errors for Calculations of Strong Shocks Using an Artificial Viscosity and an Artificial Heat Flux, *J. Comp. Phys.*, **72**, pp 78-120 (1987).
- Norman M L and Winkler K-H A, Why Ultrarelativistic Hydrodynamics is difficult. *Astrophysical Radiation Hydrodynamics*, pp 449-476, ed. by M.L. Norman and K-H.A. Winkler (Reidel, 1986).
- Pons J.A., Font J.A., Ibáñez J.M^a., Martí J.M^a., and Miralles J.A. (1998). General Relativistic Hydrodynamics with Special Relativistic Riemann Solvers. *A&A*, **339**, pp 638 - 642.
- Quirk J, A Contribution to the Great Riemann Solver Debate, *Intl. J. Numer. Meth. Fluids*, **18**, 555-574 (1994).
- Roe P L, Approximate Riemann Solvers, Parameter Vectors and Difference Schemes, *J. Comput. Phys.*, **43** 357-372, (1981).
- Schneider V, Katscher V, Rischke D H, Waldhauser B, Marhun J A and Munz C D, New Algorithms for Ultra-relativistic Numerical Hydrodynamics, *J. Comput. Phys.*, **105**, 92-107, (1993).
- Shu C W and Osher S J, Efficient Implementation of Essentially Non-Oscillatory Shock-Capturing Schemes, II, *J. Comput. Phys.*, **83**, pp 32-78 (1989).
- Steger J and Warming R F, Flux Vector Splitting of the Inviscid Gasdynamics Equations with Application to Finite Difference Methods, *J. Comput. Phys.*, v. **40**, pp 263-293, (1981).
- Spencer R E and Newell S J, *Vistas in Astronomy*, **41**, 1 (1997).
- Clare R B and Strottman D, *Phys. Reports*, **141**, 177 (1986).
- Strottman D, in *The Nuclear Equation of State, Part B*,, ed. by Greiner W and Stöcker H, Plenum Press (1989).
- Van Leer B, Towards the Ultimate Difference Scheme V. A Second Order Sequel to Godunov's Method, *J. Comput. Phys.*, **32**, pp 101-136, (1979).
- Wen L, Panaiteescu A and Laguna P, A shock-patching Code for Ultra-relativistic Fluid Flows, *Astrophys. J.*, **486**, p 919-927, (1997).

- Wilson J R, *Astrophys. J.*, **163**, 209, (1971).
- Wilson J R, Numerical Study of Fluid Flow in a Kerr Space, *Astrophys. J.*, **173**, 431-438, (1972).
- Woodward P R and Colella P, The Numerical Simulation of Two-Dimensional Fluid Flow with Strong Shocks, *J. Comput. Phys.*, **54** pp 115-173, (1984).
- Yee H C, *VKI Lecture Notes in Computational Fluid Dynamics*, von Karman Institute for Fluid Dynamics (1989).

AN ARTIFICIAL COMPRESSION PROCEDURE VIA FLUX CORRECTION

V. MARTÍNEZ

*Departament de Matemàtiques,
Universitat Jaume I,
Campus de Riu Sec,
Castelló 12071, Spain.*
Email: martinez@mat.uji.es

Abstract. We present a new procedure to sharpen contact discontinuities. Our procedure corrects the linear flux to obtain a consistent dynamical behaviour. The equivalence between the original equation and the modified equation is proved.

1. Introduction

It is well known that the contact discontinuities in the linear advection equation are usually smeared more severely than shocks when a standard numerical scheme is used. In order to sharpen these discontinuities some remarkable results have been obtained. Harten (Harten, 1989) introduced the concept of subcell resolution. Afterwards, this idea was used by Yang (Yang, 1990) and Mao (see (Mao, 1991) and (Mao, 1992)). Yang uses an artificial compression method based on a modification of the slopes in the ENO reconstruction. On the other hand, Mao uses an extrapolation technique.

When we consider

$$u_t + au_x = 0, \quad (1)$$

$$u_t + buu_x = 0, \quad (2)$$

we observe that the exact solution of both equations with $u_- > u_+$, $b = (2a)/(u_- + u_+)$ and the initial data:

$$u_0(x) = \begin{cases} u_-, & x < x_0, \\ u_+, & x > x_0, \end{cases} \quad (3)$$

is the same function: $u(x, t) = u_0(x - at)$. Actually, the characteristic curves for equation (1) are parallel straight lines, therefore we have a contact discontinuity. When standard numerical schemes are used, this discontinuity is usually smeared. However the characteristic curves for equation (2) are intersecting straight lines, therefore a shock is obtained. This shock is propagated with the same speed as the linear discontinuity obtained in equation (1), but the numerical approximation of the shock is less smeared than the other one.

In this paper, we have proposed an artificial compression procedure based on a flux modification. In order to obtain a better numerical approximation of the jump, our idea is to replace the linear flux in equation (1) by a nonlinear flux as in equation (2), so that the original solution is conserved. The choice of a suitable nonlinear flux is raised in section 2.

2. Travelling wave analysis

Given the scalar conservation law

$$u_t + f(u)_x = 0, \quad u(x, 0) = u_0(x), \quad (4)$$

for $(x, t) \in \mathbb{R} \times (0, T)$ with $T > 0$ and where $f, u_0 \in C^1(\mathbb{R})$ are supposed to be piecewise-smooth functions that are either periodic or of compact support. The initial data $u_0(x)$ are given by (3).

It is natural to consider the following viscosity equation

$$u_t + f(u)_x = \varepsilon u_{xx} \quad (5)$$

for a small $\varepsilon > 0$. Next, steady-state solutions to (5) are obtained and then, they are compared with the solution of (4). Similar studies have been achieved by E. Harabetian (Harabetian, 1992) and by J. Smoller (Smoller, 1994).

If p is a discontinuity point of the solution $x = x(t)$ for (4) in the (x, t) -plane and $s = x'(t_p)$, then, we may assume that (5) has solutions, called **travelling wave solutions** of the form: $u(x, t) = \varphi(\xi)$, $\xi = \frac{x - st}{\varepsilon}$, satisfying

$$\lim_{\xi \rightarrow -\infty} \varphi(\xi) = u_-, \quad \lim_{\xi \rightarrow +\infty} \varphi(\xi) = u_+, \quad \lim_{\xi \rightarrow -\infty} \varphi'(\xi) = 0 \text{ and } \lim_{\xi \rightarrow +\infty} \varphi'(\xi) = 0,$$

since φ is a solution of (4), we have $\varphi(\xi)_t + f(\varphi(\xi))_x = \varepsilon \varphi(\xi)_{xx}$, so that

$$\varphi'(\xi)(-\frac{s}{\varepsilon}) + f'(\varphi(\xi))\varphi'(\xi)\frac{1}{\varepsilon} = \varepsilon \left(\frac{1}{\varepsilon^2} \varphi''(\xi) \right),$$

multiplying by ε we obtain

$$\varphi'' = (f'(\varphi) - s)\varphi'. \quad (6)$$

It is easy to verify that the phase plane of equation (6) when $f(u) = au$ does not have trajectories (see (Smoller, 1994)). Consequently, (4) does not allow travelling wave solutions. On the contrary, if $f(u) = \frac{1}{2}bu^2$ as in equation (2), then, (4) has travelling wave solutions and the phase plane of (6) has trajectories.

The previous argument allows us to assume that a good representation of the jump may be obtained by replacing the linear flux of (1) by a non-linear flux that allows a quadratic trajectory in the phase plane of equation (6), such as $\psi(\varphi) = \rho(\varphi - u_-)(\varphi - u_+)$, where ρ (see Fig. 1) is a parameter verifying the following two conditions:

(C1) Consistency with the dynamical behaviour of the jump:

$$\rho > 0, \text{ if } u_- > u_+ \text{ and } \rho < 0, \text{ if } u_- < u_+.$$

(C2) The CFL restriction in the numerical scheme used is preserved.

The following result justifies the previous choice of ψ :

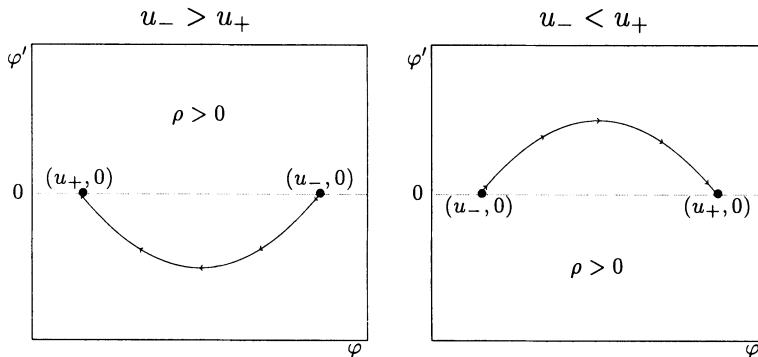


Figure 1. Trajectories of (6)-(7) for different values of u_- and u_+ .

Lemma 1 If we consider the flux

$$f(u) = au + \psi(u), \quad (7)$$

then the phase plane of (6) has a trajectory connecting $(u_-, 0)$ and $(u_+, 0)$.

Proof: From (6), we have $\varphi'' = \psi'(\varphi)\varphi'$, if we consider $y_1 = \varphi$ and $y_2 = \varphi'$, we obtain the ordinary differential system

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}' = \begin{pmatrix} 0 \\ \psi'(y_1)y_2 \end{pmatrix},$$

whose solution is $y_2 = \psi(y_1) + E$, where E is a constant.

Finally, if $E = 0$ we obtain the trajectory which connects the equilibria $(u_-, 0)$ and $(u_+, 0)$.

3. Flux Correction Procedure

The procedure we proposed before consists of the replacement of the problem (1)-(3) by the problem (3)-(4)-(7), but we need to check some results to ensure that the solution of the original problem is the same.

Lemma 2 *Problem (3)-(4)-(7) propagates the shock with speed a .*

The proof is trivial.

Theorem 1 *Problem (3)-(4)-(7) has the same solution as (1)-(3).*

Proof: We assume that u_1 is the solution of (1)-(3) and that u_2 is the solution of (3)-(4)-(7), therefore $\forall (x, t) \in \mathbb{R} \times (0, T)$ we have

$$u_1(x, t) = \begin{cases} u_-, & x - at < x_0, \\ u_+, & x - at > x_0 \end{cases} \quad \text{and}$$

$$u_2(x, t) = \begin{cases} u_-, & x - (a + \rho(u_- - u_+))t < x_0, \\ u_+, & x - (a + \rho(u_+ - u_-))t > x_0. \end{cases}$$

If $u_1 \neq u_2$, then, there is an element $(x_1, t_1) \in \mathbb{R} \times (0, T)$ such that

$$x_1 - at_1 - \rho(u_- - u_+)t_1 < x_0, \quad u_2(x_1, t_1) = u_-$$

$$\text{and} \quad x_0 < x_1 - at_1 < x_0 + \rho(u_- - u_+)t_1, \quad u_1(x_1, t_1) = u_+,$$

which contradicts the previous Lemma. Therefore, $u_1 = u_2$.

Next, we will justify how to devise an efficient implementation of the preceding recipe: to solve the equation (4) using standard finite-difference methods, we assume that u is a function of two independent variables. The (x, t) -plane is subdivided into grid lines parallel to the x - and t -axis with sides $\Delta x = h$ and $\Delta t = k$. Define $x_j = jh$, $j = \dots, -1, 0, 1, \dots$ and $t_n = nk$, $n = 0, 1, 2, \dots$. We denote the numerical approximation of u at the mesh point (x_j, t_n) by $u_j^n = u(x_j, t_n)$. In each computational cell $[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}] \times [t_n, t_{n+1}]$, we consider the numerical flux function $f_{j+\frac{1}{2}}^n = \tilde{f}(u_{j-m+1}^n, \dots, u_{j+m}^n)$ as a function of $2m$ variables, which is consistent with (4), i.e. $\tilde{f}(u, \dots, u) = f(u)$, then we have the numerical scheme in conservation form:

$$u_j^{n+1} = u_j^n - \lambda(f_{j+\frac{1}{2}}^n - f_{j-\frac{1}{2}}^n). \quad (8)$$

under a suitable CFL restriction

$$\lambda \cdot \max_u |f'(u)| \leq \lambda_0, \quad (9)$$

where $\lambda = \frac{k}{h}$ and λ_0 depends on the numerical scheme used.

Lemma 3 If

$$|\rho| \leq \frac{\lambda_0 - |a| \lambda}{\lambda |u_- - u_+|}, \quad (10)$$

then the CFL restriction (9) is satisfied.

Proof: Since $f'(u) = (2\rho)u + (a - \rho u_- - \rho u_+)$, the triangular inequality gives the result.

The following algorithm is proposed to sharpen discontinuities:

ALGORITHM 3.1

Step 1: Take λ such that $\lambda_0 - |a| \lambda > 0$.

Outside jumps:

Step 2 Progress in time by equation (8) with $f(u) = au$.

For each jump:

Step 2 Locate the cells affected by the jump: u_l^n, \dots, u_{l+q}^n .

Step 3 Take ρ such that (C1) and (10) ($u_- = u_l^n$, $u_+ = u_{l+q}^n$) are satisfied.

Step 4 For $j = l, \dots, l + q$ progress in time by equation (8) with

$$f(u) = au + \rho(u - u_l^n)(u - u_{l+q}^n).$$

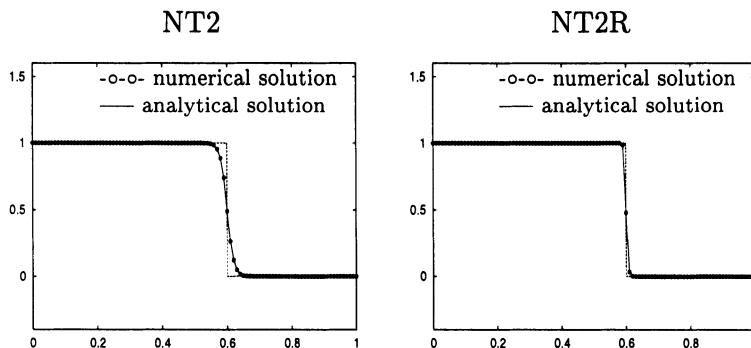


Figure 2. Comparison of NT2 and NT2R on problem 1.

4. Numerical Experiments

To study the behaviour of our procedure, we consider the Lax-Friedrichs first order numerical flux (LAXF1). Let LAXF1R denote the use of LAXF1

together with the ALGORITHM 3.1. We will also study the effects that the previous recipe produces when a high order method is used. Here we will consider (see (Nessyahu and Tadmor, 1990)) the method of Nessyahu-Tadmor (NT2). Similarly, let NT2R denote the use of NT2 together with ALGORITHM 3.1.

Furthermore, the CFL restriction (9) for LAXF1 is satisfied when $\lambda_0 = 1$ (see (Smoller, 1994)) so one may use $\lambda_0 \approx 0.5$ for NT2 (see (Nessyahu and Tadmor, 1990)).

We consider the following test problems:

$$\text{Problem 1: } u_t + u_x = 0 \text{ with } u_0(x) = \begin{cases} 1, & x < 0.3, \\ 0, & x > 0.3. \end{cases}$$

$$\text{Problem 2: } u_t + u_x = 0 \text{ with } u_0(x) = \begin{cases} 1, & 0.15 < x < 0.45, \\ 0, & \text{otherwise.} \end{cases}$$

Only one jump appears in problem 1 with $u_- > u_+$ and the movement of discontinuity has speed $a = 1$. On the other hand, two jumps appear in problem 2, the first one with $u_- > u_+$ and the second one with $u_- < u_+$. The movement of both jumps has speed $a = 1$.

In both problems we have reduced the study to the interval $(0, 1)$, we have taken $T = 0.3$ and 100 mesh points are used. For LAXF1 and LAXF1R we have taken $\lambda = 0.75$ and for NT2 and NT2R, $\lambda = 0.375$.

Numerical results with NT2 and NT2R (see Fig.2 and Fig.3) show that the use of the ALGORITHM 3.1 improves the numerical solutions getting correct speed propagation of discontinuities and producing sharper profiles for contact discontinuities. Numerical results with LAXF1 and LAXF1R are similar, but they are more smeared.

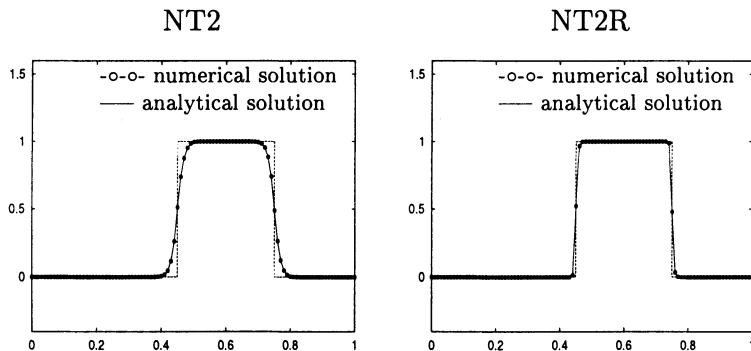


Figure 3. Comparison of NT2 and NT2R on problem 2.

5. Concluding Remarks

We present an artificial compression method, which is a new point of view for sharpening contact discontinuities. Our procedure replaces contact discontinuities by shocks. This recipe is based on the replacement of the linear flux by a nonlinear flux with a consistent dynamical behaviour. The method has been tested on some problems with discontinuous solutions and has shown to yield a good approximation. Recently, we have used initial data with several jumps (of sizes different to 0 and 1), and we have obtained results with the same quality than those shown in Fig.2 and Fig.3.

Research in order to extend this technique to systems with linearly degenerate fields is in progress.

Acknowledgements. Continuous support from DGES(MEC), project PB97-0397, and from the Universitat Jaume I Fundació Caixa Castelló, project P1B-97-23 are acknowledged.

References

- Harabetian E (1992). A Subcell Resolution Method for Viscous Systems of Conservation Laws. *J. Comp. Phys.* **103**, pp 350-358.
- Harten A (1989). ENO Schemes with Subcell Resolution. *J. Comp. Phys.* **83**, pp 148-184.
- Mao D K (1991). A Treatment of Discontinuities in Shock-capturing Finite Difference Methods. *J. Comp. Phys.* **92**, pp 422-455.
- Mao D K (1992). A Treatment of Discontinuities for Finite Difference Methods. *J. Comp. Phys.* **103**, pp 359-369.
- Nessyahu H and Tadmor E (1990). Non-oscillatory Central Differencing for Hyperbolic Conservation Laws. *J. Comp. Phys.* **87**, pp 408-463.
- Smoller J (1994). Shock Waves and Reaction-Diffusion Equations. Springer-Verlag.
- Yang H (1990). An Artificial Compression Method for ENO Schemes: The Slope Modification Method. *J. Comp. Phys.* **89**, pp 125-160.

A SECOND ORDER TIME-SPLITTING TECHNIQUE FOR ADVECTION-DISPERSION EQUATION ON UNSTRUCTURED GRIDS

A. MAZZIA, L. BERGAMASCHI AND M. PUTTI

*Dipartimento di Metodi e Modelli Matematici
per le Scienze Applicate,
Via Belzoni 7, 35131 Padova, Italy.
Emails:mazzia, berga, putti @dmsa.unipd.it*

Abstract. In this paper we present a time-splitting approach for advection-dispersion equations. We split the dispersive and advective fluxes into two separate partial differential equations (PDEs) one containing the dispersive term and the other one the advective term, respectively. On triangular elements we combine a triangle-based high resolution Finite Volume (FV) scheme for advection with a Mixed Hybrid Finite Element (MHFE) technique to solve dispersion. In this work we consider a development of the time-splitting technique to obtain second order accuracy in space and time. This is obtained by means of the combination of the Crank-Nicolson scheme for the MHFE and the explicit midpoint rule for FV and by adding a correction term in the linear reconstruction of the FV discretization of the advective term. Numerical results are used to validate the theory presented.

1. Introduction

The numerical solution of the advection-dispersion partial differential equation (PDE) is usually obtained by means of implicit time-stepping schemes because of their stability properties. Explicit schemes are characterized by strong time step size limitations due to the stability requirements posed by the discretization of the dispersive terms. When high resolution Godunov-type Finite Volume schemes are used for the spatial discretization of the advective term in combination with implicit time stepping, the intrinsically nonlinear character of these methods yields a system of nonlinear equations

to be solved at each time step. This occurs even if the original PDE is linear and has a strong influence on the performance of the overall approach in terms of both accuracy and computation time. We propose a time-splitting approach for the advection-dispersion equation that tries to overcome this problem by solving implicitly the diffusion term and explicitly the convective term. The only limitation to the time step size is the not overly restrictive CFL-type stability condition of the FV technique.

Our attention is focused on the class of methods known as fully Eulerian Godunov-Mixed Methods (GMM) (Dawson, 1991; Dawson, 1993; Dawson, 1995). They originate from the splitting of the dispersion and advection fluxes into two separate partial differential equations (PDEs) containing the dispersive and advective terms, respectively. A time-explicit, spatially second-order Godunov method is used to discretize advection, and a time-implicit, spatially second order Mixed Finite Element method is used for modeling dispersion. Using Euler (or 2nd order Runge-Kutta) time stepping, the advective term is discretized by a triangle-based high resolution Finite Volume (FV) scheme (Durlofsky, Engquist and Osher, 1992; Liu, 1993), while the dispersive flux is discretized using a Mixed Hybrid Finite Element (MHFE) technique. The choice of these two schemes is dictated on the one hand by their accuracy, robustness and efficiency in handling nonuniform meshes and highly variable coefficients. On the other hand, both FV and MHFE are based on the weak formulation of the governing equation and use similar functional spaces for the approximation of the dependent variable, making them ideally suited for combination in a time-splitting approach. The scheme can be shown to be second order accurate in space (away from sharp fronts) and time (Mazzia, Bergamaschi and Putti, 2000a; Mazzia, Bergamaschi and Putti, 2000b) if a corrected version of Runge scheme is used for the advective term. In this work we investigate the numerical behavior of the proposed time-splitting technique. The theoretical results of the proposed scheme are verified on one-dimensional test problems characterized by different mesh Peclet numbers.

2. The numerical scheme

Subsurface contaminant transport is governed by an advection-diffusion equation of the form

$$\begin{aligned} \frac{\partial \phi c}{\partial t} + \vec{\nabla} \cdot (\vec{v}c - D\vec{\nabla}c) &= f && \text{on } \Omega \times (0, T], \\ c &= c^0 && \text{on } \Omega \times 0, \\ c &= b_D && \text{on } \Gamma_D \times (0, T], \\ -D\vec{\nabla}c \cdot \vec{n} &= b_N && \text{on } \Gamma_N \times (0, T] \\ (\vec{v}c - D\vec{\nabla}c) \cdot \vec{n} &= b_C && \text{on } \Gamma_C \times (0, T] \end{aligned} \tag{1}$$

where c is the concentration of the solute, ϕ is the porosity of the medium, \vec{v} is Darcy's velocity, $D = D(\vec{v})$ is the tensor accounting for mechanical dispersion and molecular diffusion, and f is a source or sink term. Denoting by \vec{F} and \vec{G} the advective and dispersive flux, respectively, equation (1) may be written as:

$$\begin{aligned} \frac{\partial \phi c}{\partial t} + \vec{\nabla} \cdot (\vec{F} + \vec{G}) &= f && \text{on } \Omega \times (0, T] \\ \vec{F} &= \vec{v}c \\ \vec{G} &= -D \vec{\nabla} c \end{aligned} \quad (2)$$

As the geometry of the physical domain Ω is often complex when dealing with real world applications, we choose to work with unstructured meshes, and thus Ω is discretized into m triangles, T_l , $l = 1, \dots, m$. Concentration c can be approximated by $c \simeq \sum_{l=1}^m c_l \psi_l$, where ψ_l are $P_0(T_l)$ scalar basis functions, taking on the value one on triangle T_l and zero elsewhere. Multiplying equation (2) by ψ_l and integrating in space and time, with time-step Δt over the time interval $[t^k, t^{k+1}]$, the following semidiscrete equations are obtained:

$$\phi_l^{k+1} c_l^{k+1} = \phi_l^k c_l^k - \frac{\Delta t}{|T_l|} \int_{T_l} [\vec{\nabla} \cdot (\vec{F}(c^{k+1-\theta}) + \vec{G}(c^{k+\theta})) - f^{k+\theta}] \, d\Delta \quad (3)$$

$$l = 1, \dots, m$$

where c_l^k is the volume average over T_l defined by

$$c_l^k = \frac{\int_{T_l} c(\cdot, t^k) \, d\Delta}{|T_l|}, \quad (4)$$

$|T_l|$ is the area of T_l , ϕ is considered constant within each triangle, and a weighted scheme is used for the time quadrature with weighting parameter $\theta \in \{0.5, 1\}$ and $c^{k+\theta} = \theta c(\cdot, t^{k+1}) + (1 - \theta) c(\cdot, t^k) = \theta c^{k+1} + (1 - \theta) c^k$. Denoting by L_d the spatial discretization operator for dispersion, where we also include the source term f , and by L_a the spatial discretization operator for advection, the time-splitting technique can be described by the following algorithm. For each time step, we consider the advection and the dispersion step, respectively. In the advection step, for each T_l we solve, with the explicit FV scheme,

$$\phi c_l^{k+1} = \phi c_l^k + \Delta t [L_a(c^{k+1-\theta})] \quad (5)$$

determining the predictor concentration \hat{c}_l^{k+1} . In the dispersion step, for each T_l we solve with implicit MHFE method using \hat{c}_l^{k+1} as initial condition

$$\phi^{k+1} c_l^{k+1} = \phi^k \hat{c}_l^{k+1} + \Delta t [L_d(c_l^{k+\theta})] \quad (6)$$

obtaining the final approximation c_i^{k+1} . The FV scheme is stable if $\text{CFL} < 1/3$, while MHFE, because of its implicit nature, is not subject to stability restrictions. Spatial second order accuracy is obtained by the MHFE at the centroids of the triangles (Douglas and Roberts, 1985), while the FV scheme - as developed by (Durlofsky, Engquist and Osher, 1992) and then modified by (Liu, 1993) - achieves spatial second order accuracy away from sharp fronts by employing linear reconstruction plus slope limiting, combined in such a way as to locally satisfy the maximum principle. In the dispersion step, our implementation of the MHFE applied to the discretization of (6) produces a system of linear equations involving the pressure head, the dispersive flux and the Lagrange multiplier, namely the trace of the pressure at the edges of the triangulation. The Schur complement applied to this set of equations reduces the original system to a smaller positive definite linear system (Brezzi and Fortin, 1991; Mazzia, Bergamaschi and Putti, 2000a), with the Lagrange multipliers as unknowns. Such a system can be conveniently solved by a Preconditioned Conjugate Gradient method.

In time, for $\theta = 1$ we use the implicit Euler scheme for L_d and the explicit Euler for L_a and the accuracy of the overall technique is first order in time and globally second order accurate in space. For $\theta = 0.5$ the second order time accurate midpoint rule is applied in its implicit version to the dispersive and in its explicit version to the advective phase. However, the time-splitting algorithm introduces a term in the FV equation that reduces the time accuracy of the scheme to first order. To correct this behavior, the FV linear reconstruction has to be modified, as we will see in the following.

2.1. SECOND ORDER ACCURACY IN TIME

Setting $\theta = 0.5$ we use the midpoint rule for the FV scheme together with the Crank-Nicolson procedure in the MHFE method. This is not sufficient to obtain second order accuracy in time. The usual midpoint (or Runge) scheme is given by:

$$\begin{aligned} c^1 &= c^k + \frac{\Delta t}{2} E(x, c^k) \\ c^{k+1} &= c^k + \Delta t E(x, c^1) \end{aligned} \quad (7)$$

where E represent the numerical advective flux. Following this scheme, we can loose accuracy in time, because c^1 approximates $c^{k+1/2}$ without considering the dispersion term. In this way, we introduce an error term depending on dispersion, that would destroy second order accuracy in time. Therefore we approximate $c^{k+1/2}$ by Taylor expansion taking into account the influence of the dispersive term on the transport equation (1). Since the values c^k and c^{k-1} obtained by solving (6) are numerical approximations to (1) we can simply approximate the temporal partial derivative by backward

difference. In this way we get

$$c^1 = c^k + \frac{\Delta t}{2} \frac{\partial c^k}{\partial t} = \frac{3}{2}c^k - \frac{1}{2}c^{k-1} \quad (8)$$

Thus, the numerical flux $E(x, c^1)$ is constructed by using the values that approximate $c^{k+1/2}$ taking into account the dispersion term. Following the lines of (Dawson, 1993) we can prove second order accuracy not only in space but also in time.

3. Numerical results

The behavior of the proposed numerical scheme can be characterized as a function of two grid related dimensionless numbers, the Courant-Friedrichs-Lewy (CFL) number and the Peclet (Pe) number. The CFL number can be defined for each triangle T_l as $\text{CFL} = \Delta t \sup(\bar{T}_l / |T_l|) \sup |d\vec{F}/dc|$, where \bar{T}_l denote the perimeter of T_l . Stability of the FV scheme requires that $\text{CFL} \leq 1/3$.

The Peclet number represents the ratio between the advective and the dispersive term and can be defined in our case as $\text{Pe} = \text{CFL}/\gamma$, where the dispersion number γ is given by $\gamma = |D| \Delta t \sup(1/|T_l|)$ and $|D|$ is the norm of tensor D . Low Peclet numbers indicate that dispersion is predominant over advection, and vice versa.

The numerical convergence rate of the time-splitting technique is tested on a one-dimensional model problem solved in a two-dimensional grid system. We consider the partial differential equation describing the movement of a tracer in a semi-infinite column and simulate it on a rectangular domain of unit length, with $\vec{v} = (v, 0)$ and $D = \text{diag}(D_1, D_2)$. The boundary conditions $c = 1$ in $x = 0$ and $c = 0$ for $x = \infty$ are imposed, and zero concentration is used as initial condition. This situation is simulated numerically by employing a grid of unitary length and making sure that at the time at which the relative error is evaluated the solution vanishes naturally at the right boundary. The analytical solution to this problem is (Bear, 1979)

$$c(x, t) = \frac{1}{2} \left(\text{erfc} \frac{x - vt}{2\sqrt{D_1 t}} + \exp \frac{vx}{D_1} \cdot \text{erfc} \frac{x + vt}{2\sqrt{D_1 t}} \right) \quad (9)$$

The numerical convergence behavior of the scheme is evaluated by calculating the L_1 norm errors ($|e_\ell|$) at different grid levels. For all the test runs we consider the solution at $t^k = 0.1$ s.

Three grid levels are used and defined as follows. At the coarsest level ($\ell = 1$, on the square $[0, 1] \times [0, 1]$, characterized by $m = 200$ triangles and 121 edges), the domain is discretized into ten layers of rectangular elements further subdivided into triangles. The refined triangulations ($\ell = 2, 3$) are

TABLE 1. Case with $D = 8 \times 10^{-2}$ m²/s and $v = 5 \times 10^{-2}$ m/s.

ℓ	MP2		MP1		Eu	
	$ e_\ell $	rate	$ e_\ell $	rate	$ e_\ell $	rate
1	7.07e-2		7.50e-2		1.46e-1	
2	2.12e-2	1.74	2.18e-2	1.78	6.41e-2	1.18
3	6.10e-3	1.80	6.93e-3	1.63	3.07e-2	1.06

obtained by connecting the midpoints of the three edges of each triangle. The Pe number decreases by a factor of 2 in passing from a coarser to a finer level. The simulations are aimed at numerically verifying the theoretical convergence rate of the time-splitting technique scheme under different Pe numbers.

We compare the results obtained by setting $\theta = 1$ (Eu) with those obtained with $\theta = 0.5$. In the latter, the time integration scheme is given by Crank-Nicolson for MHFE while, for FV, we propose two different discretization techniques: the midpoint rule taking into account the correction term described in Section 2.1 (MP2) and the midpoint scheme without considering the correction (MP1). The time step $\Delta t = \Delta x/2$, where Δx is the grid step size, varies in the interval $[5 \times 10^{-2}, 1.25 \times 10^{-2}]$. Velocity and dispersion parameters are chosen in such a way to preserve correct boundary conditions and CFL= 0.17.

Table 1 reports the errors and convergence rates at the different levels when dispersion is $D = 8 \times 10^{-2}$ m²/s and velocity is $v = 5 \times 10^{-2}$ m/s. These values correspond to Peclet numbers varying from 0.213 ($\ell = 1$) to 0.053 ($\ell = 3$), indicating that dispersion is dominant. As expected, the Eu scheme is of first order accuracy, while MP1 and MP2 display superlinear convergence. However, MP1 rates decrease from 1.78 to 1.63, while MP2 rates tend to second order accuracy. It is evident that the introduction of the dispersion corrective term is crucial to improve accuracy when the dispersion term is not negligible.

The last table considers the case of $D = 1 \times 10^{-2}$ m²/s, so that Pe=1.71 at the first grid level. For this intermediate value of Pe, advection starts to become dominant and thus the importance of the correction decreases, as can be seen from the relatively small differences in the accuracy of MP1 and MP2. As expected, at larger Pe numbers, the differences between these two schemes decrease further.

TABLE 2. Case with $D = 1 \times 10^{-2}$ m²/s and $v = 5 \times 10^{-2}$ m/s

ℓ	MP2		MP1		Eu	
	$ e_\ell $	rate	$ e_\ell $	rate	$ e_\ell $	rate
1	2.38e-1		2.38e-1		2.69e-1	
2	8.68e-2	1.45	8.78e-2	1.44	1.24e-1	1.12
3	1.97e-2	2.14	2.13e-2	2.04	4.26e-2	1.54

References

- Bear J (1979). Hydraulics of Groundwater. McGraw-Hill, New York.
- Dawson C N (1991). Godunov-mixed methods for advective flow problems in one space dimension. *SIAM J. Num. Anal.* **28**, pp 1282-1309.
- Dawson C N (1993). Godunov-mixed methods for advection-diffusion equations in multidimensions. *SIAM J. Num. Anal.* **30**, pp 1315-1332.
- Brezzi F and Fortin M (1991). Mixed and Hybrid Finite Element Methods. Springer-Verlag, Berlin.
- Dawson C N (1995). High resolution upwind-mixed finite element methods for advection diffusion equations with variable time-stepping. *Num. Meth. PDE* **11**, pp 525-538.
- Douglas J Jr and Roberts J R (1985). Global estimates for mixed methods for second order elliptic equations. *Math. Comp.* **44**, pp 39-52.
- Durlofsky L J, Engquist B and Osher S (1992). Triangle based adaptive stencils for the solution of hyperbolic conservation laws. *J. Comp. Phys.* **98**, pp 64-73.
- Liu X D (1993). A maximum principle satisfying modification of triangle based adaptive stencils for the solution of scalar hyperbolic conservation laws. *SIAM J. Num. Anal.* **30**, pp 701-716.
- Mazzia A, Bergamaschi L and Putti M (2000). A time-splitting technique for advection-dispersion equation in groundwater. *J. Comp. Phys.* **157**, pp 181-198.
- Mazzia A, Bergamaschi L and Putti M (2000). Triangular finite volume-mixed finite element discretization for the advection-diffusion equation. Large-Scale Scientific Computations in Engineering and Environmental Problems II, NNFM. Griebel M, Margenov S and Yalamov P (Editors). VIEWEG, Vol. 73, pp 371-378.

TOWARDS IMPLICIT GODUNOV METHOD: EXACT LINEARIZATION OF THE NUMERICAL FLUX

I. MEN'SHOV AND Y. NAKAMURA

*Department of Aerospace Engineering,
Nagoya University,
Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan
Emails: menshov@nuae.nagoya-u.ac.jp
nakamura@nuae.nagoya-u.ac.jp*

Abstract.

The present paper addresses an implicit Godunov method for compressible fluid dynamics, where the exact solution to the Riemann problem (RP) is used to approximate the numerical flux. Solution of discrete equations is achieved in this approach with an iterative Newton method that requires exact linearization of non-linear terms related to the numerical flux-function. To derive this linearization, first variation of the exact solution to the RP caused by a small change in initial values is studied for the compressible Euler equations. By introducing variation matrices associated with the left- and right-hand side states of the surface of initial discontinuity, this variation takes a linear form with regard to variations of initial values, and exact expressions for these matrices are derived in a compact explicit form. These matrices are then employed to linearize the Godunov numerical flux-function. The resulting system of linear algebraic equations in delta form is solved in two steps by implementing an LU-SGS factorization method. A comparative analysis is offered between the present method and the implicit Godunov method implemented with the approximate flux linearization of Turkel and Jameson. The numerical results show that the use of exact linearization in the implicit Godunov method offers a substantial increase in performance over the commonly used Turkel and Jameson's linearization.

1. Introduction

In the field of numerical methods for gas dynamics, the idea of employing the exact solution to a Riemann initial-value problem (RP) was proposed by Godunov (Godunov, 1959). Since then, many numerical schemes that incorporate this solution, or its approximations, followed. These are now referred to as Godunov-type schemes (Harten et. al., 1983).

The Godunov method was originally developed as an explicit finite volume method, where the cell interface numerical flux-function was approximated by the flux value calculated at the exact solution to the RP with the initial data taken to be values of the state parameter vector in two cells adjacent to the interface.

The time step for any explicit scheme is restricted by the Courant-Friedrichs-Lowy (CFL) condition, which requires the domain of dependence in a numerical scheme to be included in the domain of dependence in the corresponding differential equations. In the explicit Godunov method, this time step must be bounded in such a way that characteristics produced at each grid point do not intersect each other during a given time step.

This restriction can be relaxed if the explicit time integration scheme is replaced by an implicit one. However, the difficulty of using implicit temporal integration in Godunov schemes is that the state parameter vector to be solved is defined by a system of algebraic equations which have a complicated non-linear form because of a complex non-linear relationship between the numerical flux and the solution vector. Nevertheless, the equations can be efficiently solved by the standard Newton method that gives fast, quadratic convergence providing that the linearization of the flux is virtually exact.

Until the present time, there has been widespread opinion that the exact linearization (EL) of the Godunov flux is extremely expensive, if not impossible, not only for the exact solution, but even for an approximate one (for example Roe's solution). Therefore, an approach has been employed, consisting in the replacement of the EL in the Newton iterative method with an approximate linearization (AL). A commonly used way of such approximations was proposed by Jameson and Turkel (Jameson and Turkel, 1981). This corresponds to linearization of a simplified Roe's flux-function (Roe, 1981). A shortcoming of making this approximation in the linearization process is that the quadratic convergence inherent in Newton's iterations can no longer be achieved in so far as mismatch and inconsistency appear between the right- and left-hand sides of linearized equations.

In this paper we propose a solution to the EL of the Godunov flux-function (to our knowledge it has not been done so far), and apply it to Newton's iterations for solving discrete (digitized) equations of the Euler

implicit time-integration scheme. To do so, first we consider an auxiliary problem of variation in the RP exact solution caused by a small change in the initial data. We show that this problem has a unique solution, and exact formulas expressing this solution can be analytically obtained in an explicit compact form. These formulas are then applied to attain variation matrices involved in the linearization process.

After linearization, the resultant linear equations are solved with the LU-SGS approximate factorization method of Yoon and Jameson (Yoon and Jameson, 1988). Numerical experiments are carried out for both steady and unsteady compressible flow calculations under different flow conditions ranging from low subsonic to strongly supersonic speeds. Comparisons are made for the convergence rate of the implicit LU-SGS Godunov method, between the EL of the numerical flux and the commonly employed AL of Jameson-Turkel.

2. Variational Riemann Problem

A Riemann initial-value problem for one-dimensional extended gas dynamics

$$\frac{\partial \mathbf{Q}}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{Q})}{\partial x} = 0 \quad (1)$$

is formulated as a time-dependent Cauchy problem with the piecewise constant initial values at $t = 0$, as:

$$\mathbf{Q} = \begin{cases} \mathbf{Q}_l & \text{for } x < 0 \\ \mathbf{Q}_r & \text{for } x > 0 \end{cases} \quad (2)$$

where \mathbf{Q} is the generalised state vector including density, three components of momentum, and total energy as components, and \mathbf{f} is the corresponding flux-function. \mathbf{Q}_l and \mathbf{Q}_r are considered to be constant with the space coordinate x .

For any initial data realized from the point of view of physics, this problem always has a unique analytical solution (Kochin, 1949; Godunov et. al., 1979). This solution depends on the initial values \mathbf{Q}_l and \mathbf{Q}_r , and has the form of a piecewise-analytical function with regard to the self-similar parameter λ , $\lambda = x/t$. By denoting this function \mathbf{Q}^R , the exact solution to a RP can be written as

$$\mathbf{Q} = \mathbf{Q}(\lambda) = \mathbf{Q}^R(\lambda, \mathbf{Q}_l, \mathbf{Q}_r) \quad (3)$$

The RP's solution necessarily has a certain inner structure or wave pattern. It is defined by a set of discontinuities produced by the the initial discontinuity break up. Among these there must exist only one contact

discontinuity (CD), which separates the gas initially located on the left-hand side with $x < 0$ from that on the right-hand side with $x > 0$. A constant (uniform) flow domain referred to as the contact zone must border on the CD on either side. Pressures as well as velocities take the same values in the two contact zones. There must be only either one shock wave or one expansion Taylor wave (TW) on each side of the CD, which separates the flow in the contact zone from that in the unperturbed zone with initial values \mathbf{Q}_l or \mathbf{Q}_r . The solution continuously varies throughout the TW zone in accordance with the following relations:

$$\mathbf{Q}(\lambda) : \quad u \pm a - \lambda = 0, \quad cu' \mp p' = 0, \quad v' = w' = s' = 0 \quad (4)$$

where a is the sound velocity, u, v, w are components of the velocity vector, p is the pressure, and $c = \rho a$, where ρ is the density. The prime denotes the derivative with respect to λ .

Assuming that the exact solution to an RP is known, a Variational Riemann Problem (VRP) can be put forward as follows. Let the exact solution be given for initial vectors \mathbf{Q}_l and \mathbf{Q}_r . Considering small variations to these vectors such as

$$\mathbf{Q}_l \rightarrow \mathbf{Q}_l + \delta \mathbf{Q}_l, \quad \mathbf{Q}_r \rightarrow \mathbf{Q}_r + \delta \mathbf{Q}_r \quad (5)$$

a first variation of the solution is introduced, which can be written as

$$\delta \mathbf{Q}^R = M_l \delta \mathbf{Q}_l + M_r \delta \mathbf{Q}_r \quad (6)$$

through the variation matrices (VM) M_l and M_r given by

$$M_i = M_i(\lambda, \mathbf{Q}^R) = \frac{\partial \mathbf{Q}^R}{\partial \mathbf{Q}_i}, \quad i = l, r \quad (7)$$

The VRP addressed here is how to obtain these VMs for the arbitrary initial data of the baseline RP. It is evident that both M_l and M_r are piecewise analytical matrix-functions with the same domains of analyticity as the base non-varied solution $\mathbf{Q}^R(\lambda)$. These domains are defined by a set of waves originating from the initial discontinuity, or in other words, by the wave pattern produced in the non-varied RP.

Moreover, VMs are obviously constant with respect to λ in each analyticity domain except for TW zones, where they vary with λ . In unperturbed zones, we have $M_l = I$ and $M_r = O$ on the left-hand side and $M_l = O$ and $M_r = I$ on the right-hand side, where I is the unit-matrix and O is the null-matrix. The VM for the TW zone can be obtained in an explicit form by varying eq.(4) and integrating the resultant system of differential equations (Men'shov and Nakamura, 2000).

In contact zones, the VMs are constant with λ , and should be defined by considering conjugate relations at the rear characteristic of a TW or a shock wave. Analysing these relations, one can assert that the variational vector $\delta\mathbf{Q}$ in each contact zone is determined for all possible cases in a form which includes one arbitrary constant c_i , $i = l, r$ as follows:

$$\delta\mathbf{Q} = N^i \mathbf{Q}_i + c_i \mathbf{m}^i \quad (8)$$

where the matrix N^i and the vector \mathbf{m}^i can be explicitly expressed through the solution to the base RP, in the case where both a shock and a TW occurs between the contact and the unperturbed zones. The corresponding rather tedious formulas are not given here, but can be found in Men'shov and Nakamura (Men'shov and Nakamura, 2000).

To complete the solution, it is necessary to determine two constants, c_l and c_r in eq.(8). There are just two conjugate relations for this. These link the variations in pressure and velocity at the CD as follows:

$$[\delta u] = 0, [\delta p] = 0 \quad (9)$$

where square brackets denote the variation at the CD.

With these relations, the constants c_l and c_r are determined in a linear form with respect to the initial variations, $\delta\mathbf{Q}_i$, $i = l, r$, and variational matrices are explicitly defined throughout the flow by eq.(8). The resultant formulas for the variation matrices can be represented in a simple compact form, if the above solution is carried out for the vector of the primitive parameters $\{u, p, s, v, w\}$ instead of the conservative vector \mathbf{Q} , and the concept of proper and associated parameters is used. By this concept a parameter is considered as proper or associated depending on the value of λ . For example, if this value corresponds to the left-hand side of the CD, then the left-hand side parameters are proper, and those on the right-hand side are associated, and vice versa. By denoting associated values using the superscript (asterisk), and initial variations by the subscript (zero), and keeping the same notations for matrices and vectors as introduced above, the final formulas can be written as follows (Men'shov and Nakamura, 2000):

$$\delta\mathbf{Q} = M \delta\mathbf{Q}_0 + M^* \delta\mathbf{Q}_0^* \quad (10)$$

where proper and associated matrices are given as

$$\begin{aligned} M &= I & M^* &= O && \text{for the unperturbed zone} \\ M &= M_{TW} & M^* &= O && \text{for the TW zone} \\ M &= N - \mathbf{m} \otimes \mathbf{k}^* & M^* &= \mathbf{m} \otimes \mathbf{n}^* && \text{for the contact zone} \end{aligned} \quad (11)$$

The vectors \mathbf{n} and \mathbf{k} in this equation are given by

$$\mathbf{n} = \frac{m_2 \mathbf{N}_1 - m_1 \mathbf{N}_2}{m_2 m_1^* - m_1 m_2^*}, \quad \mathbf{k} = \frac{m_2 \mathbf{N}_1^* - m_1 \mathbf{N}_2^*}{m_2 m_1^* - m_1 m_2^*} \quad (12)$$

Here M_{TW} is the variational matrix of the TW zone, \mathbf{N}_1 and \mathbf{N}_2 are the first and the second row of the matrix N , respectively, and m_1 and m_2 are the first and the second component of the vector \mathbf{m} , respectively.

3. Implicit Time Discretization

We consider the system of Navier-Stokes equations written in the conservative form as

$$\partial_t \mathbf{q} + \partial_k \mathbf{f}_k = \partial_k \mathbf{g}_k \quad (13)$$

where \mathbf{q} is the conservative vector consisting of density, momentum and total energy, and \mathbf{f}_k and \mathbf{g}_k are inviscid and viscous fluxes. These equations are discretized in time with an implicit backward scheme, and in space with a finite volume method. After that, a pseudo-time technique (Jameson, 1991) is applied to discretized equations followed by an implicit discretization in the pseudo-time parameter, and Newton's iterations, to make the unsteady residual vanish for each time step. This leads to a system of linear equations with respect to the iterative increment of the solution vector $\Delta^s \mathbf{q} = \mathbf{q}^{n+1,s+1} - \mathbf{q}^{n+1,s}$:

$$\left[V_i \left(\frac{1}{\Delta\tau} + \frac{1}{\Delta t} \right) + D_i^{n+1,s} \right] \Delta^s \mathbf{q}_i = \mathbf{R}_i^{n+1,s} - \sum_{\sigma} \mathbf{G} (\Delta^s \mathbf{q}_{\sigma}) \quad (14)$$

where superscripts n and s indicate time levels and inner Newton's iterations respectively, $\Delta\tau$ is a pseudo-timestep, Δt is the real-timestep, and the summation over σ on the right-hand side is fulfilled over all cells surrounding the cell i .

In eq.(14), \mathbf{R} represents the iterative unsteady residual,

$$\mathbf{R}_i^{n+1,s} = \sum_{\sigma} \left(-\mathbf{f}^{n+1,s} + \mathbf{g}^{n+1,s} \right) - V_i \left(\mathbf{q}_i^{n+1,s} - \mathbf{q}_i^n \right) / \Delta t \quad (15)$$

where the inviscid flux is approximated by the Godunov flux-function based on the exact RP solution: $\mathbf{f} = \mathbf{f}(\mathbf{Q}^R)$, and $\mathbf{Q}^R = \mathbf{Q}^R(\mathbf{Q}_i, \mathbf{Q}_{\sigma})$. Here \mathbf{Q} denotes a vector with cell interface contravariant components of the Cartesian conservative vector \mathbf{q} .

The matrix D and the vector \mathbf{G} in eq.(14) result from linearization of the numerical flux, and have the following form:

$$\begin{aligned} D_i^{n+1,s} &= \sum_{\sigma} \left(M_1^{n+1,s} + \rho_{dis}^{n+1,s} \right) \\ \mathbf{G} (\Delta^s \mathbf{q}_{\sigma}) &= \left(M_2^{n+1,s} - \rho_{dis}^{n+1,s} I \right) \Delta^s \mathbf{q}_{\sigma} \end{aligned} \quad (16)$$

where matrices M_1 and M_2 define linearization of the inviscid numerical flux,

$$\delta \mathbf{f} = M_1 \delta \mathbf{q}_i + M_2 \delta \mathbf{q}_{\sigma} \quad (17)$$

and ρ_{dis} is a majorant of the spectral radius of the viscous Jacobian.

Exact linearization corresponds to the matrices $M_{1,2}$ that are evaluated through the VMs of the exact RP solution:

$$M_{1/2} = A \left(\mathbf{Q}^R \right) \frac{\partial \mathbf{Q}^R}{\partial \mathbf{Q}_{i/\sigma}} \quad (18)$$

Along with this, we also consider a commonly used approximate linearization proposed by Turkel and Jameson (Jameson and Turkel, 1981), where the matrices M_1 and M_2 are represented by the inviscid Jacobian A and its spectral radius ρ_{inv} as

$$M_1 = 0.5[A(\mathbf{Q}_i) + \rho_{inv}I]; \quad M_2 = 0.5[A(\mathbf{Q}_i) - \rho_{inv}I] \quad (19)$$

These two approaches are referred to as the exact linearization (EL) scheme and the approximate linearization (AL) scheme respectively, hereinafter.

Eq.(14) is solved at each s iteration with the LU-SGS method (Yoon and Jameson, 1988) performed in two (forward and backward) iterative sweeps. Details of this approach can be found in previous papers (Men'shov and Nakamura, 2000; Men'shov and Nakamura, 1995).

4. Results

Several computational tests are considered to evaluate the convergence behavior of the implicit Godunov method with an exactly linearized numerical flux, EL scheme, in comparison with the AL scheme.

4.1. INVISCID STEADY FLOW.

The first example is an inviscid flow past a Zhukovskii airfoil at an angle of attack of 5° under different freestream conditions. Calculations have been carried out with a C-type grid of 272×100 mesh points (128 mesh points are placed on the body surface, and 100 mesh points between the body and an outer boundary located about 15 airfoil chord lengths away from the body). In these calculations, only one inner iteration was used to solve eq.(14) at each time step. The timestep Δt defined in terms of the CFL number was taken to be in the order of 1 at the beginning of calculations and then gradually increased up to the value of 10^4 after 50 time steps.

The relative L_{inf} norm of the residual, $res_n = \max_k |\Delta^n \mathbf{q}_k|$, is taken as a criterion to examine the convergence rate in these calculations. Figure 1 shows the convergence history of the EL and AL schemes versus the time steps for freestream Mach numbers of 0.085, 0.25, 0.85, 1.1, and 8.5, numbered from 1 to 5, respectively. As can be seen from this figure, the EL

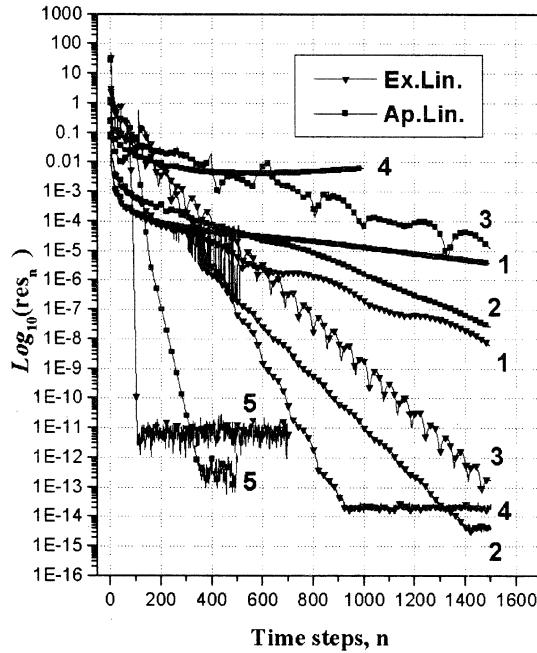


Figure 1. Residual convergence histories vs. time steps for inviscid flow around a Zhukovskii airfoil at an attack angle of 5° obtained with the AL and EL schemes: 1 – $M_{\infty} = 0.085$, 2 – $M_{\infty} = 0.25$, 3 – $M_{\infty} = 0.85$, 4 – $M_{\infty} = 1.1$, 5 – $M_{\infty} = 8.5$.

scheme offers relatively a much faster convergence rate compared to the AL scheme, in all the considered variants. However, quadratic convergence is not achieved because the system of linear equations is not solved exactly, that is, with solely one iteration at each time step, and therefore Newton's iterative process is not actually realized.

The convergence rate can be essentially improved by implementing a number of inner iterations, and the quadratic convergence can be achieved in the limit, providing these iterations are performed up to full convergence at each time step. The improvement of the convergence can be seen in Fig. 2, where convergence histories are shown regarding the EL scheme executed with no inner iterations, and with 5 sub-iterations.

4.2. VISCOUS STEADY FLOW.

The following two examples deal with calculation of laminar steady flows. In the first one, a Blasius boundary layer is computed for a finite size flat

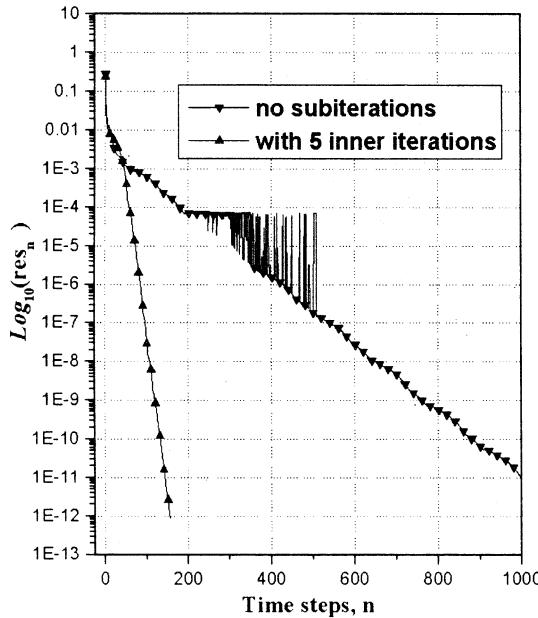


Figure 2. Effect of inner iterations on the residual in the EL scheme for inviscid flow past a Zhukovskii airfoil at $M_{\infty} = 0.25$ and an attack angle of 5° .

plate at an inflow Mach number of 0.17 and a Reynolds number of 230000 based on the plate length. A rather coarse grid is used for this calculation, consisting of 20 even-spaced grid points along the plate and 20 grid points clustered toward the plate surface. In Fig. 3, computed longitudinal velocities are shown against the Blasius similarity parameter η for all mesh points. It is seen that these computed data are well matched with the Blasius similarity law. Residual convergence histories for this problem are shown in Fig. 4. The superiority of the EL scheme over the AL scheme is clearly confirmed.

The second test is of a supersonic flow past a cylinder. The forebody transonic laminar flow is computed for a freestream Mach number of 6 and a Reynolds number of 1300000 based on the cylinder radius. The computational grid consists of 120×80 cells, where 120 even-spaced cells are placed along the cylinder surface and 80 clustered cells are in the radial direction; the minimal mesh size near the cylinder is $10^{-5} \times R_{cyl}$. Initial data used in these calculations correspond to a steady inviscid flow under the same freestream conditions. A comparison of residual convergence histories among the EL and AL schemes is given in Fig. 5. This also indicates

that the convergence rate of the EL scheme is much faster than that of the AL scheme. Note that one inner iteration is used in these calculations for each time step also, and therefore a strict Newton iterative process is not achieved, even if the exact linearization is performed.

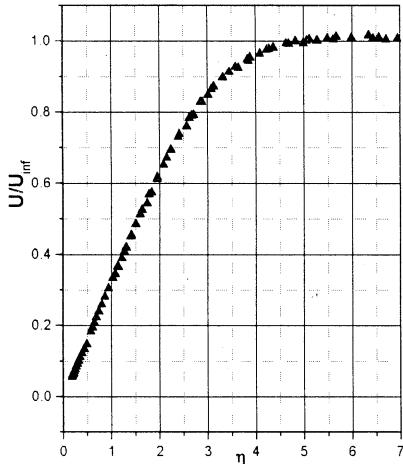


Figure 3. Computed velocities for flow over a flat plate.

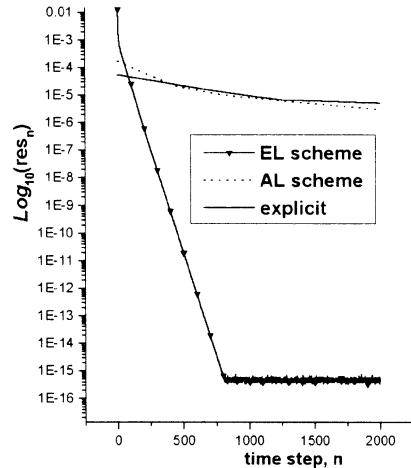


Figure 4. Residual convergence histories vs. time steps for a flat plate.

4.3. VISCOUS UNSTEADY FLOW.

This test case concerns calculations of unsteady viscous flows. For unsteady calculations, it is important to drive the unsteady residual, eq. (15), to zero for better resolution of unsteady phenomena. This process is accomplished by implementing inner iterations. To examine the convergence of inner iterations in the EL scheme, we consider a subsonic viscous flow past the same Zhukovskii airfoil as used above for inviscid calculations at the same angle of attack of 5° and a freestream Mach number of 0.5. Calculations have been carried out with the same number of mesh cells (272×100) as in the above inviscid calculations, but clustered toward the airfoil surface with a minimal normal mesh spacing at the body of $10^{-3} \times L_{chord}$. The Reynolds number in these calculations is 23000 based on the chord length. In Fig. 7, the computed vorticity field obtained with the EL scheme with 5 inner iterations is shown. As seen in this figure, an unsteady vortical flow is produced in the wake behind the airfoil because of boundary layer separation. A comparison of the convergence of inner iterations between the EL and AL schemes is given in Fig. 6 for a time step. This also displays the superiority of the EL scheme over the AL scheme. Only 5 sub-iterations are required to reduce the unsteady residual by two orders in this scheme, while more than 20 sub-iterations are needed to do this with the AL scheme.

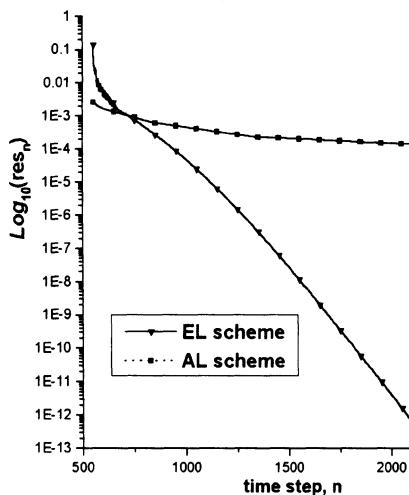


Figure 5. Residual convergence histories for supersonic laminar flow around a circular cylinder forebody.

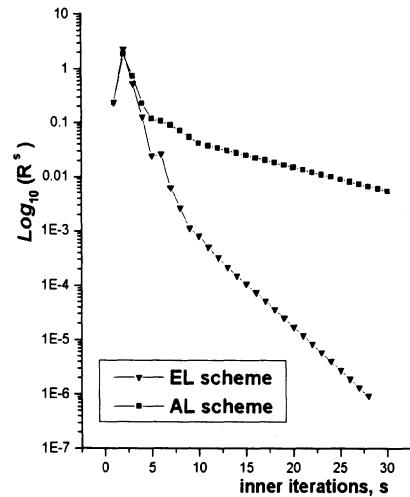


Figure 6. Convergence of inner iterations for unsteady flow past a Zhukovskii airfoil.

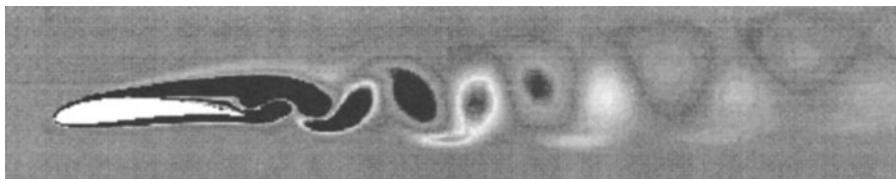


Figure 7. Computed vorticity field for viscous flow past a Zhukovskii airfoil at $M_{\infty} = 0.5$ and an attack angle of 5° .

5. Conclusions

The first variation of the exact solution to a RP caused by a small variation of initial data has been considered in order to derive the exact linearization of the Godunov numerical flux-function. Variation matrices of the RP solution associated with left- and right-hand side initial values, whereby the flux linearization is defined, have been analytically derived in an explicit compact form and implemented in an implicit Godunov method. Numerical experiments have been performed with the present method and also with the implicit Godunov method implemented with the commonly used approximate linearization of Turkel and Jameson. In the both methods, the solution to the resultant system of linear equations has been obtained with the LU-SGS approximate factorization method. The convergence processes in these two methods have been tested in calculations of several compress-

ible flows, in inviscid and viscous flow models. The results obtained show that the use of exact linearization of the numerical Godunov flux-function leads to a substantial acceleration in the residual convergence process compared with the approximate linearization scheme for both steady and unsteady flow calculations.

References

- Godunov S K (1959). A Finite-Difference Method for the Numerical Computation of Discontinuous Solutions of the Equations of Fluid Dynamics. *Matematicheskiy Sbornik* bf 47, pp 271-393 (in Russian).
- Harten A, Lax P D, and Van Leer B (1983). On Upstream Differencing and Godunov-Type Schemes for Hyperbolic Conservation Laws. *SIAM Review* **25**, pp 35-61.
- Jameson A and Turkel E (1981). Implicit Schemes and LU Decompositions. *Mathematics of Computation* **37**, pp 385-397.
- Roe P L (1981). Approximate Riemann Solvers, Parameter vectors, and Difference Schemes. *J. Comp. Physics* **43**, pp 357-372.
- Yoon S and Jameson A (1988) Lower-Upper Symmetric-Gauss-Seidel Method for the Euler and Navier-Stokes Equations. *AIAA J.* **26**, pp 1025-1026.
- Kochin N E (1949). Collected Papers II. Moskva, Izd.Akad.Nauk SSSR (in Russian).
- Godunov S K, Zabrodin A V, Ivanov M I, Kraiko A N and Prokopov G P (1979). Resolution Numerique des Problemes Multidimensionnels de la Dynamique des Gas. Moscow, Editions MIR.
- Men'shov I S and Nakamura Y (2000). Variation Matrices in the Riemann Problem with Application to Implicit Godunov's Method. AIAA Paper 2000-0920.
- Men'shov I S and Nakamura Y (2000). On Implicit Godunov's Method with Exactly Linearized Numerical Flux. *Computer & Fluids* bf 29, pp 595-616.
- Men'shov I S and Nakamura Y (1995). An Implicit Advection Upwind Splitting Scheme for Hypersonic Air Flows in Thermochemical Nonequilibrium. A Collection of Technical Papers of 6th Int.Symp.on CFD, Vol. II, Lake Tahoe, Nevada, pp 815-821.
- Men'shov I S and Nakamura Y (1995). Implementation of the LU-SGS Method for an Arbitrary Finite Volume Discretization. Proc. of 9th Conference on CFD, Tokyo, pp 123-124.
- Jameson A (1991). Time Dependent Calculations Using Multigrid, with Applications to Unsteady Flows past airfoils and wings. Technical report AIAA 91-1596, 1991.

MASS FLUX COMPUTATION AS A KEY TO THE CARBUNCLE PHENOMENON

J.-M. MOSCHETTA

Department of Aerodynamics, SUPAERO

Department Models for Aerodynamics and Energetics, ONERA

BP 4025, 31055 Toulouse CEDEX 4, France

E-mail: moscheta@oncert.fr

Abstract

Quirk's problem is used as a basis to show that the carbuncle phenomenon depends on the mass flux computation and its dependency on pressure differences as recently suggested by Liou (1997) and discussed by Xu (1998). The present paper describes a two-step procedure which aims at removing the carbuncle phenomenon without degrading the contact wave-capturing capability of a given upwind method. The resulting procedure is a quasi-conservative method in which the mass flux is modified in order to remove pressure dependency.

1. Introduction

Shock-capturing upwind methods developed since the 1980's following the pioneering work of Godunov can be classified into two distinct categories : upwind methods which exactly solve contact discontinuities (referred to as *contact-capturing* methods in the following) and upwind methods which introduce spurious diffusivity in the resolution of contact discontinuities. Schemes taken from the second category never produce the carbuncle phenomenon but are not suitable for viscous calculations. On the other hand, schemes taken from the first category are good candidates for viscous computations but usually produce the carbuncle phenomenon(Gressier, 2000). The second category includes Flux-Vector Splitting methods such as EFM (1980), Steger-Warming (1981) or Van Leer (1982). All these schemes are robust and produce nice looking shocks even with extremely high pressure ratios. An attractive property for high-speed flows applications is that

none of these schemes suffer from the *carbuncle phenomenon* which arises when intense shock waves are aligned with grid lines(Quirk, 1994). However, when applied to the Navier-Stokes equations, they tend to artificially broaden boundary layers for relatively coarse grids and actually require a huge number of points to provide acceptable viscous solutions. The first category, which includes methods such as Godunov (1959), Roe (1981), Osher (1982), AUSM (1993), EIM (1993), contains methods which are suitable for Navier-Stokes computations in the sense that they can produce accurate viscous solutions with a limited number of points in boundary layers. However, with a few exception, they all suffer from marginal stability problems such as, for instance, the carbuncle phenomenon or the kinked Mach stem(Gressier, 2000). In this context, AUSM scheme appears as an exception to the incompatibility which seems to exist between “viscous accuracy” and stable shock wave computation. In the present paper, Quirk’s problem(Quirk, 1994) is used as a basis to show that the carbuncle phenomenon depends on the mass flux computation and its dependency on pressure differences as recently suggested by Liou(Liou, 1997) and discussed by Xu(Xu, 1998).

2. Numerical procedure

In order to investigate the mass flux influence on the carbuncle phenomenon, one can start with a basic upwind scheme which is contact-capturing but suffers from the carbuncle phenomenon. A second upwind scheme is combined with the first one in order to remove the carbuncle problem using the following procedure :

1. First, the basic upwind scheme is applied to compute the intermediate state vector of conservative variables :

$$\bar{U} = (\bar{\rho}, \bar{\rho}\bar{u}, \bar{\rho}\bar{v}, \bar{\rho}\bar{E})^T \quad (1)$$

2. Second, the density is separately computed using the mass flux taken from a second upwind scheme :

$$\rho_i^{n+1} = \rho_i^n - \sigma \left(f m_{i+1/2}^{(1)}(U_i^n, U_{i+1}^n) - f m_{i-1/2}^{(1)}(U_{i-1}^n, U_i^n) \right) \quad (2)$$

3. Third, the final state vector is computed by keeping constant the intermediate values for the velocity components and the pressure as :

$$U^{n+1} = (\rho^{n+1}, \rho^{n+1}\bar{u}, \rho^{n+1}\bar{v}, \frac{\bar{p}}{\gamma - 1} + \frac{1}{2}\rho^{n+1}(\bar{u}^2 + \bar{v}^2))^T \quad (3)$$

Numerical experiments show that if the second upwind scheme does not display the carbuncle phenomenon, the above three-step procedure provides

a carbuncle-free method. This situation occurs whenever its mass flux does not explicitly depend on the pressure difference $\Delta p = p_R - p_L$. The above procedure does not guarantee that the resulting method is still contact-capturing since, for instance, the use of a Flux-Vector-Splitting mass flux will remove the carbuncle problem but will also ruin the contact-capturing property of the basic upwind scheme.

In the case of a moving contact discontinuity, Godunov's method provides the ideal mass flux form toward which any contact-capturing method should tend. First, let us consider the two-dimensional initial states in primitive form on either sides of a moving contact wave

$$W_L = \begin{pmatrix} \rho_L \\ u \\ v_L \\ p \end{pmatrix}, \quad W_R = \begin{pmatrix} \rho_R \\ u \\ v_R \\ p \end{pmatrix} \quad (4)$$

Godunov state corresponding to the exact solution to the Riemann problem at the initial interface is given by

$$W^G = \begin{pmatrix} \rho^G & = & \frac{1}{2} [\rho_L + \rho_R - \text{sgn}(u)(\rho_R - \rho_L)] \\ u^G & = & u \\ v^G & = & \frac{1}{2} [v_L + v_R - \text{sgn}(u)(v_R - v_L)] \\ p^G & = & p \end{pmatrix} \quad (5)$$

As a consequence, Godunov mass flux is

$$fm^G = \frac{1}{2} [u(\rho_L + \rho_R) - |u|(\rho_R - \rho_L)] \quad (6)$$

Therefore, the necessary condition which should be satisfied by a mass flux to preserve the contact-capturing property is

$$fm(W_L, W_R) \longrightarrow \frac{1}{2} [u(\rho_L + \rho_R) - |u|(\rho_R - \rho_L)] \quad (7)$$

when W_L, W_R tend to the initial states correponding to a moving contact discontinuity (see Eq.4).

An important feature of numerical methods applied to the resolution of Euler equations is the conservativity property. This property is crucial when intense shock waves are present in the computational domain. In the present method, the third step (Eq.3) introduces a non conservative transformation of the initial state vector. Indeed, two adjacent cells, labeled (1) and (2), are updated using the above procedure, the computation of density is still

conservative but not the computation of momentum, for instance. In other words,

$$\rho_1^{n+1} + \rho_2^{n+1} = 2\rho_{(1-2)}^{n+1}, \quad \text{but} \quad \rho_1^{n+1}\bar{u}_1 + \rho_2^{n+1}\bar{u}_2 \neq 2\rho_{(1-2)}^{n+1}u_{(1-2)}^{n+1} \quad (8)$$

where $(1 - 2)$ denotes the cell which is made of the two adjacent cells (1) and (2) .

3. Results and discussion

Quirk's problem has been selected to analyze the carbuncle phenomenon because it removes the strong grid dependency usually associated with the blunt body problem. It consists of a shock wave traveling with a Mach number $M_s = 6$ along a duct filled with air initially at rest. The 800×20 grid covers a 40×1 length unit duct with the grid centerline perturbed following

$$y_{i,j_{mid}} = y_{j_{mid}} + (-1)^i \cdot 10^{-6} \quad (9)$$

The carbuncle phenomenon, which appears with the original Godunov's method (Fig. 1), completely vanishes when the density is calculated using Van Leer's mass flux. The same result could be obtained by using any other flux-vector splitting mass flux.

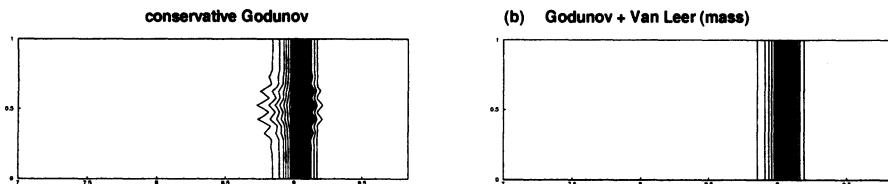


Figure 1. Pressure contours for Quirk's problem using : (a) the original Godunov's method, (b) Godunov scheme with Van Leer mass flux.

The resulting scheme does not exactly solve contact discontinuities any longer but still exactly solves pure shear waves (i.e. only tangential velocities are different). This may be acceptable for incompressible boundary layer but not for compressible boundary layers in which density gradients are not negligible. When AUSM scheme, which does not produce, by itself, the carbuncle phenomenon, is used for the momentum and the energy equations with Godunov's method for the mass flux, the result is very unstable (Fig. 2 a). When the blending is performed the other way round, the carbuncle phenomenon has disappeared (Fig. 2 b). This indicates that the mass flux in AUSM is indeed responsible for the absence of carbuncle phenomena.

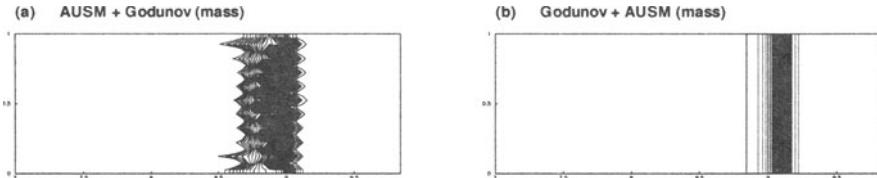


Figure 2. Pressure contours for Quirk's problem using : (a) AUSM with Godunov's mass flux, (b) Godunov scheme with AUSM mass flux.

Liou's conjecture states that a sufficient condition to remove carbuncle-like failings is that the mass flux should not depend on the pressure difference(Liou, 1997). In the present paper, Roe scheme is modified in order to remove pressure dependency *only* in the mass flux expression. The resulting mass flux has the following form :

$$fm = \frac{1}{2} [(\rho u)_L + (\rho u)_R - |\tilde{u}| \Delta \rho - \tilde{\mathcal{M}} \Delta u] \quad (10)$$

where $\Delta \Phi = \Phi_R - \Phi_L$, Roe's average state is denoted by \sim and

$$\mathcal{M} = \frac{1}{2} (|\tilde{M} + 1| - |\tilde{M} - 1|) \quad (11)$$

It should be noticed that the modified Roe mass flux still allows the exact resolution of contact waves since it reverts to Godunov's mass flux when $\Delta u = 0$ and $\Delta p = 0$ (Eq.7). The resulting "modified" Roe scheme yields a carbuncle-free solution (Fig. 3) without degrading the desirable property of exactly solving contact discontinuities.

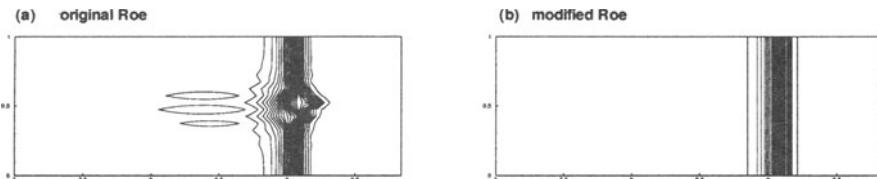


Figure 3. Pressure contours for Quirk's problem using : (a) the original Roe scheme, (b) the modified Roe scheme.

When removing the pressure dependency from the mass flux, one can expect some problems for the computation shock tube problems in which the initial states only differ from one another by the pressure difference. In that case, the mass flux writes $fm = \rho u$ regardless of the pressure difference. This is what happens for AUSM+ for which a small time step is

necessary to stabilize the computation during the first iterations. The influence of pressure differences in the mass flux is yet physical and removing this influence might degrade the numerical solution in a different situation. In order to assess the error due to the non-conservativity of the method, a stronger shock wave moving at Mach 20 has been computed using the original Van Leer conservative method and the present approach combining Van Leer flux with the modified Roe mass flux function given by Eq.(10). Numerical results are remarkably close to each other with a maximum relative error of 10^{-2} percent (Fig.4).

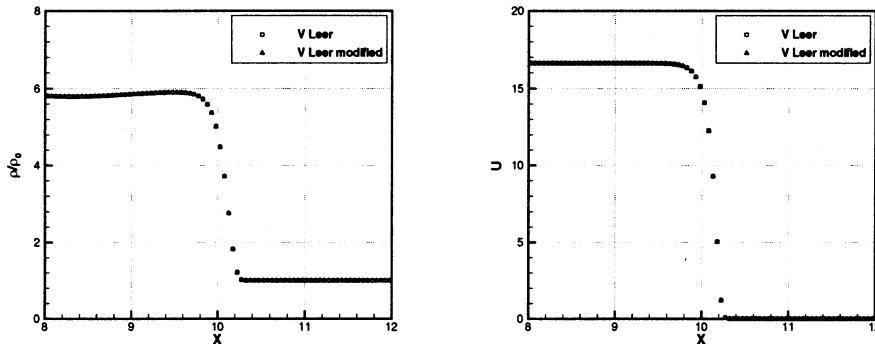


Figure 4. Quirk's problem at Mach 20 using the original Van Leer's method and the modified Van Leer mass flux (Eq.10) : density distribution (left), velocity distribution (right).

The present method, although not fully satisfactory because of the additional computing effort required is superior to the standard primitive approach which yields an oscillation-free solution with a wrong shock speed (Fig.5). The primitive approach is the one proposed by Toro(Toro, 1995) using the vector of primitive variables $W = (\rho, u, v, p)$.

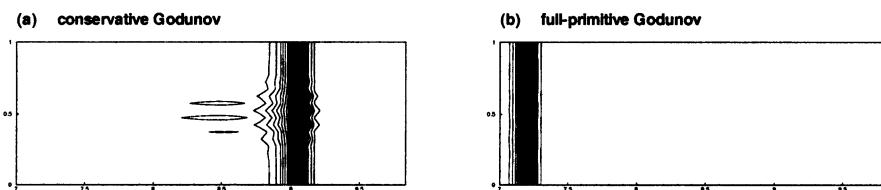


Figure 5. Comparison between the conservative Godunov solution (left) and the primitive Godunov formulation (right).

Finally, when the modified mass flux (Eq. 10) is abruptly used in place of the original Roe mass flux formulation, instability problems have been observed, leading to the above three-step quasi-conservative procedure.

4. Conclusion

The present study aims at assessing Liou's conjecture which suggests that pressure dependencies on the mass flux computation is responsible for the carbuncle phenomenon. A quasi-conservative approach has been proposed to investigate this conjecture using various upwind schemes. The present method does not intend to be used as an alternative to standard entropy fixes but aims at illustrating the importance of the mass flux computation in the onset of the carbuncle phenomenon. Although it is not strictly conservative, the present approach does not significantly affect the computation of correct shock speeds even for intense shock waves. Under some restrictions, suitable forms of the mass flux can allow carbuncle-free solutions and maintain the contact-capturing property if this was already satisfied by the original numerical flux.

References

- Gressier J, Moschetta J M (2000). Robustness versus Accuracy in Shock Wave Computations. *International Journal for Numerical Methods in Fluids* to appear.
- Liou M S (1997). Probing Numerical Fluxes : Mass Flux, Positivity, and Entropy-Satisfying Property. *AIAA Paper 97-2035*.
- Quirk J J (1994). A Contribution To the Great Riemann Solver Debate. *Int. J. Numer. Meth. Fluids* **18** pp 555-574.
- Toro E F (1995). On Adaptable Primitive-Conservative Schemes for Conservation Laws. in M. Hafez, editor, *Proceedings of the 6th International Symposium on Computational Fluid Dynamics, Vol. III* Lake Tahoe, Nevada, pp 1294-1299.
- Xu K (1998). Gas Kinetic Schemes for Unsteady Compressible Flow Simulations. *VKI Lecture Series 1998-03*.

ON THE POSITIVITY OF FVS SCHEMES

J.-M. MOSCHETTA, J. GRESSIER

Department of Aerodynamics, SUPAERO

*Department Models for Aerodynamics and Energetics, ONERA
BP 4025, 31055 Toulouse CEDEX 4, France*

E-mail: moscheta@oncert.fr

Abstract

Over the last ten years, robustness of schemes has raised an increasing interest among the CFD community. One mathematical aspect of scheme robustness is the positivity preserving property. At high Mach numbers, solving the conservative Euler equations can lead to negative densities or internal energy. Some schemes such as the flux vector splitting (FVS) schemes are known to avoid this drawback. In this study, a general method is detailed to analyze the positivity of FVS schemes. As an application, three classical FVS schemes (Van Leer's, Hänel's variant and Steger and Warming's) are proved to be positively conservative under a CFL-like condition. Finally, it is proved that for any FVS scheme, there is an intrinsic incompatibility between the desirable property of positivity and the exact resolution of contact discontinuities.

1. Introduction

In highly accelerated flows, the total energy is mainly composed of kinetic energy. Yet, in conservative formulation, both total and kinetic energy are computed independently, and their difference may yield negative internal energy, aborting the computation. In order to give some mathematical interpretation of schemes robustness or weakness in such severe configurations, it is useful to introduce the positivity property: a scheme is said to be positively conservative if, starting from a set of physically admissible states, it can only compute new states with positive densities and internal energies. Perthame (Perthame, 1990) first proposed a scheme which satisfies this property. Afterwards, Einfeldt *et al.* (Einfeldt, 1991)

gave some results concerning Godunov-type schemes. They proved that Godunov scheme is positively conservative while Roe's scheme is not, and derived the HLLE method, a positive variant of HLL schemes family. Later, Villedieu and Mazet (Villedieu, 1995) proved that Pullin's EFM kinetic scheme (Pullin, 1980) is positively conservative under a CFL-like condition. Recently, Dubroca (Dubroca, 1998) proposed a positive variant of Roe's method. Since any scheme is positively conservative for a zero time step, it is absolutely essential to specify a time step condition when defining the positivity property.

2. Defining scheme positivity

Since one can formally extend any first-order one-dimensional positively conservative method to a second-order multidimensional positively conservative method (Perthame, 1996; Linde, 1997), we will restrict ourselves to the case of first-order schemes for the one-dimensional Euler equations in the following analysis. A conservative explicit method applied to the Euler equations can be expressed as

$$\mathbb{U}_i = \mathcal{U}_i - \frac{\Delta t}{\Delta x} [F_{i+1/2} - F_{i-1/2}] \quad (1)$$

where \mathcal{U}_i is the average value over cell Ω_i of the vector of conservative variables ${}^T(\rho, \rho u, \rho E)$ at a given time step. \mathbb{U}_i is the updated state vector. Δx is the measure of cell Ω_i and $F_{i+1/2}$ is the numerical flux between the cells Ω_i and Ω_{i+1} . The numerical flux is a function $F_{i+1/2} = F(\mathcal{U}_i, \mathcal{U}_{i+1})$ of the states in both neighboring cells. The numerical flux must satisfy the consistency condition $F(\mathcal{U}, \mathcal{U}) = \mathcal{F}(\mathcal{U})$, where \mathcal{F} is the exact Euler flux. The discretized conservation equation Eq. (1) can be rewritten as

$$\mathbb{U}_i = \mathcal{U}_i - \frac{\chi_{loc,i}}{\lambda_i} [F_{i+1/2} - F_{i-1/2}] \quad (2)$$

where $\lambda(\mathcal{U})$ is the characteristic wave speed defined by $\lambda(\mathcal{U}) = |u| + a$ and $\chi_{loc,i} = \lambda(\mathcal{U}_i)\Delta t/\Delta x$. For physical reasons, the state \mathcal{U} cannot take any arbitrary value in \mathbb{R}^3 . Its density and internal energy must be both strictly positive. One can define $\Omega_{\mathcal{U}}$ as the open set of physically admissible states

$$\Omega_{\mathcal{U}} = \{\mathcal{U} = {}^T(u_1, u_2, u_3) \mid u_1 > 0 \text{ and } 2u_1u_3 - u_2^2 > 0\} \quad (3)$$

Vacuum is an admissible state for the closure $\bar{\Omega}_{\mathcal{U}}$ but not for $\Omega_{\mathcal{U}}$ since it is not expected to be reached in practical computations.

Definition 1 A scheme is said to be positively conservative if and only if there exists a constant χ , such that if both conditions are satisfied

$$\bullet \quad \forall i \in \mathbb{Z}, \quad \mathcal{U}_i \in \Omega_{\mathcal{U}} \quad (4a)$$

$$\bullet \quad \Delta t \leq \chi \frac{\Delta x}{\max_{i \in \mathbb{Z}} \lambda(\mathcal{U}_i)} \quad (4b)$$

then

$$\forall i \in \mathbb{Z}, \quad \mathbb{U}_i \in \Omega_{\mathcal{U}} \quad (5)$$

For $\Delta t = 0$, according to Eq. (1), one has $\forall i \in \mathbb{Z}, \quad \mathbb{U}_i = \mathcal{U}_i \in \Omega_{\mathcal{U}}$ for any flux function used. So, for any continuous flux function F , since $\Omega_{\mathcal{U}}$ is an open subset of \mathbb{R}^3 , whatever initial conditions \mathcal{U}_i are in $\Omega_{\mathcal{U}}$, one can find Δt small enough which will preserve positivity of states \mathbb{U}_i . Consequently, the property of positivity consists of proving that Δt is not too small compared to the maximum time step given by the CFL condition. Otherwise, one can find a situation in which a physical admissible state can only be obtained by a vanishing time step, which is not acceptable for practical gas dynamics applications. On the contrary, a scheme is said to be *non-positive* if

$$\forall \chi > 0, \quad \exists (\mathcal{U})_{i \in \mathbb{Z}} \in \Omega_{\mathcal{U}}, \quad \mathbb{U}_i \notin \Omega_{\mathcal{U}} \quad (6)$$

For a non-positive scheme (e.g. Roe, AUSM), one may have to use an extremely small time step to update the solution and may not be able to produce a physically admissible solution after a finite period of time.

3. Positivity of FVS methods

This study has been restricted to a class of FVS schemes in which the fluxes F^\pm satisfy the symmetry property

$$\overline{F^-(\mathcal{U})} = -F^+(\overline{\mathcal{U}}) \quad (7)$$

where \overline{X} is the symmetric vector ${}^T(x_1, -x_2, x_3)$ of $X = {}^T(x_1, x_2, x_3)$. It should also satisfy

$$\forall u, a \in \mathbb{R} \times \mathbb{R}^+, \quad \lim_{\rho \rightarrow 0} F^\pm(\rho, u, a) = 0 \quad (8)$$

Since $F^\pm(\mathcal{U})$ is generally an homogeneous function of ρ , Eq. (8) is not a restrictive assumption in practice.

Theorem 1 A given consistent FVS scheme satisfying properties (7) and (8) is positively conservative if and only if its F^\pm functions satisfy both properties:

$$\bullet \quad \forall \mathcal{U} \in \Omega_{\mathcal{U}}, \quad F^+(\mathcal{U}) \in \bar{\Omega}_{\mathcal{U}} \quad (9a)$$

$$\bullet \quad \exists \chi > 0, \quad \forall \mathcal{U} \in \Omega_{\mathcal{U}}, \quad \mathcal{U} - \frac{\chi}{\lambda(\mathcal{U})}[F^+(\mathcal{U}) - F^-(\mathcal{U})] \in \bar{\Omega}_{\mathcal{U}} \quad (9b)$$

In that case, the less restrictive positivity condition is expressed as

$$\forall i \in \mathbb{Z}, \quad \chi^{loc}_i < \chi_{opt} \quad (10)$$

where χ_{opt} is the greatest constant χ satisfying (9b).

Proof A detailed proof can be found in (Gressier, 1999).

The condition (9b) leads to a maximum time step which has then to be put into a CFL-like form $\chi^{loc} < \chi_{opt}$. This is the case for VL, VLH and SW schemes since $\chi_{opt} = \inf_M (\chi_{max})$ is not zero. It turns out that the inter-

		VL	VLH	SW
F^+	Supersonic		$ M \geq \sqrt{\frac{\gamma-1}{2\gamma}}$	
	Subsonic		$\gamma \geq 1$	$1 \leq \gamma \leq 3$
\mathcal{W}	Supersonic		$\chi^{loc} < \frac{ M +1}{ M + \sqrt{\frac{\gamma-1}{2\gamma}}}$	
	Subsonic	$\chi^{loc} < \chi_{max}^{VL}$	$\chi^{loc} < \chi_{max}^{VLH}$	$\chi^{loc} < \chi_{max}^{SW}$

TABLE 1. Internal energy positivity conditions

nal energy positivity conditions are always more stringent than the mass positivity conditions. Therefore, it is the internal energy positivity condition which actually drives the scheme positivity. Moreover, it means that zero values cannot be reached simultaneously by density and internal energy. Since expressions of χ_{max}^{VL} , χ_{max}^{VLH} and χ_{max}^{SW} are intricate, they are not detailed but these coefficients can be easily computed as a function of the local Mach number. The smallest values of these conditions have been computed and lead to the optimal CFL condition χ_{opt} which ensures that the scheme is positively conservative in all configurations. These constants χ_{opt} are summarized in table 2 and lead to an optimal CFL number of one for

VL	VLH	SW
1	$\min\left(1, \frac{2}{\gamma}\right)$	1

TABLE 2. Optimal CFL number
 χ_{opt} .

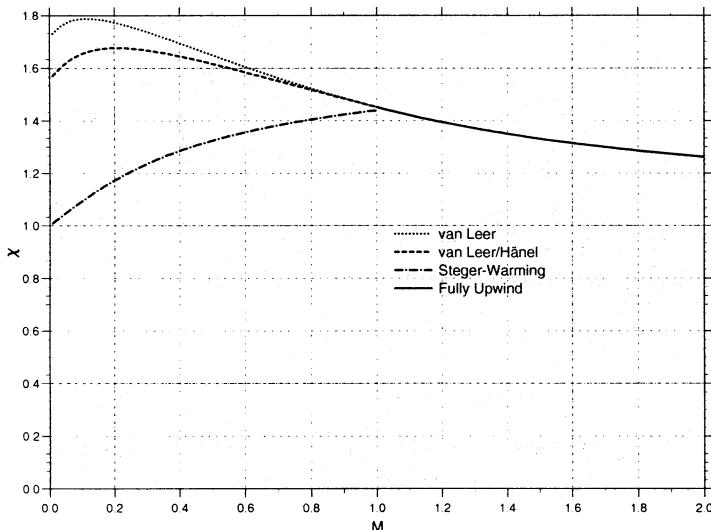


Figure 1. Maximum CFL number χ^{loc} to ensure internal energy positivity.

usual gases where $1 < \gamma < 2$. Since necessary and sufficient conditions have been derived, it can be interesting to plot the local CFL conditions. The different χ_{max} functions are plotted on Fig. 1 for $\gamma = 1.4$. The SW scheme yields the most severe condition while the VL scheme allows a greater local CFL condition in the subsonic range. All three curves merge in the supersonic range where the CFL condition implies that χ should decrease to 1 for high Mach numbers (Fig. 1). As a consequence, a CFL number of one *a fortiori* ensures positivity of the three schemes. According to Fig. 1, a CFL number of 1.45 (for $\gamma = 1.4$) can be used with Van Leer's method if the flow is expected to remain subsonic. Note that this condition only ensures the scheme positivity, but not its stability. Using too high CFL numbers might produce oscillations even though the updated solution would still be an admissible state.

4. Accuracy versus Positivity

Most FVS schemes have proved to be robust in many flow configurations but none of them are able to exactly resolve contact discontinuities since it remains a non-vanishing dissipation which smears out an initial discontinuity of density. Van Leer (Van Leer, 1991) pointed out that preventing numerical diffusion of contact discontinuities may lead to a marginally stable or unstable behavior for slow flows.

Theorem 2 *If a FVS scheme exactly preserves stationary contact discontinuities, then it cannot be positively conservative.*

Proof Consider a FVS scheme given by its flux functions F^\pm , and assume it exactly preserves stationary contact discontinuities. Then, the interface flux between $\mathcal{U}_L = {}^T \left(\rho_L, 0, \frac{p}{\gamma-1} \right)$ and $\mathcal{U}_R = {}^T \left(\rho_R, 0, \frac{p}{\gamma-1} \right)$ must satisfy

$$F^+(\mathcal{U}_L) + F^-(\mathcal{U}_R) = (0, p, 0)^T \quad (11)$$

Since ρ_L and ρ_R are independent variables, $F^+(\mathcal{U}_L)$ must be a function of only p . Hence, for all $\mathcal{U} = {}^T \left(\rho, 0, \frac{p}{\gamma-1} \right)$,

$$F^+(\mathcal{U}) = (f_1(p), f_2(p), f_3(p))^T \quad (12a)$$

Moreover, considering the symmetry property (7) and using $\bar{\mathcal{U}} = \mathcal{U}$, one has $F^-(\mathcal{U}) = -\overline{F^+(\mathcal{U})}$. Then,

$$F^-(\mathcal{U}) = (-f_1(p), +f_2(p), -f_3(p))^T \quad (12b)$$

Substituting expressions (12a) and (12b) in Eq. (11), one obtains $f_2(p) = p/2$. Moreover, $f_1(p)$ must be positive or null to satisfy the condition (9a) of positivity. If $f_1(p) = 0$, condition (9a) is not satisfied since $f_2(p)$ is not null.

If $f_1(p) > 0$, then the first component of $\mathcal{W}_r = \mathcal{U} - \frac{\chi^{loc}}{\lambda} [F^+(\mathcal{U}) - F^-(\mathcal{U})]$ may be expressed as

$$\rho - \frac{\chi^{loc}}{a} 2f_1(p) = \rho - \sqrt{\rho} \left(2\chi^{loc} \frac{f_1(p)}{\sqrt{\gamma p}} \right) \quad (13)$$

Hence, for all functions $f_1(p)$ and for all $\chi^{loc} > 0$, one can always find p and ρ such that expression (13) is negative.

5. Conclusion

A general method to prove the positivity of FVS schemes has been proposed and applied to standard FVS schemes, namely the van Leer scheme, one of

its variants, and the Steger and Warming scheme. Although these schemes have been known for a long time to be robust, they are now proved to be positively conservative under a CFL condition of 1, for usual values of the specific heat ratio γ in the range [1; 2]. In particular, this shows that all these FVS schemes can be confidently applied to gas dynamics problems including real gas effects for which γ may range between 1.4 and 1. Moreover, these conditions have been proved to be incompatible with the particular form of FVS schemes which would be able to exactly preserve stationary contact discontinuities. In other words, one cannot develop a robust and accurate scheme within the FVS family.

References

- Dubroca B (1998). Positively Conservative Roe's Matrix for Euler Equations. *Lecture Notes in Physics*, pp 272–277. 16th ICNMFD, Springer Verlag, 1998.
- Einfeldt B, Munz C D, Roe P L, Sjögren B (1991). On Godunov-Type Methods near Low Densities. *J. Comput. Phys.* **92**, pp 273–295.
- Gressier J, Villedieu P, Moschetta J M (1999). Positivity of Flux-Vector Splitting Schemes. *J. Comput. Phys.* **155**, pp 199–220.
- Linde T, Roe P L (1997). Robust Euler Codes. *AIAA Paper* 97-0209.
- Perthame B (1990). Boltzmann type Schemes for Gas Dynamics and the Entropy Property. *SIAM J. Numer. Anal.* **27**, pp 1405–1421.
- Perthame B, Shu C (1996). On Positive Preserving Finite Volume Schemes for the Compressible Euler Equations. *Numer. Math.* **76**, pp 119–130.
- Pullin D I (1980). Direct Simulation Methods for Compressible Inviscid Ideal-Gas Flow. *J. Comput. Phys.* **34**, pp 231–244.
- Van Leer B (1991). Flux-Vector Splitting for the 1990s. *NASA CP-3078*.
- Villedieu P, Mazet P (1995). Kinetic Schemes for the Euler Equations Out of Thermochemical Equilibrium. *La Recherche Aerospatiale* **2**, pp 85–102.

THE CARBUNCLE PHENOMENON : A GENUINE EULER INSTABILITY ?

J.-M. MOSCHETTA, J. GRESSIER, J.-C. ROBINET, G. CASALIS

Department of Aerodynamics, SUPAERO

Department Models for Aerodynamics and Energetics, ONERA

BP 4025, 31055 Toulouse CEDEX 4, France

E-mail: moscheta@oncert.fr

Abstract

Owing to the strong link frequently observed between the carbuncle phenomenon and Quirk's problem, the stability of a shock wave propagating down a duct has been recently revisited (Robinet, 2000), leading to the existence of an unstable mode which only occurs for a specific value of the upstream Mach number. The purpose of the present paper is to establish a relationship between the unstable mode given by the stability theory and the carbuncle phenomenon. The numerical results suggest that the carbuncle phenomenon is not a pure numerical pathology but reveals an instability mechanism intrinsically associated to the Euler equations.

1. Linear stability analysis

The so-called *carbuncle phenomenon*, which has been numerically observed with many classical upwind schemes such as Godunov (1959), Roe (1981), Osher (1982), to name a few, has been attributed so far to a "pathological" behaviour associated with the numerical methods and lots of efforts have been devoted in the development a suitable "cure" for this instability. As proposed by Quirk (Quirk, 1994), the carbuncle phenomenon can be studied in the simple geometry of a rectangular shock tube, in which a shock propagates with a slightly perturbed symmetry line. This problem has been recently studied in the framework of a linear stability analysis of the continuous Euler equations for the problem of a planar moving shock wave (Robinet, 2000). A small perturbation technique has been carried out using the continuous Euler equations written for a perfect gas and the

Rankine-Hugoniot shock relations. Each physical quantity q is written in the following form :

$$q = \bar{q} + \hat{q}e^{i(k_x x + k_y y - \omega t)} \quad (1)$$

where \bar{q} represents the mean value of q downstream of the shock, x and y being the Cartesian coordinates. Due to the confined geometry, the wave number k_y can only have discrete values. The wave number k_x is a complex number, its imaginary part must be such that the perturbation decays far away from the shock. Finally, ω is a complex number, its real part ω_r represents the frequency of the perturbation and its imaginary part ω_i a temporal growth rate. If a *normal mode form* of the perturbations is initially assumed, no unstable modes can exist as already established by previous authors. However, in the special case when

$$\omega = \pm ik_y \overline{U_1}, \quad k_x = ik_y \quad (2)$$

a Jordan transform of the resulting algebraic system must be applied and provides perturbations in *non normal mode* form. The dispersion relation leads to the existence of an unstable mode which only exists for a critical value $M_{0,c}$ of the upstream Mach number defined as :

$$M_{0,c} = \sqrt{\frac{5 + \gamma}{3 - \gamma}} \quad (3)$$

2. Numerical Results and Discussion

First, the temporal evolution of the unstable mode is numerically studied. Quirk's problem is computed with $M_0 = 6$, $\gamma = 1.4$ and a CFL number of 0.7. The evolution in time of the transverse velocity amplitude is plotted in Fig.1 for two different values of the initial grid perturbation (triangles for 10^{-4} and circles for 10^{-6}). Let alone the forced zone and the non linear zone, an exponential growth rate, independent from the initial perturbation amplitude, is obtained. The numerical slope on Fig.1 corresponds to the temporal amplification ω and can then be "experimentally" measured. Since the temporal amplification is related to the spatial behaviour by Eq.(2), one can write the spatial evolution of the fluctuating transverse velocity eigenfunction as

$$v_f^* = \operatorname{Re} [Axe^{ik_x x}] \xrightarrow{\text{Dispersion relation}} v_f^* = \operatorname{Re} \left[Axe^{-\frac{\omega_i}{U_1} x} \right], \quad (4)$$

where A is an arbitrary complex amplitude and ω_i is the temporal amplification factor. The spatial evolution is then compared to the computed

transverse velocity amplitude using the measured value of ω plotted for different values of y (Fig.2). Note that, since the theoretical perturbation is defined within a constant of proportionality, the amplitude A of the theoretical eigenfunction has been tuned to fit the numerical evolution in x , this amplitude has been determined for each y independently. The good agreement observed in Fig.2 between the theoretical and the numerical evolutions in space, can be interpreted as a confirmation of the dispersion relation, provided that the spatial evolution $xe^{-\alpha x}$ is fairly unusual in stability problems. Numerically, for values less than $M_{0,c}$ ($M_{0,c} = 2$ for $\gamma = 1.4$), the shock remains stable, even for numerical scheme which usually display the carbuncle phenomenon. Fig. 3 shows the computed values of the transverse velocity component as function of the upstream Mach number M_0 , for three different numerical schemes, which are known to produce the carbuncle, namely Roe, Osher and HLLC (Gressier, 2000). If a numerical scheme which does not exactly solve the contact discontinuity is used (e.g. any member of the Flux-Vector Splitting family), the shock remains stable regardless of the shock wave Mach number. Although three upwind schemes based on totally different approaches have been tested, a single threshold value of M_0 is clearly obtained. This numerical value is in close agreement with the theoretical value $M_{0,c} = 2$ predicted by Eq. (3) when $\gamma = 1.4$. Numerical thresholds computed from Roe's method are plotted in terms of the upstream Mach number (Fig. 4, left) above which the carbuncle instability appears. These values are compared with the theoretical ones and a close agreement is obtained for various gamma values. The corresponding downstream Mach number is $M_1 = 1/\sqrt{3}$ which does not depend on γ as observed on Fig. 4 (right). Furthermore, Roe scheme is applied to Quirk's problem and a standard entropy fix is selectively applied to the different characteristic waves (Fig. 5). Namely, for the transverse numerical flux, Roe's matrix eigenvalues are:

$$\tilde{\lambda}_{1,4} = \tilde{v} \pm \tilde{a} ; \quad \tilde{\lambda}_2 = \tilde{v} ; \quad \tilde{\lambda}_3 = \tilde{v} \quad (5)$$

and the corresponding right eigenvectors are:

$$\tilde{r}_{1,4} = \begin{pmatrix} 1 \\ \tilde{u} \\ \tilde{v} \pm \tilde{a} \\ \tilde{H} \pm \tilde{a}\tilde{v} \end{pmatrix}, \quad \tilde{r}_2 = \begin{pmatrix} 1 \\ \tilde{u} \\ \tilde{v} \\ \frac{1}{2}(\tilde{u}^2 + \tilde{v}^2) \end{pmatrix}, \quad \tilde{r}_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \tilde{u} \end{pmatrix} \quad (6)$$

When an entropy fix is applied to the entropy wave (i.e. to the eigenvalue associated to \tilde{r}_2), oscillations compare well with the case where the entropy fix is applied to the acoustic waves (i.e. to the eigenvalues associated to $\tilde{r}_{1,4}$). However, when applied to the vorticity wave (i.e. to the

eigenvalues associated to \tilde{r}_3), oscillations are significantly damped and can even be totally removed if a greater Harten's parameter is used. This result is consistent with the theoretical analysis in which the unstable mode is associated to the vorticity wave. Finally, one could argue that the natural viscosity would damp the unstable mode, explaining why the shock instability has not been experimentally observed so far. However, a viscous computation has been carried out to show that the carbuncle phenomenon still appears for fairly low Reynolds numbers (based on the duct width). Fig.6 shows that unless a tremendous quantity of molecular viscosity is added to the flow ($Re = 100$), the carbuncle still appears.

3. Conclusion

The stability analysis proposed by (Robinet, 2000) indicates that there is an unstable mode of the continuous inviscid equations which cannot be predicted by the normal mode form. Numerical results show that there is a strong link between the theoretical form of the perturbation and the numerical behaviour observed in the carbuncle phenomenon onset: Although the theory predicts a singular value for the unstable mode, numerical calculations give a threshold value above which the instability develops. This difference is yet unexplained. Furthermore, even if the occurrence of the carbuncle phenomenon seems intrinsic to the Euler equations, the influence of the numerical parameters is far from being negligible. In particular, the CFL number, the order of accuracy of the method and the dissipative mechanism included in each scheme (e.g. FVS methods) significantly affect the numerical threshold value above which the carbuncle appears. For many years, it was tacitly assumed in the CFD community that the carbuncle phenomenon was a purely numerical problem. The present work suggests that this is not true.

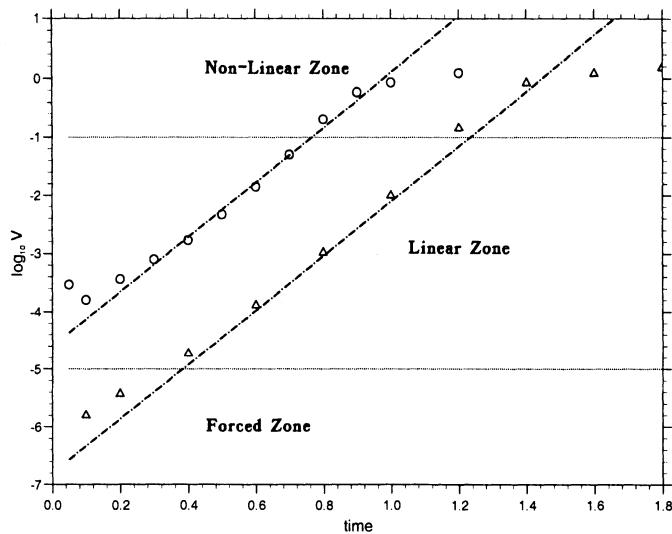


Figure 1. Temporal amplification ($N_y = 20$, $CFL = 0.7$).

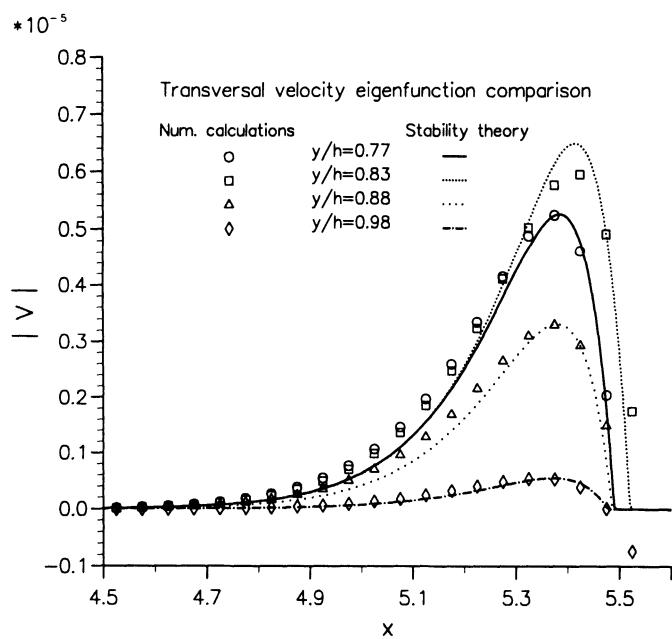


Figure 2. Transversal velocity comparison, ($t=1$ s).

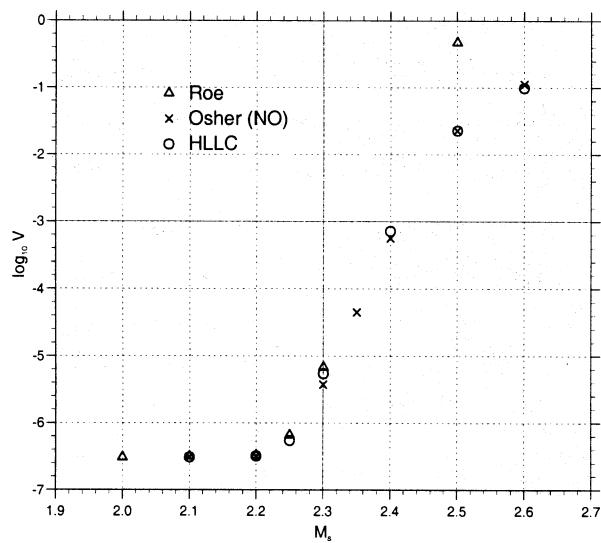


Figure 3. Shock instability threshold for three different upwind schemes

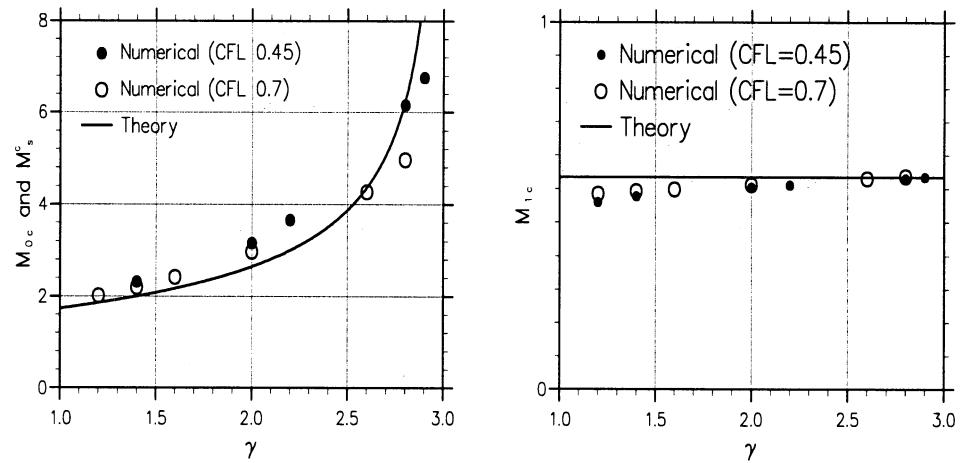


Figure 4. Instability thresholds as a function of γ : upstream Mach number (left), downstream Mach number (right)

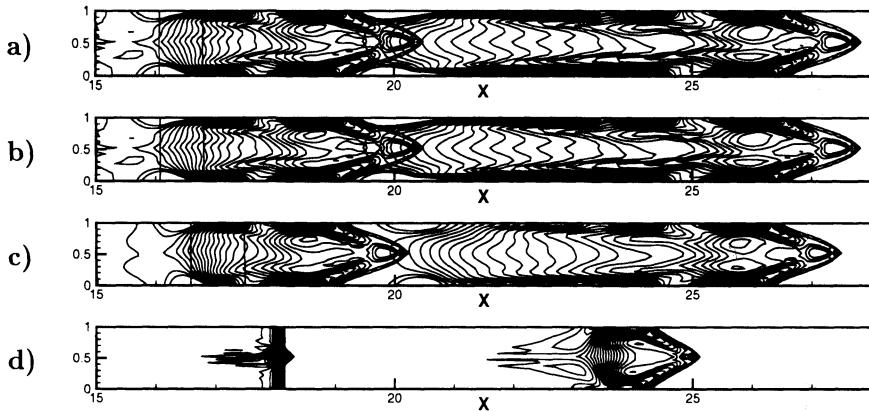


Figure 5. Effect of an entropy fix applied on different waves: none (a), acoustic waves (b), entropy wave (c), vorticity wave (d)

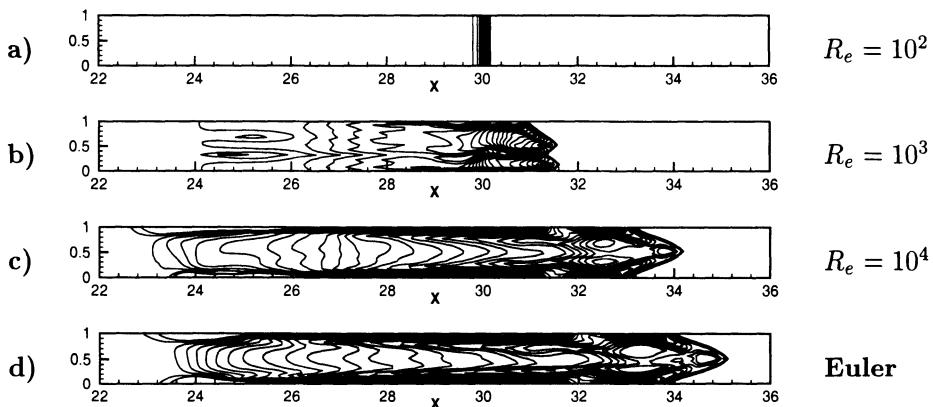


Figure 6. Density contours for the viscous Quirk's problem using different Reynolds numbers.

References

- Gressier J, Moschetta J M (2000). Robustness versus Accuracy in Shock Waves Computations. *International Journal for Numerical Methods in Fluids* **33**, pp 313-332.
 Quirk J J (1994). A Contribution To the Great Riemann Solver Debate. *International Journal for Numerical Methods in Fluids* **18**, pp 555-574.
 Robinet J C, Gressier J, Casalis G, Moschetta J M (2000). Shock Wave instability and carbuncle phenomenon: same intrinsic origin ? *Journal of Fluid Mechanics* **417**, pp 237-263.

A GODUNOV-TYPE SOLVER FOR THE MAXWELL EQUATIONS WITH DIVERGENCE CLEANING

C.-D. MUNZ

*Institut für Aerodynamik und Gasdynamik,
Universität Stuttgart.
Pfaffenwaldring 21, 70550 Stuttgart, Germany
Email: munz@iag.uni-stuttgart.de*

P. OMNES

*CEA Saclay,
DCC/DPE/SPCP,
91191 Gif sur Yvette Cedex, France.
Email: pomnes@cea.fr*

AND

R. SCHNEIDER

*Institut für Hochleistungsimpuls- und Mikrowellentechnik,
Forschungszentrum Karlsruhe — Technik und Umwelt
Postfach 3640, 76021 Karlsruhe, Germany.
Email: Rudolf.Schneider@ihm.fzk.de*

Abstract. We present a high-resolution finite-volume Godunov-type Maxwell solver for three-dimensional unstructured meshes, based on the purely hyperbolic Maxwell (PHM) system, which is established by introducing two additional degrees of freedom into the evolutionary part of the Maxwell equations and coupling them with the elliptical constraints given by Gauß' law and the $\nabla \cdot \mathbf{B} = 0$ statement. This model allows for possible errors in the charge conservation equation as may occur in particle-in-cell simulations, and yields approximative solutions of the conventional Maxwell equations. Numerical results demonstrate the relevance of the correction approach when the charge conservation equation is violated.

1. Introduction

The operation behavior of devices such as microwave tubes and light ion or electron sources is substantially influenced by the interaction of charged particle flow with electromagnetic fields. Modelling the physical phenomena occurring inside these devices requires the numerical solution of the three-dimensional Vlasov-Maxwell equations in the time domain for complex geometries. The most popular numerical approach used to solve this non-linear set of equations is the so-called particle-in-cell (PIC) method (Birdsall and Langdon, 1985). However, it is well-known that the different approximation steps in the PIC procedure introduce numerical errors and that, consequently, the charge conservation equation may not be satisfied on the discrete level. Then, ignoring Gauß' law in the numerical solution of the Maxwell equations may lead to damaging errors in the self-consistent movement of the particles and, especially, long-time computations are not trustworthy. To avoid the increase of numerical errors caused by suppressing the information contained in Gauß' law, we propose in the present paper a finite-volume scheme for the three-dimensional time-dependent Maxwell equations on unstructured grids, in which the divergence cleaning technique we introduce does not require the solution of a Poisson equation, a decisive advantage over the usual one (Birdsall and Langdon, 1985).

2. Governing equations

In the following, we consider a three-dimensional bounded domain $\Omega \subset \mathbb{R}^3$ in vacuum with Cartesian coordinates $\mathbf{x} = (x_1, x_2, x_3) = (x, y, z)$. When ρ and \mathbf{j} satisfy the charge conservation equation ($\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0$) for all times $t \geq 0$, and when the initial fields satisfy the elliptical part of the Maxwell equations, the system consisting of the evolution equations (Maxwell and Faraday equation) has a unique solution in Ω which automatically fulfills Gauß' law and the condition $\nabla \cdot \mathbf{B} = 0$ for all times. However, in numerical computations, a discrete analogue of the charge conservation equation does not hold in general (as known from particle-in-cell simulations) and the initial data may not fulfil the discrete elliptical requirement exactly. To get rid of this numerically caused lack, a generalized mathematical model for the Maxwell equations has been proposed (Munz et al., 1999) recently, allowing errors in the divergence constraints for the initial data as well as in the continuity equation. It is established by introducing two additional variables $\Phi(\mathbf{x}, t)$ and $\Psi(\mathbf{x}, t)$ into the evolution equations of the electromagnetic fields, and coupling them with the elliptical constraints, and is

shortly abbreviated as purely hyperbolic Maxwell (PHM) model:

$$\frac{\partial \mathbf{E}}{\partial t} - c^2 \nabla \times \mathbf{B} + \chi c^2 \nabla \Phi = -\frac{\mathbf{j}}{\epsilon_0}, \quad (1)$$

$$\frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathbf{E} + \gamma \nabla \Psi = 0, \quad (2)$$

$$\frac{1}{\chi} \frac{\partial \Phi}{\partial t} + \nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}, \quad (3)$$

$$\frac{1}{\gamma c^2} \frac{\partial \Psi}{\partial t} + \nabla \cdot \mathbf{B} = 0, \quad (4)$$

where χ and γ are dimensionless positive parameters. This PHM system has to be supplemented with appropriate initial conditions on Γ and in Ω for the electromagnetic \mathbf{E} and \mathbf{B} as well as for the additional variables Φ and Ψ (Munz, Omnes and Schneider, 2000). It is straightforward to check that Φ (resp. Ψ) satisfies a wave equation which advects the possible errors to the boundaries with the velocity χc (resp. γc). Thus, the values of χ and γ may be chosen sufficiently large to ensure that the errors are transported several times faster than the physical information. However, in practical calculations we did not run into problems using $\chi = \gamma = 1$ which means, that the usual CFL condition is not affected. Otherwise a sub-cycling procedure should be introduced as proposed in (Munz et al., 2000). Small approximation errors in the divergence constraints seem not to be a serious problem as long as the local increase is avoided. We note that if the charge conservation equation holds exactly and the divergence conditions are satisfied by the initial fields, the PHM model system is equivalent to the conventional Maxwell equations. Moreover, it was shown in (Munz et al., 1999) that the PHM system along with appropriate initial and boundary conditions admits a unique solution, and that the energy norm of Φ and Ψ is uniformly bounded in time. Consequently, it is guaranteed that the divergence errors do not increase in time and that the conventional Maxwell equations are approximately satisfied by (1)-(4) during the whole computation.

In order to construct a finite-volume scheme, we recast the PHM equations (1)-(4) into the relevant conservation form

$$\frac{\partial \mathbf{u}}{\partial t} + \sum_{j=1}^D \frac{\partial \mathbf{f}_j(\mathbf{u})}{\partial x_j} = \mathbf{g}, \quad (5)$$

where $D = 3$ in the present three-dimensional considerations. The vector of the evolving variables $\mathbf{u} = \mathbf{u}(\mathbf{x}, t)$ is composed by the electric and magnetic fields as well as by the two supplementary variables:

$$\mathbf{u} = (E_1, E_2, E_3, \Psi, B_1, B_2, B_3, \Phi)^T.$$

The physical fluxes \mathbf{f}_j are determined from $\mathbf{f}_j(\mathbf{u}) = \mathcal{K}_j \mathbf{u}$ with $j = 1, 2, 3$, where the block-structured matrices $\mathcal{K}_j \in \mathbb{R}^{8 \times 8}$ possess constant entries. The source of the conservation equation (5) is given by

$$\mathbf{g} = -\frac{1}{\epsilon_0} (j_1, j_2, j_3, 0, 0, 0, 0, -\chi\rho)^T$$

and contains the current as well as the charge density.

Some properties of the linear combination

$$\mathcal{A} = \sum_{j=1}^D n_j \mathcal{K}_j, \quad \text{with} \quad \sum_{j=1}^D n_j^2 = 1,$$

are essential. The matrix \mathcal{A} possesses a complete set of linearly independent right eigenvectors associated to its eight real eigenvalues

$$(c, c, -c, -c, \gamma c, -\gamma c, \chi c, -\chi c).$$

The matrices of the right and left eigenvectors $\mathcal{R} = (\mathbf{r}_1, \dots, \mathbf{r}_8)$ and $\mathcal{R}^{-1} = (\mathbf{l}_1, \dots, \mathbf{l}_8)^T$ can be expressed as functions of an orthonormal basis and are explicitly given in (Munz, Omnes and Schneider, 2000). As all eigenvalues of \mathcal{A} are real numbers and as its right eigenvectors form a basis of \mathbb{R}^8 , the PHM system is strictly hyperbolic.

3. Numerical approximation of the PHM system

In the following section, we introduce a finite-volume (FV) scheme on unstructured mesh arrangements for the PHM equations introduced above. The bounded domain of computation Ω is partitioned into N non-overlapping cells $(C_i)_{i \in [1, N]}$. The surface ∂C_i bounding each grid zone C_i consists of σ_i faces $S_{i,\alpha}$ with area $L_{i,\alpha}$, where α runs from one to σ_i . The solution is computed at a set of discrete $t^n = n \Delta t$, where Δt is determined with respect to the CFL condition. In order to solve the inhomogeneous PHM evolution equations (5) numerically, we apply the so-called Strang (Strang, 1968) splitting for the source term, which preserves second-order accuracy with respect to time. Now, we consider the integration of the homogeneous conservation equations (5) over the space-time element $C_i \times [t^n, t^{n+1}]$. Applying Gauß theorem, we obtain an exact evolution equation, which approximation yields the FV scheme, usually written in the form

$$\mathbf{u}_i^{n+1} = \mathbf{u}_i^n - \frac{\Delta t}{V_i} \sum_{\alpha=1}^{\sigma_i} \mathbf{G}_{i,\alpha}^{n+1/2}, \quad (6)$$

where V_i denotes the volume of C_i and \mathbf{u}_i^n is the cell average of \mathbf{u} at t^n . This scheme is completely defined if the numerical flux $\mathbf{G}_{i,\alpha}^{n+1/2}$ is specified as an

approximation of the physical flux through the boundary face $S_{i,\alpha}$ in the normal direction. This may be realized in several ways, like the Godunov-type schemes or the flux-vector splitting approaches, which lead, for linear evolution equations, to the same numerical flux:

$$\mathbf{G}_{i,\alpha}^{n+1/2} = L_{i,\alpha} \left(\mathcal{A}_{i,\alpha}^+ \mathbf{u}_L + \mathcal{A}_{i,\alpha}^- \mathbf{u}_R \right). \quad (7)$$

Here, \mathbf{u}_L and \mathbf{u}_R are approximate states at the interface $S_{i,\alpha}$ of C_i and its neighbouring cell $C_{\nu_{i,\alpha}}$, respectively. The matrices \mathcal{A}^\pm are explicitly given by

$$\mathcal{A}^\pm = \begin{pmatrix} \mathcal{D}^\pm(\chi, \gamma) & c^2 \mathcal{E}(\chi, \gamma) \\ \mathcal{E}^T(\chi, \gamma) & \mathcal{D}^\pm(\gamma, \chi) \end{pmatrix} \quad (8)$$

with

$$\mathcal{E}(\chi, \gamma) = \begin{pmatrix} 0 & n_3 & -n_2 & n_1 \chi \\ -n_3 & 0 & n_1 & n_2 \chi \\ n_2 & -n_1 & 0 & n_3 \chi \\ n_1 \gamma & n_2 \gamma & n_3 \gamma & 0 \end{pmatrix} \quad (9)$$

and

$$\mathcal{D}(\mu, \nu) = \begin{pmatrix} [1 + n_1^2(\mu - 1)]c & n_1 n_2 (\mu - 1)c & n_1 n_3 (\mu - 1)c & 0 \\ n_1 n_2 (\mu - 1)c & [1 + n_2^2(\mu - 1)]c & n_2 n_3 (\mu - 1)c & 0 \\ n_1 n_3 (\mu - 1)c & n_2 n_3 (\mu - 1)c & [1 + n_3^2(\nu - 1)]c & 0 \\ 0 & 0 & 0 & \nu c \end{pmatrix}, \quad (10)$$

where we dropped the index "i, α " and used the abbreviation $\mathcal{D}^\pm(\mu, \nu) = \pm \mathcal{D}(\mu, \nu)$. As seen from (7), the total flux $\mathbf{G}_{i,\alpha}^{n+1/2}$ through the face $S_{i,\alpha}$ is composed of a flux coming from the "left", having positive eigenvalues only, and a flux from the "right", having negative eigenvalues only, respectively associated to $\mathcal{A}_{i,\alpha}^-$ and $\mathcal{A}_{i,\alpha}^+$. Furthermore, the accuracy of the numerical flux (7) and, hence, the order of the FV scheme (6) depends on the choice of \mathbf{u}_L and \mathbf{u}_R . If the approximate average values at t^n are taken, the scheme is only first-order accurate in both space and time. Second-order accuracy is obtained by applying the MUSCL approach proposed by van Leer (van Leer, 1979). This ansatz relates \mathbf{u}_L and \mathbf{u}_R with the cell averages \mathbf{u}_i^n and $\mathbf{u}_{\nu_{i,\alpha}}^n$ and the gradients at the barycenters B_i and $B_{\nu_{i,\alpha}}$ of the two adjacent grid cells C_i and $C_{\nu_{i,\alpha}}$. (see (Munz, Omnes and Schneider, 2000)).

4. Numerical results

To demonstrate the qualities of the FV method on unstructured grid arrangements for the PHM model, we discuss a numerical calculation where the charge conservation equation is artificially violated. We restrict ourselves here to two space dimensions and set $\gamma = 0$ and $\chi = 1$. Furthermore,

we assume that the current density vanishes during the whole course of the numerical simulation. The violation of the continuity equation is explicitly accomplished by introducing a square-shaped charge density distribution around the center of the computational domain $\Omega = [0, 1] \times [0, 1]$ defined by $\rho(x, y, t) = \rho_0 \omega t F(x, y)$, $x \in \Omega$, $t \geq 0$, with some constants ρ_0 and ω . The function F is given by $F(x, y) = [H(x - x_1) - H(x - x_2)] \cdot [H(y - y_1) - H(y - y_2)]$, where H denotes the Heaviside function and $x_1 = y_1 = 0.449$ m and $x_2 = y_2 = 0.551$ m are the coordinates of the lower left and upper right corner of the charge distribution. The electromagnetic fields are set to zero initially and we impose absorbing boundary conditions on the four edges of Ω during the whole computation.

Under these circumstances, a traditional Maxwell solver based on the evolution equations only, will compute $\mathbf{E}(\mathbf{x}, t) = \mathbf{B}(\mathbf{x}, t) = \mathbf{0}$ as the solution inside Ω for $t > 0$ and will, consequently, ignore the existence of the electric field due to the charge density. Such a solution is not expected from the proposed PHM field solver, because the information available from Gauß' law is incorporated into the computation via the variable Φ . This correction mechanism, inherently provided by the PHM solver, transports the errors due to the violation of the charge conservation out of the computational domain with the finite propagation velocity χc , and the numerical solution tends toward the correct physical one. This is demonstrated by Figure 1, where two snapshots, recorded at $t = 1$ ns (upper row) and $t = 2$ ns (lower row), of the spatial distribution of the E_1 field and the function Φ are seen. In this and the subsequent numerical simulations, the function Φ is set to zero initially and at the border of the domain Ω during the whole computation. Plots of the quantity $\frac{1}{\chi} \|\frac{\partial \Phi}{\partial t}\|_{L_2} = \|\nabla \cdot \mathbf{E} - \frac{\rho}{\epsilon_0}\|_{L_2}$ are shown by Figure 2. On the left plot, $\chi = 10^{-6}$ and, consequently, the PHM field solver is approximately equivalent to a traditional Maxwell solver where no charge correction is applied, and we observe the expected linear growth of the L_2 norm of the error in Gauß' law. On the right plot, $\chi = 1$ and we observe that the L_2 -norm $\|\nabla \cdot \mathbf{E} - \frac{\rho}{\epsilon_0}\|_{L_2}$ decays in an oscillatory manner till it reaches after approximately 50 ns a stable and bounded state. This profile indicates the relevance and the essential property of our new approach: The PHM solver perceives the inconsistency between the current and charge densities and restores the divergence equation up to a certain stabilized level which is bounded in time, leading finally to a very reasonable solution for the model problem on hand which may be considered as a "worst case problem" in the context of self-consistent charged particle treatment in electromagnetic fields.

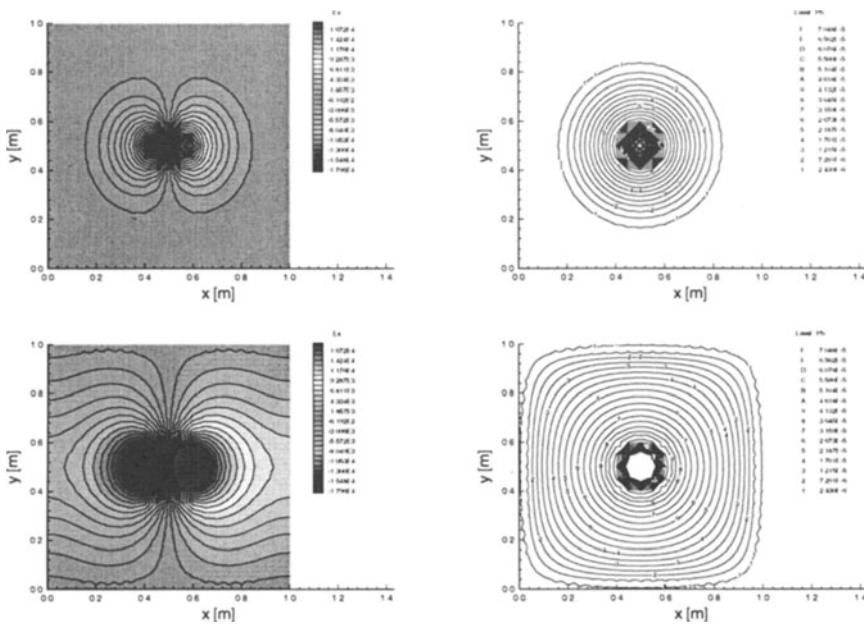


Figure 1. Plots of the E_1 component of the electric field (left) and of the correcting variable Φ (right) at $t = 1 \text{ ns}$ (upper row) and $t = 2 \text{ ns}$ (lower row).

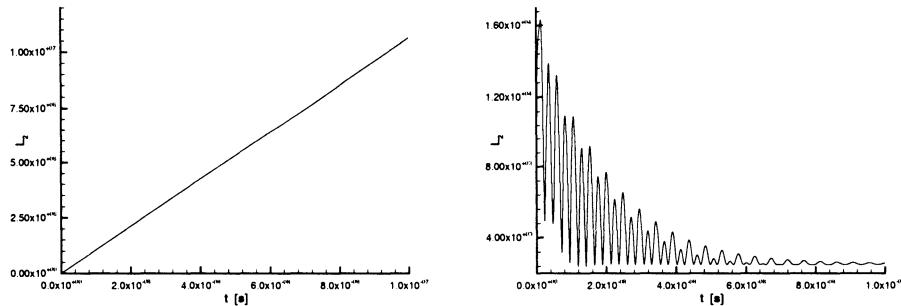


Figure 2. Temporal evolution of $\|\nabla \cdot \mathbf{E} - \frac{\rho}{\epsilon_0}\|_{L_2}$. Left plot: uncorrected case; right plot: corrected case.

References

- Birdsall C K and Langdon A B (1985). *Plasma Physics via Computer Simulation*. McGraw-Hill.
 Munz C-D, Schneider R, Sonnendrücker E and Voß U (1999). Maxwell's Equations when the Charge Conservation Equation is not satisfied. *C. R. Acad. Sci. Paris.* **328**. pp

- 431-436.
- Munz C-D, Omnes P and Schneider R (2000). A three-dimensional finite-volume solver for the Maxwell equations with divergence cleaning on unstructured meshes. *Accepted for publication in Comp. Phys. Comm.*
- Munz C-D, Omnes P, Schneider E, Sonnendrücker E and Voß U (2000). Divergence correction techniques for Maxwell solvers based on a hyperbolic model. *Accepted for publication in J. Comput. Phys.*
- Strang G (1968). On the construction and comparison of difference schemes. *SIAM J. Num. Anal.*. **5**. pp 505-517.
- van Leer B (1979). Towards the ultimative conservative difference scheme V, A second order sequel to Godunov's method. *J. Comput. Phys.*. **32**. pp 101-136.

CONVERGENCE OF KINETIC APPROXIMATION TO NONLINEAR PARABOLIC PROBLEMS

G. NALDI

*Dipartimento di Matematica e Applicazioni,
Università di Milano-Bicocca,
via Bicocca degli Arcimboldi 8, I-20126 Milano, Italy
Email: naldi@matapp.unimib.it*

L. PARESCHI

*Dipartimento di Matematica,
Università di Ferrara,
via Machiavelli 35, I-44100 Ferrara, Italy
Email: pareschi@dm.unife.it*

AND

G. TOSCANI

*Dipartimento di Matematica,
Università di Pavia,
via Ferrata 1, I-27100 Pavia, Italy
Email: toscani@dimat.unipv.it*

Abstract. In a recent paper (Naldi et. al., 1999) we have introduced a general way of constructing hyperbolic relaxation approximations to nonlinear system of parabolic equations. Here, we propose a way to construct one-dimensional relaxing systems which have a clear interpretation as kinetic systems with two velocities with total mass that relaxes towards the solution to the parabolic equation strongly in $L^1_{loc}(\mathbb{R})$. Numerical results for a Porous-Fischer's equation are also presented.

1. Introduction

The numerical passage from hyperbolic systems with diffusive relaxation towards the corresponding parabolic equilibrium limit equations has been

studied recently by the authors (Naldi and Pareschi, 2000), (Naldi et. al., 1999).

For many of such systems the relaxing hyperbolic equations have a clear kinetic interpretation as generalized Carleman models (Carleman, 1957) or Broadwell models (Broadwell, 1964) with source terms. A theoretical approach that justifies the passage from the kinetic system to the second order (nonlinear) parabolic equation has been developed mainly in (Lions and Toscani, 1997) for a generalized Carleman model, in absence of sources, and in (Gabetta and Perthame, 1996) for the so-called Ruijgrok-Wu model of the Boltzmann equation. Here we extend their analysis to cover the diffusive limit towards a parabolic equation of the form

$$\partial_t u + \partial_x f(u) = \partial_{xx} p(u) + s(u), \quad (1)$$

where $(x, t) \in \mathbb{R} \times \mathbb{R}^+$, p , f and s are given smooth functions such that $p(0) \geq 0$ and $p'(u) > 0$, with initial data

$$u(x, 0) = u_0(x). \quad (2)$$

By introducing a new variable v , one can couple v and u through a linear hyperbolic system, which in the simplest situation reads

$$\begin{aligned} \partial_t u + \partial_x v &= s(u), \\ \partial_t v + \frac{1}{\epsilon^2} \partial_x u &= -\frac{1}{\epsilon^2} k(u)(v - f(u)), \end{aligned} \quad (3)$$

with the additional initial condition $v(x, 0) = v_0(x)$. Here ϵ is a small positive parameter called the *relaxation time* and $k(u) = (p'(u))^{-1}$. As usual when $\epsilon \ll 1$ system (3) is said to be *stiff*.

In the small relaxation limit, $\epsilon \rightarrow 0^+$, the relaxation system (3) formally approximates to leading order

$$v = f(u) - \partial_x p(u), \quad (4)$$

$$\partial_t u + \partial_x f(u) = \partial_{xx} p(u) + s(u). \quad (5)$$

As usual, the state satisfying (4) is called the *local equilibrium* while (5) corresponds to the equation of continuum mechanics generated by the kinetic model.

It is clear that the relaxing system (3) is not unique, and that a proof of the convergence of the system to (1) is deeply connected with the particular choice of the relaxing system. In the next section we will propose a general way to construct a kinetic system with two velocities with total mass that relaxes towards the solution to equation (1) strongly in $L^1_{loc}(\mathbb{R})$.

With respect to the previous studies of (Lions and Toscani, 1997) and (Gabetta and Perthame, 1996) on classical discrete kinetic models, the relaxation analysis of systems of type(3) contain some remarkable differences. First of all, due to the source terms, there is no mass conservation. Second, there are no entropy principles. We remark that both mass conservation and entropy decay have been at the basis of the analysis on Lions and Toscani (Lions and Toscani, 1997) and (Gabetta and Perthame, 1996). Hence, their analysis is not directly applicable to the present situation.

The main advantage of numerically solving the relaxation system (3) over the original nonlinear convection-reaction-diffusion equation (1) lies in the localized lower order term. This permits to reduce a nonlinear second order system to a semi-linear first order one and to avoid the use of Riemann solvers. However, as observed recently (Jin et. al., 1999), (Jin et. al., 2000), (Naldi and Pareschi, 1998), (Naldi and Pareschi, 2000), the numerical solution to (3) is challenging due to the stiffness of the problem for both the convection and the relaxation terms. In our numerical examples, we will make use of the splitting algorithm proposed in (Jin et. al., 1999). This method has been shown to be robust in the diffusive limit providing a stability restriction independent of the small relaxation parameter and avoiding the solution of nonlinear system of algebraic equations.

2. Kinetic Relaxation Approximation to Parabolic Equations

The present section deals with the construction of a kinetic system with two velocities with mass that approach in the limit $\epsilon \rightarrow 0$ the solution to (1). As premised in the introduction, a kinetic relaxing system can be constructed in many ways. Looking at the previous papers concerned with the passage to the limit in the diffusive scaling (Lions and Toscani, 1997), (Gabetta and Perthame, 1996), (Muller and Ruggeri, 1999), we observe that this limit procedure can be rigorously justified if the kinetic system under consideration is such that the solution to the initial value problem for a system of type (3) corresponding to initial data of bounded variation is of bounded variation, and the family $\{u_\epsilon\}$ of mass densities is uniformly bounded in $BV(\mathbb{R})$ in any interval of time, while the family $\{v_\epsilon\}$ of fluxes is uniformly bounded in $L^2(\mathbb{R} \times [0, T])$. In fact, these conditions ensure the passage to the limit for the density in strong L^1 -norm.

Having this in mind, we consider the kinetic system

$$\begin{aligned}\partial_t U + \frac{1}{\epsilon} \partial_x U &= \frac{1}{2\epsilon^2} [k(U + V)(V - U) + 2\epsilon Uh(U + V) + 2\epsilon^2 Vg(U + V)], \\ \partial_t V - \frac{1}{\epsilon} \partial_x V &= \frac{1}{2\epsilon^2} [k(U + V)(U - V) - 2\epsilon Uh(U + V) + 2\epsilon^2 Ug(U + V)].\end{aligned}\tag{6}$$

System (6) models a fictitious gas composed of two kinds of particles that can move parallel to the x -axis with constant and equal speeds, either in the positive x -direction with a density U , or in the negative x -direction with a density V . To maintain positivity, the function h is required to be nonnegative. The right-hand side represents the variation of the densities due to collisions. We have here four types of collisions $+ \rightarrow -$ and $- \rightarrow +$ with coefficient $2/\epsilon^2$, $+- \rightarrow +$ and $-- \rightarrow -$ with coefficients respectively $2/\epsilon$ and 2. The third type of collision represent a source term for both densities.

The macroscopic equations generated by system (6) are

$$\begin{aligned} \partial_t u + \partial_x v &= ug(u), \\ \partial_t v + \frac{1}{\epsilon^2} \partial_x u &= -\frac{1}{\epsilon^2} [k(u)v - uh(u) - \epsilon vh(u) - \epsilon^2 vg(u)]. \end{aligned} \tag{7}$$

In the small relaxation limit we have the local equilibrium

$$v = \frac{1}{k(u)} [uh(u) - \partial_x u]. \tag{8}$$

It is clear that, if

$$g(u) = \frac{s(u)}{u}, \quad h(u) = \frac{f(u)k(u)}{u}, \quad k(u) = \frac{1}{p'(u)}, \tag{9}$$

at least formally, the relaxation system (7) approximates to leading order (5). We want to prove that this limit can be proven rigorously, as far as we impose conditions on the functions k, h and g , that we fix smooth from now on. In what follows we suppose $k(u) \geq 0$, $u \geq 0$, strictly increasing from $k(0) = 0$ or strictly decreasing, with $k(0)$ unbounded, $h(u) \geq 0$, $u \geq 0$, strictly increasing from $h(0) = 0$ and

$$\inf_{[0, \bar{u}]} u \frac{h'(u)}{h(u)} = M(\bar{u}), \quad \bar{u} > 0, \tag{10}$$

with $M(\bar{u}) > 0$ ($h(u) = u^p$, $p > 1$ is a typical example). Finally, let $g(u) \geq 0$ when $0 \leq u \leq u_{max}$.

System (7) is quite general. It contains many different models, and with it we can approach a variety of different linear or nonlinear equations of parabolic type. For example, the choice $k(u) = 1$, $h(u) = u$, $g(u) = 0$ gives in the limit Burgers equation

$$\partial_t u + \partial_x u^2 = \partial_{xx} u, \tag{11}$$

while the choice $k(u) = 1$, $h(u) = 0$ and $g(u) = 1 - u$ gives in the limit Fisher's equation

$$\partial_t u = \partial_{xx} u + u(1 - u). \quad (12)$$

The main result is the following

Theorem 2.1 *Let $0 \leq U_0, V_0 \leq M \in BV(\mathbb{R})$, and let us set*

$$\omega_\delta(x) = (1 + x)^\delta, \quad 0 < \delta < \frac{1}{4},$$

with u_0 such that

$$\int_{\mathbb{R}} \omega_\delta(x) u_0(x) dx = c_\delta < \infty.$$

Let k, h and g satisfy the previous growth conditions, and let us consider the global solution $U^\epsilon(x, t), V^\epsilon(x, t)$ to system (7), corresponding to initial data $0 \leq U_0, V_0 \in BV(\mathbb{R}) \cap L^1$. Then

$$u^\epsilon(x, t) \rightarrow u(x, t)$$

in $L^1(\mathbb{R})$

$$\epsilon v^\epsilon(x, t) \rightarrow 0$$

in $L^2(\mathbb{R} \times [0, T])$.

The limit $u(x, t)$ is a solution to the equation (1).

The complete proof is an extension of analogous ones for a generalized Ruijgrok-Wu model, namely the model obtained for $g(u) = 0$ (see (Gabetta and Perthame, 1996)) and for the case of the limiting Fisher's equation (see (Cavazzoni, 1999)). The main argument is to show uniform BV -bounds, and the existence of a entropy function by which one gets uniform L^2 -bounds for the flux. The presence of the source term, while dropping the mass conservation and the time monotonicity of the entropy, does not destroy the uniform bounds both for the mass and for the flux. The details will be presented elsewhere.

3. Numerical results

In this section we present a numerical test for a nonlinear reaction-diffusion equation, namely the following Porous-Fisher's equation (Newman, 1980), (Witelski, 1995), (Smoller, 1983)

$$\partial_t u = \partial_x (u^\alpha \partial_x u) + u(1 - u^\alpha), \quad (13)$$

on $-\infty < x < \infty$, where α is a positive real parameter.
Such equation can be written in the form (1) by taking

$$f(u) = 0, \quad p(u) = \frac{1}{\alpha+1} u^{\alpha+1}, \quad s(u) = u(1-u^\alpha).$$

Equations of Porous-Fisher type arise, for example, in some model of combustion phenomena and in population dynamics when the interaction between individuals and the behavior of the population flow depend on the density function $u(x, t)$ (Gurney and Nisbet, 1975), (Murray, 1990).

Newman (Newman, 1980) has shown that (13) has the unique traveling wave solution

$$u(x, t) = \left(\left(1 - \exp \left[\frac{\alpha z}{\sqrt{(\alpha+1)}} \right] \right) \right)^+, \quad z = x - ct, \quad (14)$$

where $(x)^+ \equiv \max(x, 0)$ and the velocity c is defined as

$$c = \frac{1}{\sqrt{(\alpha+1)}}.$$

Using suitable translations and reflections (Witelski, 1995) of the wave (14) it is possible to write a family of solutions of (13). For example, we can construct a traveling wave $u_0(x, t)$ moving to the right, starting from some point x_0 , and a traveling wave $u_1(x, t)$ starting from the point x_1 and moving to the left. Using these two wave solutions, we can produce merging populations. While the two single populations, represented by $u_0(x, t)$, and $u_1(x, t)$, remain separated there will be no interaction and the whole population density, at each point, is merely the maximum between them. At time $\hat{t} = (x_1 - x_0)/2c$ and position $\hat{x} = (x_0 + x_1)/2$, the two populations first meet and begin their interaction. An analysis of such a behavior was done by Witelski (Witelski, 1995) by using perturbation theory and the method of matched asymptotic expansions.

The approximation of the macroscopic system (7), which has the Porous-Fisher's equation as the limit state, was performed by using the splitting scheme described in (Jin et. al., 1999), (Jin et. al., 2000), (Naldi and Pareschi, 2000). Such a scheme is obtained as a generalization of the relaxation schemes, recently introduced by Jin and Xin (Jin, 1995), (Jin and Xin, 1995), in the case of a kinetic system with diffusive relaxation (Jin et. al., 1999), (Naldi and Pareschi, 1998). Taking $\alpha = 2$ in the original equation (13) the velocity c of the traveling waves, the critical time t and the critical position \hat{x} are given, respectively, by $c = 1/\sqrt{3}$, $\hat{t} = 15\sqrt{3}$ and $\hat{x} = 5$. The system is solved in the interval $x \in [-20, 30]$ with suitable fixed values of density $u(x, t)$ at the boundary which agree with the values of the interacting waves. In Figure (1) is shown the short time, no interaction

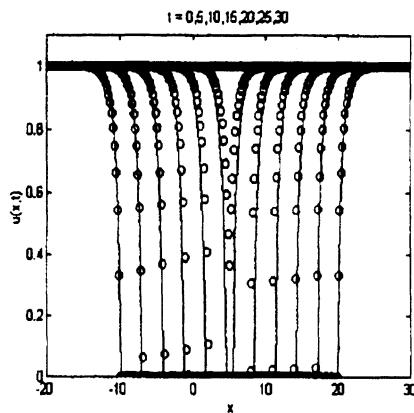


Figure 1. Evolution of two traveling waves solutions of the Porous-Fisher's equation for different times before the interaction.

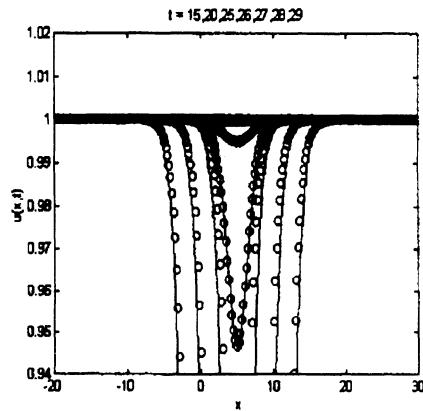


Figure 2. Evolution of two travelling waves solutions of the Porous-Fisher's equation for different times after the critical time \hat{t} when the two waves first meet.

behavior of the waves, while in Figure (2) the long time behavior of the numerical solution is plotted. In the computations we have used an uniform space-time mesh where $\Delta x = 0.2$, $\Delta t = 0.01$ and $\epsilon = 10^{-8}$. As the true solution, represented by a continuous line in the figures, we have considered the asymptotic solution computed by Witelski in (Witelski, 1995).

Acknowledgements

This work was supported by Progetto Nazionale di Ricerca MURST "Analisi numerica: metodi e software matematico" and by the European TMR Project Asymptotic Methods in Kinetic Theory, contract ERBFRMXCT 970157.

References

- J.E. Broadwell (1964). Shock structure in a simple discrete velocity gas. *Phys. Fluids* **7** (8), pp 1243-1247.
- T. Carleman (1957). Problèmes mathématiques dans la théorie cinétique des gaz. (French) Publ. Sci. Inst. Mittag-Leffler. 2 Almqvist & Wiksell Boktryckeri Ab, Uppsala.
- R. Cavazzoni (1999). Diffusive approximation of Fisher's equation. *Computers and Mathematics with Applications*, in press.
- E. Gabetta and B. Perthame (1996). Scaling limits of the Ruijgrok-Wu model of the Boltzmann equation. Proc. of the International Conference on Nonlinear Equations and Applications, Bangalore 19-23 August 1996. Springer-Verlag.
- W.S. Gurney and R.M. Nisbet (1975). The regulation of inhomogeneous populations. *J. Theor. Biol.*, **52**, pp 441-457.
- S. Jin (1995). Runge-Kutta methods for hyperbolic conservation laws with stiff relaxation terms. *J. Comput. Phys.* **122**, pp 51-67.
- S. Jin, L. Pareschi and G. Toscani (1999). Diffusive relaxation schemes for multiscale discrete-velocity kinetic equations. *SIAM J. Num. Anal.* **35**, pp 2405-2439.
- S. Jin, L. Pareschi and G. Toscani (2000). Uniformly accurate diffusive relaxation schemes for transport equations. *SIAM J. Num. Anal.* to appear.
- S. Jin and Z. Xin (1995). The relaxation schemes for systems of conservation laws in arbitrary space dimension. *Comm. Pure and Appl. Math.* **48**, pp 235-276.
- P.L. Lions and G. Toscani (1997). Diffusive limit for two-velocity Boltzmann kinetic models. *Revista Matematica Iberoamericana* **13**, pp 473-513.
- I. Muller and T. Ruggeri (1999). Rational Extended Thermodynamics. Springer-Verlag.
- J.D. Murray (1990). Mathematical Biology. Springer-Verlag, Berlin.
- G. Naldi and L. Pareschi (1998). Numerical schemes for kinetic equations in diffusive regimes. *Appl. Math. Letters* **11**, pp 29-35.
- G. Naldi and L. Pareschi (2000). Numerical schemes for hyperbolic systems of conservation laws with stiff diffusive relaxation. *SIAM J. Num. Anal.* **37**, pp 1246-1270.
- G. Naldi, L. Pareschi and G. Toscani (1999). Hyperbolic relaxation approximation to nonlinear parabolic problems, pp 747-756. Intern. Series of Numerical Math. **130**, Birkhäuser Verlag.
- W.I. Newman (1980). Some exact solutions to a nonlinear diffusion problem in population genetics and combustion. *J. Theor. Biol.* **85**, pp 325-334.
- A. Pulvirenti and G. Toscani (2000). A kinetic formulation of the fast diffusion equation. *Math. Mod. and Meth. in App. Scie.*, to appear.
- J. Smoller (1983). Shock waves and reaction-diffusion equations. Springer-Verlag, New-York.
- T.P. Witelski (1995). Merging traveling waves for the Porous-Fisher's equation. *Appl. Math. Lett.* **8**, pp 57-62.

ON OPTIONS FOR THE NUMERICAL MODELLING OF THE DIFFUSION TERM IN RIVER POLLUTION SIMULATIONS

S. NEELZ

*Royal Institute of Technology,
Stockholm, Sweden.
Email: sneelz@voila.fr*

S. G. WALLIS

*Heriot-Watt University, UK.
Email: stevew@civ.hw.ac.uk*

AND

J.R. MANSON

*Rensselaer Polytechnic Institute, Troy, USA.
Email: mansoj@rpi.edu*

Abstract. In this paper five numerical methods for modelling the diffusion term in a one-dimensional advection-diffusion equation are compared. The motivation behind the work is to find a computationally efficient method for modelling diffusion for incorporating in a semi-Lagrangian approach for advection. Three of the schemes are traditional Eulerian implicit methods (backward, Crank-Nicholson, optimised time-weighted); the other two are based on work by Teixeira (Teixeira, 1998) who proposed a method based on a discrete view of diffusion.

Numerical tests were undertaken in a non-advective uniform flow and concerned the evolution of an initial Gaussian spatial distribution of a conservative solute. Tests covered a range of initial spatial resolutions and a range of dimensionless diffusion numbers. For each test, the number of time steps required until the numerical solution agreed to within a specified tolerance of the exact solution to the problem was noted.

The results showed that an optimized time-weighted implicit scheme (used with a weighting coefficient taking values close to 0.7) was twice as efficient as the backward implicit scheme. Teixeira's original method proved to be rather inefficient in general, but a modified form of it was as efficient as the optimum time-weighted scheme. The Crank-Nicholson scheme was

the most efficient unless grid-scale oscillations were present, when it became very inefficient.

1. Introduction

Mathematical models of river pollution incidents are usually based on a one-dimensional advection-diffusion equation which describes the longitudinal transport of a solute in terms of simultaneous translation (advection) and spreading (diffusion). In the Water Industry, this equation is known as the Advection-Dispersion Equation (ADE) because the spreading is caused by hydrodynamic dispersion. However, it is modelled by an analogy to Fickian diffusion (Taylor, 1954) (Fischer et al., 1979), and since the results presented here may be relevant to the simulation of diffusion in general, "diffusion" rather than "dispersion" is used in the remainder of this paper.

The motivation behind the work reported here is to find a computationally efficient method for modelling the diffusion term in the ADE for using within a semi-Lagrangian approach for advection developed previously namely, DISCUS (Wallis and Manson, 1997). The use of a semi-Lagrangian method for rivers is logical because they tend to be advection dominated and, following the lead of other application areas of computational fluid dynamics (e.g. numerical weather forecasting (Staniforth and Côté, 1991) and groundwater transport (Celia et al., 1990)), the potential advantages of such an approach are now being realised. In these other fields Eulerian finite difference treatments of diffusion have become the norm, but for the river case the authors' earlier studies indicated some advantages and disadvantages of these schemes. For example, explicit central differences (Wallis et al., 1998) imposed a relatively severe restriction on computational efficiency by requiring small time steps for reasons of stability. Backward implicit central differences (Wallis et al., 1999) were more robust in this regard, of course. However, the accuracy of the simulation inevitably degraded as the time step increased and this indicated that the full potential of the semi-Lagrangian treatment of advection might not be realised when undertaking simulations of the ADE. Time centred central differences (Crank-Nicholson) although formally more accurate are prone to grid-scale oscillations, particularly when large time steps are used.

Teixeira (Teixeira, 1998) proposed an alternative modelling approach, based on a discrete view of diffusion, which promised good accuracy and stability at large time steps. Teixeira viewed his method as being analogous to a semi-Lagrangian method for advection. However, it is also closely aligned to simulations of diffusion based on random walk theory.

The objective of this paper is to report on the computational efficiencies of three traditional Eulerian methods for the simulation of diffusion and two methods based on Teixeira's approach.

2. Numerical schemes

The ADE is given below for the case of steady uniform flow:

$$\frac{\partial C}{\partial t} + u \frac{\partial C}{\partial x} = D \frac{\partial^2 C}{\partial x^2} \quad (1)$$

where C is the solute concentration, u is the flow velocity , D is the diffusion coefficient, x is the spatial co-ordinate and t is time.

2.1. EULERIAN METHODS

The Eulerian methods used are all implicit. They are all well-known and can be summarised using the following general algorithm:

$$\begin{aligned} -\theta\alpha C_{j-1}^{n+1} + (1 + 2\theta\alpha)C_j^{n+1} - \theta\alpha C_{j+1}^{n+1} = \\ (1 - \theta)\alpha C_{j-1}^n + (1 - 2(1 - \theta)\alpha)C_j^n + (1 - \theta)\alpha C_{j+1}^n \end{aligned} \quad (2)$$

where $\alpha = D\Delta t/(\Delta x)^2$ and θ is a weighting coefficient which determines the degree of implicitness of the scheme. Here Δt and Δx are the time step and distance step, respectively. When $\theta = 0.5$ we have the Crank-Nicholson scheme and when $\theta = 1$ we have the backward implicit scheme. The general scheme described by equation (2) is of first order accuracy and is unconditionally stable provided that $\theta > 0.5$. The Crank-Nicholson scheme is unconditionally stable and is of second order accuracy. However, it suffers from grid-scale oscillations, increasingly so with increasing values of α . Hence, it is common practice to use equation (2) with $0.5 < \theta < 1$.

2.2. TEIXEIRA AND RELATED METHODS

Teixeira (Teixeira, 1998) suggested that diffusion could be modelled by an explicit method that is stable for all values of α . He derived it by analogy with a semi-Lagrangian treatment of advection and it has the following simple update algorithm:

$$C_j^{n+1} = \frac{C_l^n + C_r^n}{2} \quad (3)$$

With this, the "arrival" concentration in control volume j at time n+1 is calculated as the mean of "departure" concentrations at time n from

locations an equal distance, Δs , to the left (l) and to the right (r) of the control volume j, where:

$$\Delta s = \sqrt{2D\Delta t} \quad (4)$$

Equation (4) infers that:

$$\frac{\Delta s}{\Delta x} = \sqrt{2\alpha} \quad (5)$$

which is a dimensionless distance expressed as a function of α . Since equation (5) does not, in general, yield an integer the "departure" concentrations are interpolated from the spatial concentration distribution at time n. Some results with linear and cubic interpolation are reported in (Teixeira, 1998).

A more classical way to derive this method is to make use of the particle random walk theory (see e.g. (Einstein, 1905) or (Van Dam, 1994)). In this, the necessary and sufficient conditions for a distribution Φ of random steps Δ to be suitable for modelling diffusion are:

$$D = \frac{1}{\Delta t} \int_{-\infty}^{\infty} \frac{\Delta^2}{2} \Phi(\Delta) d\Delta \quad (6)$$

which implies that the variance of the distribution increases by $2D\Delta t$ during one time step; $\int_{-\infty}^{\infty} \Phi(\Delta) d\Delta = 1$ and $\Phi(\Delta) = \Phi(-\Delta)$ which are the condition of conservativeness, and the condition of symmetry, respectively.

Now we see that Teixeira's method is simply a control volume interpretation of the random walk method in which the distribution $\Phi(\Delta)$ is defined by $\Phi(\Delta) = 0.5$ for $\Delta = \pm\Delta s$ and $\Phi(\Delta) = 0$ otherwise. The highly discrete nature of this is a rather poor model of the diffusion process.

A more continuous (and therefore theoretically better) treatment of diffusion is obtained using the following definition of the distribution $\Phi(\Delta)$, namely: $\Phi(\Delta) = (2\Delta d)^{-1}$ for $-\Delta d \leq \Delta \leq \Delta d$ and $\Phi(\Delta) = 0$ otherwise. Equation (6) then gives:

$$\Delta d = \sqrt{6D\Delta t} \quad (7)$$

The control volume interpretation of this is:

$$C_j^{n+1} = \frac{\Delta x}{2\Delta d} \sum_i C_i^n \quad (8)$$

where the "departure" control volumes i are all located within a distance Δd from the "arrival" control volume j. Similar to before, equation (7) infers that (Néelz, 2000) :

$$\frac{\Delta d}{\Delta x} = \sqrt{6\alpha} \quad (9)$$

3. Numerical experiments

Numerical simulations were undertaken with the five schemes shown in Table (1). The departure points in Teixeira's method were interpolated using a cubic (4 point) interpolation scheme (Néelz, 2000).

Each experiment consisted of the pure diffusion of a Gaussian spatial concentration profile in a one-dimensional domain divided into control volumes of length Δx . Experiments were undertaken over ranges of dimensionless diffusion numbers, α , and dimensionless spatial resolutions, $\sigma/\Delta x$, where σ is the standard deviation of the Gaussian. α took values between 1 and 1000; $\sigma/\Delta x$ took values between 0.3 and 10. The aim of each experiment was to determine the number of time steps, N_c , required for the numerical solution to converge sufficiently close (< 5% difference over the central, ie +/- 2σ from centroid, part of the distribution) to the exact solution. Comparison of N_c for all schemes at the same values of α and $\sigma/\Delta x$ yielded information on the relative computational efficiencies of the schemes. Smaller values of N_c indicate greater efficiency through the use of larger time steps and with correspondingly shorter execution times.

Scheme	Description
1	θ -weighted semi-implicit, $\theta = 0.5$ (Crank-Nicholson)
2	θ -weighted semi-implicit, optimized θ
3	θ -weighted implicit, $\theta = 1.0$ (Backward implicit)
4	Teixeira with cubic interpolation
5	Modified Teixeira

TABLE 1. Numerical schemes used in experiments.

4. Results

Illustrative results are shown in Table (2) for low and high spatial resolution.

Two general trends were expected and found for all the schemes. Firstly, the use of smaller time steps (ie lower α) allows a faster convergence (ie smaller N_c). Secondly, the use of finer spatial resolution (ie higher $\sigma/\Delta x$) allows a faster convergence. Indeed, at the higher spatial resolutions for small α all schemes were very efficient, often converging after just a single time step. There are a few minor anomalies at low spatial resolution. For example, scheme 5 has a significant local maximum when $\alpha = 2$, schemes 2 & 3 have a small increase in N_c with decreasing α when $\alpha < 5$ and scheme 4 showed a rather odd behaviour for low values of α .

$\sigma/\Delta x$	0.3	0.3	0.3	0.3	0.3	10	10	10	10	10
Scheme	1	2	3	4	5	1	2	3	4	5
α										
1000	*	9	14	*	7	52	5	14	*	7
550	*	9	14	*	7	23	5	14	*	7
320	*	9	14	*	7	11	5	14	*	7
180	*	7	14	*	7	4	5	14	*	7
100	*	7	14	*	7	3	4	13	*	6
55	*	7	14	*	7	1	3	12	21	5
32	*	7	14	*	7	1	1	11	8	4
18	117	7	14	*	7	1	1	1	1	1
10	59	7	14	*	8	1	1	1	1	1
5.5	30	7	15	*	9	1	1	1	1	1
3.2	15	8	15	81	7	1	1	1	1	1
1.8	8	8	15	97	15	1	1	1	1	1
1	4	9	16	15	6	1	1	1	1	1

TABLE 2. Values of N_c for two spatial resolutions (* : more than 120 time steps needed).

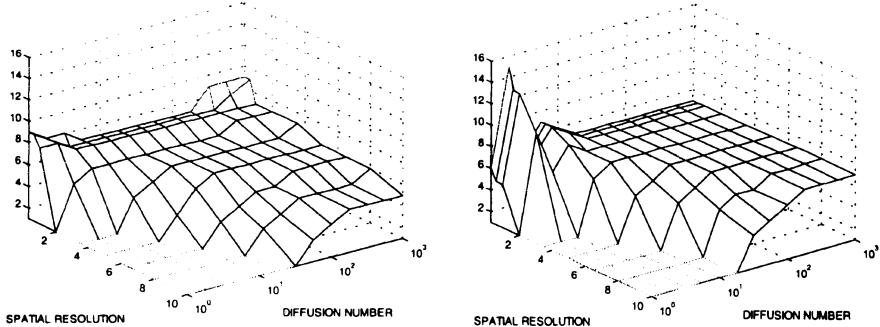


Figure 1. 3D-plots showing efficiency (N_c) as function of α and $\sigma/\Delta x$ for schemes 2 (left) and 5 (right).

In general, schemes 2 & 5 were of similar efficiency and were about twice as efficient as scheme 3. Scheme 4 was generally the least efficient and scheme 1 fell between schemes 3 and 4. Schemes 2,3 & 5 behaved in a similar (robust) fashion with each of them showing relatively little variation in N_c with α or $\sigma/\Delta x$ compared to schemes 1 & 4. Schemes 1 & 4 showed

the greatest variability in efficiency. Scheme 1 was very efficient at low α (due to its second-order accuracy) provided that grid-scale oscillations did not appear. The efficiency reduced markedly, however, in the presence of oscillations which were prevalent at large α . For scheme 2 the optimum value of θ was found to decrease slightly with increasing spatial resolution (0.74 to 0.68, for $\sigma/\Delta x = 0.3$ to 10). Note that the modified Teixeira method was significantly more robust and efficient than the original. A good impression of the efficiencies of the best two schemes for all α and $\sigma/\Delta x$ is shown in Figure (1).

5. Conclusions

At high spatial resolution, all five schemes showed similar efficiencies at low dimensionless diffusion numbers, but at high dimensionless diffusion numbers the modified Teixeira method, the implicit scheme and the optimised time-weighted schemes were superior to the other methods (with the modified Teixeira method and the optimised time-weighted schemes being best).

At low spatial resolution, the modified Teixeira method and the optimised time-weighted schemes were superior to the other methods except at very low dimensionless diffusion numbers for which the Crank-Nicholson scheme was the best.

From a more practical point of view, this study brings into light the constantly robust behaviour of the modified Teixeira method, the implicit scheme and the optimised time-weighted scheme (fast convergence with any time step). The modified Teixeira method is an interesting explicit alternative, but it should be used carefully at high diffusion numbers where it may be computationally expensive due to the use of a large number of grid points.

With an eye on future work, there seems to be no impediment, in principle, to extending any of the schemes tested here to two dimensions, which might be necessary for some applications. For example, two-dimensional finite difference schemes for diffusion are common (Abbott and Basco, 1989) and since the Teixeira type schemes are algorithmically similar to these (although using information at more grid points) their extension appears to be straightforward. For river modelling, however, a simplified governing equation using a natural coordinate system offers a more appropriate two-dimensional approach than a fully two-dimensional advection-diffusion equation. In this, differential longitudinal advection and transverse diffusion are the dominant physical processes, and the solution can be obtained using a streamtube approach (Manson and Wallis, 1998) without the need of a two-dimensional diffusion solver.

References

- Abbott M B and Basco D R (1979). Computational Fluid Dynamics : an Introduction for Engineers. Longman.
- Celia M A, Russell T F, Herrera I, and Ewing R E (1990). An Eulerian-Lagrangian Localized Adjoint Method for the Advection-Diffusion Transport Equation. *Advances in Water Resources* **13**, pp 187-206.
- Einstein A (1905). Über die von der Molekulärkinetischen Theorie der Wärme Geforderte Bewegung von in Ruhenden Flüssigkeiten Suspendierte Teilchen. *Annalen der Physik* **17**, pp 549-.
- Fischer H B, List E J, Kohm R C Y, Imberger J and Brooks N H (1979). Mixing in Inland and Coastal Waters. Academic Press, New York.
- Manson J R and Wallis S G (1998). Accurate Simulation of Transport Processes in Two-Dimensional Shear Flow. *Communications in Numerical Methods in Engineering* **14**, pp 863-869.
- Néelz S (2000). Numerical Prediction of Pollutant Fate in Streams. Licensiat Thesis, Department of Civil and Environmental Engineering, KTH, Stockholm, Sweden.
- Staniforth A and Côté J (1991). Semi-Lagrangian Integration Schemes for Atmospheric Models - a Review. *Monthly Weather Review* **119**, pp 2206-2223.
- Taylor G I (1954). The Dispersion of Matter in Turbulent Flow through a Pipe. *Proceedings of the Royal Society of London A* **223**, pp 446-468.
- Teixeira J (1998). Numerical Stability of the Physical Parametrizations in Atmospheric Models and a New Method to Solve the Diffusion Equation. *Proceedings of the 6th Conference on Numerical Methods for Fluids Dynamics, ICFD, Oxford*, pp 523-529.
- Van Dam G C (1994). Study of Shear Dispersion in Tidal Waters by Applying Discrete Particle Techniques. *Mixing and Transport in the Environment*, pp 269-. Reven K J, Chatwin P C and Millbank J H (Editors). John Wiley & Sons Ltd.
- Wallis S G and Manson J R (1997). Accurate Numerical Simulation of Advection Using Large Time Steps. *International Journal for Numerical Methods in Fluids* **24**, pp 127-139.
- Wallis S G, Manson J R and Filippi L (1998). A Conservative Semi-Lagrangian Algorithm for One-Dimensional Advection-Diffusion. *Communications in Numerical Methods in Engineering* **14**, pp 671-679.
- Wallis S G, Manson J R and Rafique S (1999). Limitations of Advection-Dispersion Calculations in Rivers. *Proceedings of 28th IAHR Congress, Graz, Austria*.

MULTIDIMENSIONAL FLUX-VECTOR-SPLITTING AND HIGH-RESOLUTION CHARACTERISTIC SCHEMES

SEBASTIAN NOELLE

*Institute for Applied Mathematics
Bonn University, Germany
Email: noelle@iam.uni-bonn.de*

1. Introduction

Since the work of Godunov, Van Leer, Harten-Lax and Roe, the numerical solution of systems of hyperbolic conservation laws is dominated by Riemann-solver based schemes. These one-dimensional schemes are usually extended to several space-dimensions either by using dimensional-splitting on cartesian grids or by the finite-volume approach on unstructured grids. The first systematic criticism of using one-dimensional Riemann-solvers for multi-dimensional gas-dynamics goes back to Roe himself: the Riemann-solver is applied in the grid- rather than the flow-direction, which may lead to a misinterpretation of the local wave-structure of the solution.

Among the genuinely multi-dimensional alternatives, which were developed since the mid-eighties, let us mention the fluctuation-splitting schemes of Roe, Deconinck, Van Leer et.al. (see (Deconinck et.al., 1993) for references), the Corner-Transport-Upwind (CTU) scheme (Colella, 1990), CLAWPACK (LeVeque, 1997), the Weighted-Average-Flux (WAF) scheme (Billet and Toro, 1997) and the Evolution-Galerkin-Method (Lukacova et.al., 1997).

In this contribution we focus on yet another multi-dimensional approach, Fey's Method of Transport (**MoT**) (Fey, 1998; Fey, 1998), which belongs to the family of flux-vector-splitting schemes. The starting point of Fey's algorithm is a multi-dimensional wave-model, which leads to a reformulation of the system of conservation laws as a finite set of coupled nonlinear advection equations. At the beginning of each timestep, these coupled nonlinear equations are decomposed into a set of linear scalar advection

equations with variable coefficients, which are then solved numerically using characteristic schemes.

One-dimensional first-order flux-vector-splitting schemes split the flux-vector in the center of a cell into components which are transported across the left and right boundaries of that cell. At sonic points, where eigenvalues change their sign, this procedure (which we would like to call **Cell-Centered-Evolution**) may lead to inconsistencies (Steger and Warming, 1981). Indeed, we have constructed a linear advection equation with smooth variable coefficients and smooth solutions for which Fey's first-order scalar characteristic scheme diverges at the sonic points.

Motivated by this discovery we have developed a new version of the MoT based on **Interface-Centered-Evolution**, the **MoT-ICE** (Noelle, 1999). The new method draws ideas from the flux-vector-splitting and the flux-difference-splitting approaches: the multi-dimensional wave-models are inherited from Fey's Method of Transport or other flux-vector-splitting schemes, while a predictor-step which provides auxiliary transport-velocities on the cell-interfaces uses flux-difference-splitting techniques.

For the new method, we have proved uniform first- resp. second-order consistency, including sonic points. Numerical experiments confirm second-order-accuracy for smooth solutions and high-resolution nonoscillatory shock-capturing properties for discontinuous solutions.

The second-order version of the new MoT-ICE is four to five times faster than Fey's second-order scheme and seems to be as fast as standard second-order algorithms. This gain of efficiency is partly due to an improved linearisation and decomposition of the nonlinear system into advection equations, and to our particularly simple characteristic transport algorithm for the resulting linear advection equations.

In ongoing work with C. von Törne, we are developing the MoT-ICE into a distributed-parallel, fully adaptive code, capable of handling general geometries.

2. Multi-dimensional Linearisation

We consider multidimensional systems of n conservation laws in d space dimensions,

$$\partial_t U + \nabla \cdot \mathbf{f}(\mathbf{U}) = \mathbf{0}$$

which can be written in Fey's advection form

$$\sum_{l=1}^L \left(\partial_t S_l(U) + \nabla \cdot (S_l(U) \mathbf{a}_l(\mathbf{U})^\mathbf{T}) \right) = 0, \quad (1)$$

where the $L \geq n$ waves $S_l : \mathbf{R}^n \rightarrow \mathbf{R}^n$ and $\mathbf{a}_l : \mathbf{R}^n \rightarrow \mathbf{R}^d$ satisfy the consistency conditions

$$\sum_{l=1}^L S_l(U) = U$$

and

$$\sum_{l=1}^L S_l(U) \mathbf{a}_l(\mathbf{U})^\mathbf{T} = \mathbf{f}(\mathbf{U}).$$

Examples of such systems include the wave-equation, the shallow-water-equations, the Euler-equations and the MHD-equations (Fey et.al., 1997).

Let us discuss how to update (1) from time t_n to $t_{n+1} = t_n + k$. Our first remark is that if one linearises this system by freezing the velocities at the half-timestep $t_{n+1/2} = t_n + k/2$ and solves the resulting coupled linear system, then the linearisation-error during a single timestep is $\mathcal{O}(k^3)$, so the overall linearisation is second-order-accurate in time.

The next step is to decompose the coupled linear system by setting each of the L summands to zero separately. This leads to a first-order decomposition-error. The leading-order term of this error can be removed by modifying the initial data for the decoupled linear equations. The modified initial data are of the form

$$Z_l(t_n) = S_l(U(t_n)) + \frac{k}{2} R_l(U(t_n), \nabla \cdot U(t_n))$$

with

$$\begin{aligned} R_l(U, \nabla \cdot U) &:= -S'_l(U) \nabla \cdot \mathbf{f}(\mathbf{U}) + \nabla \cdot (\mathbf{S}_l(\mathbf{U}) \mathbf{a}_l(\mathbf{U})^\mathbf{T}) \\ &= \partial_t S_l(U) + \nabla \cdot (S_l(U) \mathbf{a}_l(\mathbf{U})^\mathbf{T}). \end{aligned} \quad (2)$$

Note that the non-zero term R_l is the residual of the decomposition. In (Noelle, 1999) we prove that the resulting linearisation and decomposition is second-order-accurate in space and time.

For the Euler equations and the shallow-water equations, Morel, Fey and Maurer have also derived a second-order-accurate decomposition into linear advection equations. Instead of evaluating the velocities at the half-timestep they freeze them at the original timestep. As a consequence, they have to add correction terms for a mixed linearisation- and decomposition-error, which are computationally more expensive than (2).

3. Interface Centered Evolution

The linearisation and decomposition given in the previous section lead us, at the beginning of each timestep, to a set of transport equations of the

form

$$\partial_t \varphi + \nabla \cdot (\varphi \mathbf{a}^T) = 0 \quad (3)$$

with initial data φ given at time t_n .

Equation (3) may be solved by introducing the characteristics $\mathbf{z}(\tau; \mathbf{x}, \mathbf{t})$ via

$$\begin{aligned} \mathbf{z} : \mathbf{R}_+ \times \mathbf{R}^d \times \mathbf{R}_+ &\rightarrow \mathbf{R}^d \\ \mathbf{z}(\mathbf{t}; \mathbf{x}, \mathbf{t}) &= \mathbf{x} \\ \partial_\tau \mathbf{z}(\tau; \mathbf{x}, \mathbf{t}) &= \mathbf{a}(\mathbf{z}, \tau). \end{aligned}$$

Since the flux-vector $\varphi \mathbf{a}^T$ in (3) is always parallel to the characteristics, we have the following identity for the update: For all $K \subset \mathbf{R}^d$,

$$\int_K \varphi(\mathbf{x}, \mathbf{t}_{n+1}) d\mathbf{x} = \int_{\mathbf{z}(\mathbf{t}_n; K, \mathbf{t}_{n+1})} \varphi(\mathbf{x}, \mathbf{t}_n) d\mathbf{x}. \quad (4)$$

We now approximate the characteristic flow as follows: On each edge, or interface, we define an auxiliary transport-velocity in the direction normal to the interface. Using these auxiliary velocities, we move the interfaces backwards in time for a timestep k . For a two-dimensional cartesian grid, this results in a subdivision of each cell into up to nine subcells. These rectangular subcells approximate the subcells which would occur when tracing the interfaces backwards in time using the exact characteristic flow. The update is computed by replacing the exact characteristic flow in (4) by our approximate flow. All we need to do now is to define the auxiliary transport velocities. Up to higher-order terms, there is a unique choice which guarantees second-order consistency in 2D:

$$\hat{a}_{i+\frac{1}{2}, j} = \left[a^* + \frac{k}{2}(a^* a_x - b^* a_y) \right]_{i+\frac{1}{2}, j} + \mathcal{O}(k^2)$$

on the interfaces in x -direction and

$$\hat{b}_{i, j+\frac{1}{2}} = \left[b^* + \frac{k}{2}(b^* b_y - a^* b_x) \right]_{i, j+\frac{1}{2}} + \mathcal{O}(k^2)$$

on the interfaces in y -direction, where

$$\begin{aligned} a^* &= a + \frac{k}{2}a_t + \mathcal{O}(k^2) \\ b^* &= b + \frac{k}{2}b_t + \mathcal{O}(k^2) \end{aligned}$$

are predicted velocities at the half timestep on the interfaces. In (Noelle, 1999) we prove that this procedure is uniformly second-order-consistent for scalar equations (3) with smooth coefficients $\mathbf{a} = (\mathbf{a}, \mathbf{b})$.

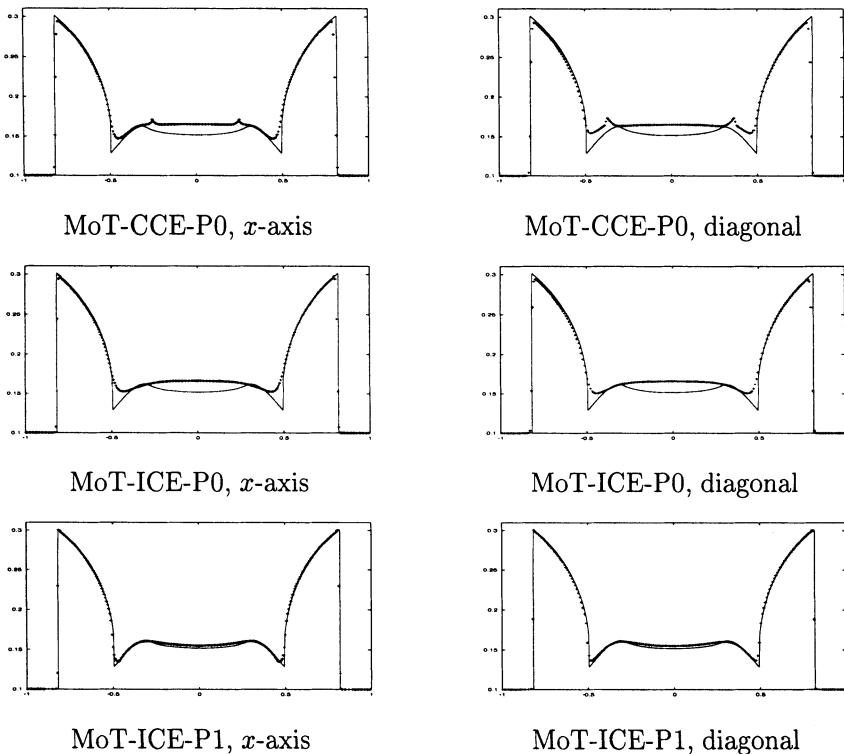


Figure 1. Cross-sections of 2D-computations of a radially-symmetric explosion for the shallow-water-equations. Plot of water-height for grids of 320×320 points. Fey's first-order MoT-CCE-P0 (top), our new first-order MoT-ICE-P0 (middle) and second-order MoT-ICE-P1 (bottom). Left column: cuts along the x -axis. Right column: cuts along the diagonal. Solid line: exact solution (one-dimensional computation with 3200 points). Note the kinks at the sonic points for the MoT-CCE-P0.

For nonlinear systems of conservation laws, where the velocity-field (a, b) is not known a-priori, it remains to define $a^*, b^*, a_x, a_y, b_x, b_y$ on the interfaces. We predict the values at the half-timestep by a one-dimensional flux-difference-splitting step, for example simply by a Lax-Friedrichs-step. Based on these predicted values a^*, b^* , we choose an upwind-weighted mean of the derivatives on both sides of an interface for a_x, a_y, b_x, b_y . These algorithmic details are crucial for stability in the presence of shocks. See (Noelle, 1999) for details.

4. Numerical Experiments

In Figure 1 we display 2D-computations of a radially-symmetric explosion for the shallow-water equations. One can see glitches at the sonic points for Fey's first-order scheme, which uses Cell-Centered-Evolution (MoT-CCE-

P0) common to other flux-vector-splitting schemes. These kinks do not occur for the first-order version of our new scheme (MoT-ICE-P0), and our second-order version (MoT-ICE-P1) resolves this problem very well.

Let us give a first comparison of efficiencies. Morel (Morel, 1997) reports that the MoT-CCE-P0, Van Leer's flux-vector-splitting and CLAWPACK $T^{1,0}$ (using the first-order Roe-solver without transverse wave-propagation) all use the same amount of cpu-time (say one time unit). CLAWPACK $T^{1,1}$ takes 1.5 units, and a first-order fix at sonic points proposed by Morel 2.9 units. Our preliminary experience with the MoT-ICE is the following: the MoT-ICE-P0 takes 0.9 to 1.0 units and, hence, it is as fast as standard first-order schemes. The MoT-ICE-P1 takes 2.0 - 2.2 units, which is the same as the second-order CLAWPACK $T^{2,2}$. This compares favorably with Fey's MoT-CCE-P1 (Fey, 1998), which is consistent at sonic points, but needs 10.5 units of cpu-time.

Acknowledgement: This work was supported by DFG-SPP ANumE and SFB256 at Bonn University.

References

- Billet S and Toro E (1997). On WAF-type schemes for multidimensional hyperbolic conservation laws. *J. Comput. Phys.* **130**, pp 1-24.
- Colella P (1990). Multidimensional upwind methods for hyperbolic conservation laws. *J. Comput. Phys.*, **87**, pp 171-200.
- Deconinck H, Paillère H Struijs R and Roe P (1993). Multidimensional upwind schemes based on fluctuation-splitting for systems of conservation laws. *Comput. Mech.*, **11**, pp 323-340.
- Fey M (1998). Multidimensional upwinding. I. The method of transport for solving the Euler equations. *J. Comput. Phys.* **143**, pp 159-180.
- Fey M (1998). Multidimensional upwinding. II. Decomposition of the Euler equations into advection equations. *J. Comput. Phys.* **143**, pp 181-199.
- Fey M, Jeltsch R, Maurer J and Morel AT (1997). The method of transport for nonlinear systems of hyperbolic conservation laws in several space dimensions. *Research Report No. 97-12, Seminar for Applied Mathematics, ETH Zürich*.
- LeVeque RJ (1997). Wave Propagation Algorithms for Multidimensional Hyperbolic Systems. *J. Comput. Phys.* **131**, pp 327-353.
- Lukacova-Medvidova M, Morton K and Warnecke G (1997). Evolution Galerkin methods for hyperbolic systems in two space dimensions. *Report 97-44, Univ. Magdeburg, Germany To appear in Math. Comp., 2000*.
- Morel AT (1997). A genuinely multidimensional high-resolution scheme for the shallow-water equations. *Dissertation, ETH Zürich Diss. No. 11959*.
- Noelle S (1999). The MoT-ICE: a new high-resolution wave-propagation algorithm based on Fey's Method of Transport. Invited plenary lecture. *"Proceedings of the Second International Symposium on Finite Volumes for Complex Applications - Problems and Perspectives", Duisburg, Germany*, p. 95.
For details see Preprint no.1999-028 at
<http://www.math.ntnu.no/conservation/1999/028.html>.
- Steger J and Warming R (1981). Flux vector splitting of the inviscid gas-dynamic equations with applications to finite difference methods. *J. Comput. Phys.* **40**, pp 263-293.

A COMPARISON OF ROE, VFFC AND AUSM+ SCHEMES FOR TWO-PHASE WATER/STEAM FLOWS

H. PAILLERE, A. KUMBARO, C. VIOZAT AND S. CLERC

CEA Saclay, Département de Mécanique et de Technologie,

91191 Gif-sur-Yvette Cedex, France

Email: henri.pailiere@cea.fr

AND

A. BROQUET AND C. CORRE

ENSAM/SINUMEF, 151 Boulevard de l'Hôpital,

75013 Paris, France

Email: corre@paris.ensam.fr

Abstract.

We discuss the extension of state-of-the-art Godunov schemes such as Roe's Approximate Riemann solver or Liou's Advection Upstream Splitting Method (AUSM+) to two-phase flow. The motivation for extending these methods to such flows is to benefit from well-known properties of characteristic-based upwind solvers, namely low numerical dissipation, sharp capture of shock and contact discontinuities, conservation through a finite volume formulation, and easy extension to unstructured meshes by directional splitting.

1. Introduction

The most simple two-phase flow model is the so-called Homogeneous Equilibrium Model (HEM), in which the two phases are assumed to be in kinematic and thermodynamical equilibrium. In that case, the model governing the flow of the mixture is analogous to the Euler equations of gas dynamics, with a general equation of state (EOS) of the form $\rho = \rho(p, h)$. In one space dimension, the equations read:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} = 0 \quad (1)$$

with $\mathbf{U} = (\rho, \rho u, \rho E)^\tau$ and $\mathbf{F} = (\rho u, \rho u^2 + p, \rho u H)^\tau$, $E = H - p/\rho$ and $H = h + \frac{u^2}{2}$. The two-phase flow character of the model is solely contained in the equation of state, given in tabulated form. Here, we will consider thermodynamic tables representative of liquid water/steam flow over a large range of pressure and temperature. These tables give in particular the specific enthalpy of each phase at saturation as a function of pressure, $h_l^{sat}(p)$ and $h_v^{sat}(p)$, where l and v denote respectively the liquid and vapour phases. From these and the value of the specific enthalpy of the mixture, one can determine the quality of the mixture,

$$x = \frac{h - h_l^{sat}}{h_v^{sat} - h_l^{sat}} \quad (2)$$

If $x \leq 0$, then the mixture consists only of the liquid phase and $\rho(p, h) = \rho_l(p, h)$; if $x \geq 1$, then only the vapour phase is present, and $\rho(p, h) = \rho_v(p, h)$. In the intermediate two-phase domain, the two phases are at saturation and the density is given as:

$$\rho(p, h) = \alpha \rho_v^{sat} + (1 - \alpha) \rho_l^{sat} \quad (3)$$

where $\rho_k^{sat} = \rho_k(p, h_k^{sat}(p))$ for each phase, and α is the volume fraction of vapour phase, known as the void fraction and equal to $x\rho/\rho_v$.

The tables also provide the partial derivatives of density needed to compute the speed of sound of the two-phase mixture:

$$a = \left[\left(\frac{\partial \rho}{\partial p} \right)_h + \frac{1}{\rho} \left(\frac{\partial \rho}{\partial h} \right)_p \right]^{-\frac{1}{2}} \quad (4)$$

The partial derivatives of pressure are discontinuous across the saturation line which divides the single and two-phase flow domains. This ‘kink’ represents a difficulty when linearising the equations near saturation states and is also responsible for the large variations of sound speed, typically a thousand meters per second in the liquid phase, and a few meters per second in the mixture.

2. Upwind differencing schemes for equilibrium two-phase flow

As is well-known, the extension of Roe’s approximate Riemann solver (Roe, 1981) to arbitrary equations of state is not trivial since it requires particular care in the construction of the linearisation and the Roe matrix (Toumi, 1992; Clerc, 2000). On the other hand, the AUSM family of schemes, in which pressure is separated from the convective flux allows for easier generalisation (Liou, 1996; Liou and Edwards, 1999).

2.1. THE ROE SCHEME

The Roe scheme is probably one of the better known upwind differencing scheme for the Euler equations for perfect gases, and may be written as:

$$\mathbf{F}_{\frac{1}{2}} = \frac{1}{2} [\mathbf{F}_L + \mathbf{F}_R] - \frac{1}{2} |\mathbf{A}(\hat{\mathbf{U}})| (\mathbf{U}_R - \mathbf{U}_L) \quad (5)$$

where \mathbf{A} is the Jacobian matrix $\partial \mathbf{F} / \partial \mathbf{U}$, evaluated at the Roe average state $\hat{\mathbf{U}}$ (Roe, 1981). One of the main difficulties in extending the scheme to the two-phase homogeneous equilibrium flow model lies in the construction of the so-called Roe matrix and associated linearisation. Problems can occur when \mathbf{U}_L and \mathbf{U}_R are not in the same region (pure liquid, two-phase, pure vapour) because the left and right states are not described by the same equation of state, and thus averaging between the two states may lead to an unphysical state. In (Toumi, 1992), Toumi introduced a sophisticated averaging procedure based on an optimal choice of path $\Phi(s, \mathbf{U}_L, \mathbf{U}_R)$ in order to construct $\hat{\mathbf{U}}_\Phi$. Here, we use a simpler averaging of the form:

$$\begin{aligned} \hat{\rho} &= \sqrt{\rho_L \rho_R} & \hat{u} &= \omega u_L + (1 - \omega) u_R \\ \hat{H} &= \omega H_L + (1 - \omega) H_R & \hat{E} &= \omega E_L + (1 - \omega) E_R \end{aligned}$$

with $\omega = \sqrt{\rho_L} / (\sqrt{\rho_L} + \sqrt{\rho_R})$. The average pressure is then computed as $\hat{p} = \hat{\rho}[\hat{H} - \hat{E}]$, which, together with $\hat{h} = \hat{H} - \frac{1}{2}\hat{u}^2$, allows us to compute the speed of sound of the two-phase mixture using the equation of state. For the test cases considered here, this linearisation has proved satisfactory, though more sophisticated averaging may be required for more difficult cases.

2.2. THE VFFC SCHEME

The VFFC scheme ('Volumes Finis à Flux Caractéristiques' *en français dans le texte*) is another approximate Riemann solver (Huang, 1981), which has been extended in recent years to two-phase flow models, and in particular to non-conservative, non-hyperbolic two-fluid models (Ghidaglia et al., 1995). Recently, Ghidaglia reformulated the scheme as one in a family of upwind schemes called 'Flux Schemes' (Ghidaglia, 1998), of the form:

$$\mathbf{F}_{\frac{1}{2}} = \frac{1}{2} [\mathbf{F}_L + \mathbf{F}_R] - \frac{1}{2} \mathbf{Q} (\mathbf{F}_R - \mathbf{F}_L) \quad (6)$$

\mathbf{Q} may be regarded as a dissipation matrix, which in the case of the VFFC scheme is given by:

$$\mathbf{Q} = \text{sign}(\mathbf{A}(\bar{\mathbf{U}})) \quad (7)$$

where $\bar{\mathbf{U}}$ is the arithmetic average between the left and right states \mathbf{U}_L and \mathbf{U}_R (on a uniform mesh). Taking precisely as average the Roe average

state $\hat{\mathbf{U}}$, one also recovers the Roe scheme, showing that the Roe scheme is a Flux Scheme.

2.3. THE AUSM+ SCHEME

The AUSM+ scheme (Liou, 1996) was proposed by Liou as an inexpensive flux splitting scheme able to capture exactly steady contact discontinuities, a property usually associated with approximate Riemann solvers such as the Roe or VFFC schemes. However, unlike the latter, the flux splitting does not require any characteristic analysis or field by field decomposition. Rather, it is simply based on a clever splitting of the flux into a convective part associated to the mass flux $\dot{m} = \rho u = \rho a M$, and a pressure part:

$$\mathbf{F} = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ \rho u H \end{pmatrix} = \mathbf{F}^c + \mathbf{P} = \dot{m} \begin{pmatrix} 1 \\ u \\ H \end{pmatrix} + \begin{pmatrix} 0 \\ p \\ 0 \end{pmatrix} = \dot{m} \Psi + \mathbf{P} \quad (8)$$

The numerical flux function may thus be written as:

$$\mathbf{F}_{\frac{1}{2}}(\mathbf{U}_L, \mathbf{U}_R) = \mathbf{F}_{\frac{1}{2}}^c + \mathbf{P}_{\frac{1}{2}} = \dot{m}_{\frac{1}{2}} \Psi_{\frac{1}{2}} + \begin{pmatrix} 0 \\ p_{\frac{1}{2}} \\ 0 \end{pmatrix} \quad (9)$$

where simple upwinding based on the sign of $\dot{m}_{\frac{1}{2}}$ is used to compute $\Psi_{\frac{1}{2}}$, i.e. $\Psi_{\frac{1}{2}} = \Psi(\mathbf{U}_L)$ if $\dot{m}_{\frac{1}{2}} \geq 0$, and $\Psi_{\frac{1}{2}} = \Psi(\mathbf{U}_R)$ if $\dot{m}_{\frac{1}{2}} < 0$. There remains of course to define the two scalars $p_{\frac{1}{2}}$ and $\dot{m}_{\frac{1}{2}}$.

In a first step, Liou defines a numerical speed of sound at the interface, $a_{\frac{1}{2}} = \sqrt{a_L a_R}$, and a left and right Mach number based on the normal flow velocity in the left and right cells and the numerical speed of sound, $M_L = u_L/a_{\frac{1}{2}}$ and $M_R = u_R/a_{\frac{1}{2}}$.

In a second step, the interface pressure is computed as a weighted average of pressure in the left and right cells,

$$p_{\frac{1}{2}} = \mathcal{P}^+(M_L)p_L + \mathcal{P}^-(M_R)p_R \quad (10)$$

where \mathcal{P}^\pm are polynomial functions, satisfying consistency, differentiability and symmetry conditions. Note however that $p_{\frac{1}{2}}$ does not necessarily lie between p_L and p_R , since $\mathcal{P}^+(x) + \mathcal{P}^-(y) \neq 1$ in general.

In the third and final step, an interface numerical Mach number $M_{\frac{1}{2}}$ is computed as a polynomial function of the left and right Mach numbers,

$$M_{\frac{1}{2}} = \mathcal{M}^+(M_L) + \mathcal{M}^-(M_R), \quad (11)$$

where \mathcal{M}^\pm are polynomial functions. Then, the mass flux at the interface is simply defined as:

$$\dot{m}_{\frac{1}{2}} = a_{\frac{1}{2}} \left(\rho_L \frac{M_{\frac{1}{2}} + |M_{\frac{1}{2}}|}{2} + \rho_R \frac{|M_{\frac{1}{2}} - M_{\frac{1}{2}}|}{2} \right) \quad (12)$$

Choosing a data structure based on primitive variables (ρ, u, a, p, H) , the AUSM+ scheme can be coded independently of the flow model, perfect gas or two-phase HEM. Indeed, the algorithm simply reads:

1. do $i=1, N_{\text{cells}}$: compute and store variables $(\rho, u, a, p, H)_i$
2. do $f=1, N_{\text{faces}}$: compute $\mathbf{F}_{\frac{1}{2}}^{\text{AUSM}+}$
 - compute $a_{\frac{1}{2}} = \sqrt{a_L a_R}$
 - compute M_L and M_R
 - compute $p_{\frac{1}{2}}, M_{\frac{1}{2}}$ and $\dot{m}_{\frac{1}{2}}$
 - return flux $\mathbf{F}_{\frac{1}{2}}^{\text{AUSM}+} = \mathbf{F}_{\frac{1}{2}}^c + \mathbf{P}_{\frac{1}{2}}$

The first step is common to all schemes and requires a Newton-type algorithm with calls to the EOS. For the Roe and VFFC schemes where local linearisations and eigen decompositions are performed at each interface, an additional call to the EOS is also required in the flux evaluation.

3. Numerical results

3.1. WATER FAUCET PROBLEM

This test case is a well-known benchmark problem for two-phase flow solvers, initially designed for two-fluid models in which the liquid and vapour phases are governed by separate mass and momentum equations. The initial conditions for the two-fluid model are $p = 1$ bar, $\alpha = 0.2$, $u_v = 0$ and $u_l = 1$ m/s. A homogeneous two-phase flow problem can thus be defined based on the following conditions:

$$\begin{aligned} p &= 1 \text{ bar}, \quad \alpha = 0.2 \\ \rho &= \alpha \rho_v^{sat}(p) + (1 - \alpha) \rho_l^{sat}(p) \\ u &= [\alpha \rho_v u_v + (1 - \alpha) \rho_l u_l] / \rho \\ h &= [\alpha \rho_v h_v^{sat} + (1 - \alpha) \rho_l h_l^{sat}] / \rho \end{aligned}$$

Figure 1 shows the profiles of void fraction obtained with the 3 different numerical methods, as well as the time-evolution of the void fraction profiles for the AUSM+ scheme. Results are virtually indistinguishable from one

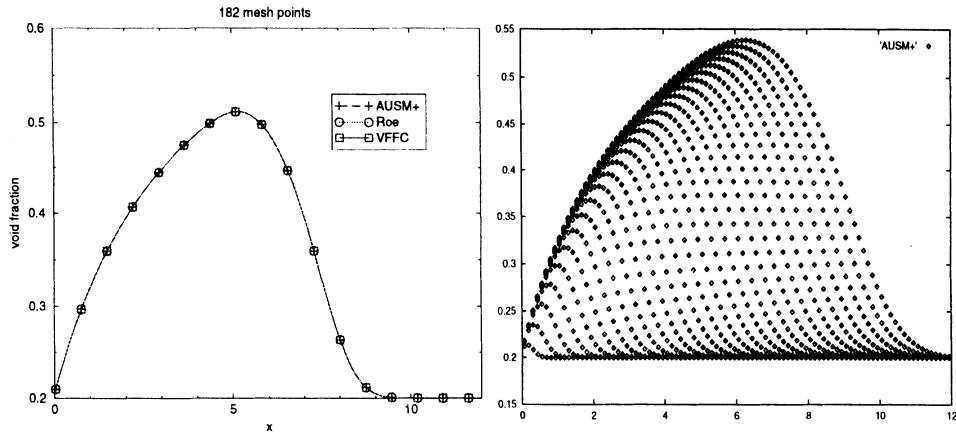


Figure 1. Water faucet problem. The left figure represents the void fraction profile at $t = 0.5$ s computed with the Roe, VFFC and AUSM+ schemes; the right figure represents the time-evolution of the void fraction profile computed with the AUSM+ scheme

another, showing that for this test case, the AUSM+ scheme performs as well as the Roe or VFFC schemes.

3.2. TWO-PHASE SHOCK TUBE PROBLEM

A two-phase shock tube problem is considered with initial conditions:

$$\begin{aligned}\alpha_L &= 0.25, p_L = 200 \text{ bars}, u_L = 0, h_L = 1886847 \text{ J/kg} \\ \alpha_R &= 0.70, p_R = 150 \text{ bars}, u_R = 0, h_R = 1886847 \text{ J/kg}\end{aligned}$$

Figure 2 shows the void fraction profiles at $t = 10^{-3}$ s, computed with

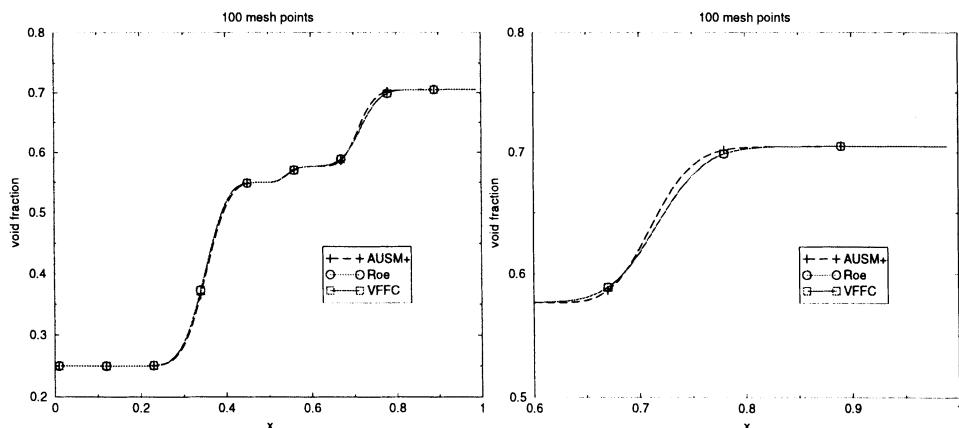


Figure 2. Two-phase shock tube problem

the AUSM+, Roe and VFFC schemes. The results are again very close to one another, though the close up shows a slightly sharper resolution of the AUSM+ scheme compared to the Roe or VFFC schemes.

4. Conclusions and future work

In this paper, we have briefly discussed the extension of three upwind schemes to a hyperbolic system of conservation laws corresponding to the two-phase homogeneous equilibrium flow model. As noted by Liou, the extension of the AUSM+ scheme to arbitrary equations of state requires little effort. One of the reasons for this simplicity lies in the fact that the scheme makes use of linearised variables (sound speed, Mach number, pressure) constructed from the corresponding values in the left and right states, and does not require any other calls to the equation of state.

In the case of the Roe or VFFC schemes, simple linearisations may also be constructed so as to separate the construction of the numerical dissipation from the thermo-dynamics. Still, compared to the AUSM+ scheme, an additional call to the equation of state and eigen decompositions are needed to construct the upwind fluxes, resulting in a higher CPU cost.

The test cases have shown that all three schemes provide the same level of accuracy. It remains to be seen whether the AUSM+ scheme is as robust as the Roe scheme or the VFFC scheme when applied to more difficult test cases, involving fast depressurisations for example. The extension of the AUSM+ scheme to non-conservative equal-pressure two-fluid models also remains to be developed before more comparisons with the Roe and VFFC schemes can be performed. This work is currently in progress.

References

- Clerc, S. (2000) Numerical Simulation of Equilibrium Two-Phase Flows. *J. Comput. Phys.* to appear
- Ghidaglia, J.-M. (1998) Flux Schemes for Solving Non-linear Systems of Conservation Laws. in Proc. Meeting in Honour of P.L. Roe, Arcachon, France
- Ghidaglia, J.M., Kumbaro, A., Le Coq, G. and Tajchman, M. (1995) A Finite Volume Implicit Method Based on Characteristic Flux For Solving Hyperbolic Systems of Conservation Laws. in Proc. Conference on Nonlinear Evolution Equations and Infinite-Dimensional Dynamical Systems, Shanghai
- Huang, L.C. (1981) Pseudo-unsteady difference schemes for discontinuous solutions of steady-state one-dimensional fluid dynamics problem. *J. Comput. Phys.* **42**:195
- Liou, M.S. (1996) A Sequel to AUSM: AUSM⁺. *J. Comput. Phys.*, **129**:364
- Liou, M.S. and Edwards, J.R. (1999) AUSM Schemes and Extensions for Low Mach and Multiphase Flow. *VKI LS 1999-03, Computational Fluid Dynamics*
- Roe, P.L. (1981) Approximate Riemann Solvers, parameter vectors & difference schemes. *J. Comput. Phys.*, **43**
- Toumi, I. (1992) A Weak Formulation of Roe's Approximate Riemann Solver. *J. Comput. Phys.* **102**

LOW DISSIPATION ENTROPY FIX FOR POSITIVITY PRESERVING ROE'S SCHEME

M. PELANTI, L. QUARTAPELLE AND L. VIGEVANO

Dipartimento di Ingegneria Aerospaziale

Politecnico di Milano

Via La Masa 34, 20159 Milano, Italy

Email: vigevano@aero.polimi.it

Abstract. We present a general formulation for the entropy fix which encompasses both Harten and Hyman entropy fixes and HLLE/M schemes. Such a formulation is found to be a simple tool for benefitting from the properties of different methods so as to obtain a positivity preserving version of entropy fix with low dissipation.

1. Introduction

Roe's linearization (Roe, 1981) represents a widespread method for solving Euler equations. However, as well known, this method may suffer from possible entropy violations and the prediction of unphysical intermediate states in presence of strong rarefactions.

The generation of nonentropic solutions can be prevented by means of suitable entropy fixes. This can be accomplished by writing the upwind scheme so as to put into evidence its numerical dissipation matrix and by operating on this matrix to assure a non zero "viscous" contribution to the numerical flux, as proposed for instance by Harten (Harten, 1983). It results a very simple and computationally convenient correction to the original Roe's scheme.

To avoid the violation of positivity of density or internal energy is a much more difficult task. The classical Roe's linearization in Jacobian form has not enough degrees of freedom to impose positivity together with consistency with the conservation laws, as demonstrated in (Einfeldt *et al.*, 1991), at least for a certain class of symmetric Riemann problems. The difficulty is avoided by resorting to a positivity preserving approximation, as the HLLE scheme (Einfeldt, 1988), (Einfeldt *et al.*, 1991). Unfortunately the HLLE method is characterized by a numerical dissipation larger than

in Roe's scheme, particularly near contact discontinuities. To overcome this drawback in recent years some modifications of the original HLLE scheme have been proposed (Einfeldt *et al.*, 1991), (Wada, 1993). An alternative approach is suggested by Dubroca (1998), that allows for the required degrees of freedom by departing from the classical Jacobian-based Roe's linearization.

The present work follows a different line. The starting point is the observation that the lack of positivity of Roe's scheme has been explained (Einfeldt *et al.*, 1991) as the consequence of an underestimation by Roe's approximate solver of the absolute value of the minimum and maximum physical signal velocities. By means of the approach we propose here, the entropy fix can be regarded as a tool to properly correct the numerical signal velocities computed by Roe's method in order to preserve the positivity of the solution. Moreover, since the entropy fix allows introducing additional parameters in the numerical scheme, it guarantees a sufficient number of degrees of freedom to impose the positivity condition, while still retaining the classical Jacobian form in the Roe's linearization.

The outline of this paper is as follows. In section 2 a general formulation of the entropy fix is presented, which encompasses both some of the existing entropy fixes (Harten and Hyman, 1983), (Harten, 1983), (LeVeque, 1992) and the HLLE/M schemes. In section 3 we use the established general framework to propose a positivity preserving version of entropy fix with low dissipation, by exploiting the properties of different methods. Some numerical examples are shown in section 4.

2. Unitary framework

2.1. A GENERAL EXPRESSION FOR THE ENTROPY FIX

In the Introduction, the approach to the entropy fix based on the notion of a “viscosity” matrix has been recalled. A rather different viewpoint is that adopted by LeVeque in (LeVeque, 1992), where the entropy fix is presented as a remedy for the difficulties encountered by Roe's approximate solver in case of transonic rarefactions.

LeVeque's approach provides us the guidelines for obtaining a general formulation of the entropy fix in terms of propagating velocities. According to the the usual notation, $\hat{a}_k = \hat{a}_k(\mathbf{u}_\ell, \mathbf{u}_r)$ and $\hat{\mathbf{r}}_k = \hat{\mathbf{r}}_k(\mathbf{u}_\ell, \mathbf{u}_r)$ will denote here respectively the eigenvalues and the eigenvectors of the Roe's matrix $\hat{\mathbf{A}} = \hat{\mathbf{A}}(\mathbf{u}_\ell, \mathbf{u}_r)$. The proposed general formulation is expressed by the numerical flux:

$$\mathbf{F}(\mathbf{u}_\ell, \mathbf{u}_r) = \frac{1}{2}[\mathbf{f}(\mathbf{u}_\ell) + \mathbf{f}(\mathbf{u}_r)] - \frac{1}{2} \sum_{k=1}^3 \alpha_k q(\hat{a}_k) \hat{\mathbf{r}}_k , \quad (1)$$

where

$$q(\hat{a}_k) = \begin{cases} \frac{[\mathcal{P}(a_{kr}) + \mathcal{N}(a_{k\ell})]\hat{a}_k - 2\mathcal{P}(a_{kr})\mathcal{N}(a_{k\ell})}{\mathcal{P}(a_{kr}) - \mathcal{N}(a_{k\ell})} + \frac{\mathcal{P}(a_{kr})\mathcal{N}(a_{k\ell})}{\mathcal{P}(a_{kr}) - \mathcal{N}(a_{k\ell})}\sigma_k & \text{if } \mathcal{P}(a_{kr}) \neq \mathcal{N}(a_{k\ell}) \\ |\hat{a}_k| & \text{if } \mathcal{P}(a_{kr}) = \mathcal{N}(a_{k\ell}), \end{cases} \quad (2)$$

with the propagation velocities $a_{k\ell}$ and a_{kr} and the parameter σ_k to be suitably defined. We use $\mathcal{P}(\cdot)$ and $\mathcal{N}(\cdot)$ to denote the positive and negative parts of their respective argument.

Following Harten and Hyman (Harten and Hyman, 1983), it is possible to demonstrate that in the scalar case, and for $\sigma = 0$, the above formulation guarantees consistency with the entropy condition if the propagation velocities a_ℓ and a_r satisfy the inequalities:

$$a_\ell \leq a(u_\ell) \quad \text{and} \quad a_r \geq a(u_r). \quad (3)$$

The problem of imposing more general conditions on the parameters $a_{k\ell}$, a_{kr} and σ_k to have entropic solutions is still to be investigated.

2.2. RECOVERING HARTEN AND HYMAN ENTROPY FIXES

Formulation (2) allows recovering Harten and Hyman entropy fixes and the version of entropy fix proposed by LeVeque by defining properly the velocities $a_{k\ell}$ and a_{kr} and the parameter σ_k . If we choose

$$a_{k\ell} = \hat{a}_k - \delta_k \quad \text{and} \quad a_{kr} = \hat{a}_k + \delta_k, \quad (4)$$

with $\delta_k = \max\{0, \hat{a}_k - a_k(\mathbf{u}_\ell), a_k(\mathbf{u}_r) - \hat{a}_k\}$, and we set $\sigma_k = 0$, $\forall k$, we obtain the method derived by Harten and Hyman (1983, p. 243). With the same definition for $a_{k\ell}$ and a_{kr} , but now setting $\sigma_k = 1$, we find the alternative method also proposed by Harten and Hyman (1983, p. 266). LeVeque's version of entropy fix (LeVeque, 1992) is obtained by restricting correction (2) of Roe's scheme to the sonic eigenvalues $k = 1$ and $k = 3$, setting $\sigma_k = 0$, and by defining

$$a_{k\ell} = a_k(\hat{\mathbf{u}}_{k,\ell}) \quad \text{and} \quad a_{kr} = a_k(\hat{\mathbf{u}}_{k,r}), \quad (5)$$

where, as usual, $\hat{\mathbf{u}}_{1,\ell} = \mathbf{u}_\ell$, $\hat{\mathbf{u}}_{1,r} = \hat{\mathbf{u}}_1 = \mathbf{u}_\ell + \alpha_1 \hat{\mathbf{r}}_1 = \hat{\mathbf{u}}_{2,\ell}$, $\hat{\mathbf{u}}_{2,r} = \hat{\mathbf{u}}_2 = \mathbf{u}_\ell + \alpha_1 \hat{\mathbf{r}}_1 + \alpha_2 \hat{\mathbf{r}}_2 = \hat{\mathbf{u}}_{3,\ell}$, and $\hat{\mathbf{u}}_{3,r} = \mathbf{u}_r$.

2.3. RECOVERING HLLE AND HLLEM SCHEMES

We find that HLLE and HLLEM schemes fall in the general formulation (2) of entropy fix just presented. We will use in the following the quantities

b_ℓ and b_r introduced by Einfeldt *et al.* (1991):

$$b_\ell = \min\{\hat{a}_1, v_\ell - c_\ell\} \quad \text{and} \quad b_r = \max\{\hat{a}_3, v_r + c_r\}, \quad (6)$$

where v and c denotes respectively the fluid velocity and the sound speed. For both the HLLE and HLLEM methods we have

$$a_{k\ell} = b_\ell \quad \text{and} \quad a_{kr} = b_r, \quad \forall k, \quad (7)$$

and $\sigma_1 = \sigma_3 = 0$. We obtain the HLLE scheme (Einfeldt *et al.*, 1991) imposing $\sigma_2 = 0$, while we recover the HLLEM scheme (Einfeldt *et al.*, 1991) if we set $\sigma_2 = 2\hat{\delta}$, with $\hat{\delta} = \frac{\hat{c}}{\hat{c} + |\bar{v}|}$, being $\hat{c} = c(\hat{h}^t, \hat{v})$ and $\bar{v} = \frac{b_\ell + b_r}{2}$.

Here \hat{h}^t and \hat{v} are the well known Roe-average of the total enthalpy per unit mass and of velocity.

Placing the HLLE scheme in the same setting of the classical entropy fix formulations supports the introduction of the idea of *positivity preserving* entropy fix and also suggests how to correct Roe's scheme to impose the positivity.

3. The proposed method

The general formulation (2) is found to be a useful and simple tool to benefit from the properties of the different methods considered here in order to guarantee: i) consistency with the entropy condition, ii) positivity, iii) low numerical dissipation. Indeed, we can suitably define the quantities $a_{k\ell}$, a_{kr} and σ_k depending on the local solution to assure the aforementioned properties. For fixed values of $a_{k\ell}$ and a_{kr} , an increase in the slope σ_k implies a lower numerical dissipation. Nevertheless, this cannot be the only criterion to adopt to set properly the value of σ_k , and in particular we still need to investigate the relation of this parameter with the entropy condition, as anticipated in section 2. Therefore, in the following we restrict our analysis to the case $\sigma_k = 0, \forall k$, for semplicity.

We start distinguishing the case in which Roe's intermediate states $\hat{\mathbf{u}}_1 = \mathbf{u}_\ell + \alpha_1 \hat{\mathbf{r}}_1$ and $\hat{\mathbf{u}}_2 = \mathbf{u}_\ell + \alpha_1 \hat{\mathbf{r}}_1 + \alpha_2 \hat{\mathbf{r}}_2$ are physically admissible from the case in which one of them or both are not. The condition discriminating the two cases consists in checking the positivity of the density and internal energy of states $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$. If the two computed intermediate states are physical, the positivity of the solution is naturally preserved, and we correct Roe's numerical flux by means of (2) only to avoid entropy violations, as usual. In such a case, for $a_{k\ell}$ and a_{kr} we use the definitions (5), as in LeVeque's method. According to the physical interpretation of the entropy fix suggested by LeVeque, this choice of the propagating velocities allows a better approximation of the exact solution of the Riemann problem and is

found to introduce the lowest level of numerical viscosity, with respect to the other possible definitions of $a_{k\ell}$ and a_{kr} , for $\sigma_k = 0$.

If on the contrary negative values of density or internal energy or both are detected, definition (5) for the propagation velocities cannot be used, since they depend at least on one not physically admissible state. Moreover, in this case we need to define $a_{k\ell}$ and a_{kr} so as to force a suitable enlargement of the numerical signal velocities, thus avoiding the underestimation of the limiting physical velocities caused by Roe's approximate solver. Following the HLLE idea, we use definition (7) for the propagation velocities. This choice guarantees both consistency with entropy condition and positivity, as demonstrated in (Einfeldt *et al.*, 1991). We remark that, if we use a nonzero value for the parameter σ_2 , still having $\sigma_1 = \sigma_3 = 0$ and the same definition (7) of $a_{k\ell}$ and a_{kr} , it is in principle possible to find out sufficient conditions on σ_2 guaranteeing positivity. These conditions are presently under investigation.

The proposed version of entropy fix proves to be a positivity preserving correction of Roe's scheme that allows an easy implementation and requires an additional computation of no relevant cost with respect to Roe's method augmented by LeVeque's entropy fix (LeVeque, 1992).

4. Numerical results

Figure 1 compares the first order numerical results obtained with the presented method and the HLLE/M methods for a Riemann problem proposed in (Einfeldt *et al.*, 1991), consisting in two symmetric rarefactions. The initial data are $\rho_\ell = 1$, $v_\ell = -2$, $P_\ell = 0.4$ for the left state, $\rho_r = 1$, $v_r = 2$, $P_r = 0.4$ for the right state.

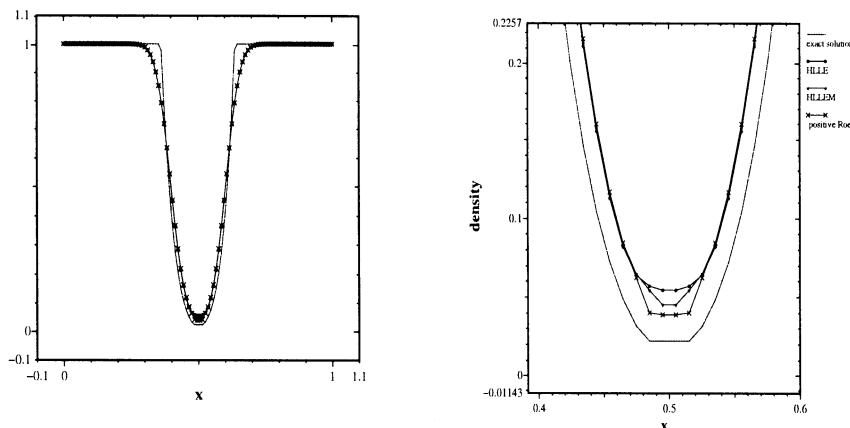


Figure 1. Strong rarefaction test problem. Comparison of the solutions computed by Roe method augmented with the proposed entropy fix and HLLE/M methods.

The proposed entropy fix allows resolving without difficulties this strong rarefaction test, which causes the failure of Roe's classical scheme (Roe, 1981), and is found to be slightly less dissipative than the HLLE scheme.

Figure 2 shows the solutions computed using the same methods for a Riemann problem obtained from the former, by replacing the value of the pressure of the left state with $P_\ell = 2$. In such a case the problem is non-symmetric. The present method and the HLLEM method, being less dissipative than the HLLE scheme, feature a small undershoot with respect to the exact solution, which does not prevent, however, to compute a positive solution.

In solving Riemann problems different from those implying low density regions, the presented method preserves all the properties of Roe's scheme.

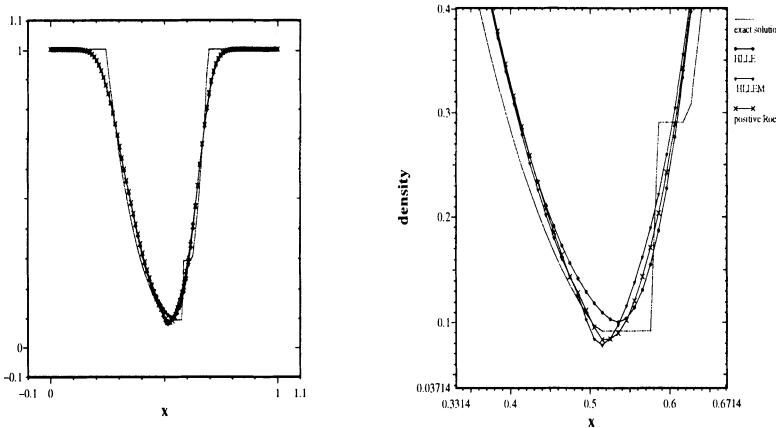


Figure 2. Strong rarefaction problem modified so as to obtain a non-symmetric solution.

References

- DUBROCA B (1998). Positively conservative Roe's matrix for Euler equations. Proc. 16th International Conference on Numerical Methods in Fluid Dynamics, Bruneau C-H (Editor), Springer-Verlag, p 272.
- EINFELDT B (1988). On Godunov-type methods for gas dynamics. *SIAM J. Numer. Anal.*, **25**, p 294.
- EINFELDT B, MUNZ C D, ROE PL, AND SJÖGREEN B (1991). On Godunov-type methods near low densities. *J. Comput. Phys.*, **92**, p 273.
- HARTEN A (1983). High resolution schemes for hyperbolic conservation laws. *J. Comput. Phys.*, **49**, p 357.
- HARTEN A, AND HYMAN J M (1983). Self adjusting grid methods for one-dimensional hyperbolic conservation laws. *J. Comput. Phys.*, **50**, p 253.
- LEVEQUE R J (1992). Numerical Methods for Conservation Laws. Birkhäuser Verlag.
- ROE P L (1981). Approximate Riemann solvers, parameter vectors and difference schemes. *J. Comput. Phys.*, **43**, p 357.
- WADA Y (1993). An improvement of the HLLEM scheme and its extension to chemically reacting gas. 2nd U.S. Nat. Congr. on Computational Mechanics, Washington, D.C.

BICHARACTERISTIC METHODS FOR MULTI-DIMENSIONAL HYPERBOLIC SYSTEMS

M. H. PHAM, M. RUDGYARD[†] AND E. SÜLI[‡]

*Oxford University Computing Laboratory, Wolfson Building,
Parks Road, Oxford, OX1 3QD, UK*

emails: minh.pham@comlab.ox.ac.uk,

† mike.rudgyard@comlab.ox.ac.uk,

‡ endre.suli@comlab.ox.ac.uk

Abstract. The main distinction between one-dimensional and multi-dimensional hyperbolic systems is that in the case of the latter information propagates in an infinite number of directions. In this paper we present a new class of bicharacteristic methods which take all the infinitely many directions of propagation into account. The principal issue involved in constructing bicharacteristic methods reduces to evolving the system of equations approximately along the bicharacteristics. For the spatial approximation we mainly concentrate on a *discontinuous Galerkin* finite element discretisation. We exemplify the technique on Maxwell's equations with constant coefficients in two space dimensions.

1. Introduction

Bicharacteristic theory, which is also commonly referred to as *characteristic* theory in the case of two independent variables, plays an important role in the design of numerical methods for hyperbolic problems. For problems with two independent variables, characteristic theory can be used for the design and analysis of numerical methods. However, the extension of the theory to systems of hyperbolic equations in several space dimensions is neither simple nor straightforward. The general argument is that multi-dimensional problems are qualitatively different from those in one-dimension, in the sense that there are infinitely many (rather than two) possible directions along which waves can propagate, and in particular the solution can no longer be regarded as constant along the characteristics (or bicharacteristics). Descriptions of bicharacteristic-based schemes up to the 1960s can

be found in (Butler, 1960; Chushkin, 1968; Cline and Hoffman, 1972; Holt, 1956). More recent work is discussed in (Lukáčová-Medvid'ová *et al.*, 2000; Ostkamp, 1995; Reddy *et al.*, 1982; Roe, 1998).

We shall not proceed with general hyperbolic systems, but concentrate on Maxwell's equations in two space dimensions.

2. Maxwell's Equations in Two Dimensions

Maxwell's equations in two space dimensions, with *electric field* vector $\vec{E} = (E_1, E_2)$ and scalar *magnetic field* H , can be written as

$$\frac{\partial \mathbf{U}}{\partial t} + \mathbf{A}_1 \frac{\partial \mathbf{U}}{\partial x} + \mathbf{A}_2 \frac{\partial \mathbf{U}}{\partial y} = 0, \quad (1)$$

where

$$\mathbf{U} = \begin{bmatrix} H \\ E_1 \\ E_2 \end{bmatrix} (\vec{x}, t), \quad \mathbf{A}_1 = \begin{bmatrix} 0 & 0 & c^2 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 0 & -c^2 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

with c assumed to be a positive constant (for simplicity).

Let $\vec{\mathbf{A}} := (\mathbf{A}_1, \mathbf{A}_2)$, $\vec{n} := (\cos \theta, \sin \theta)$, and $[\vec{\mathbf{A}} \cdot \vec{n}] := \mathbf{A}_1 \cos \theta + \mathbf{A}_2 \sin \theta$ for $\theta \in [0, 2\pi]$. The system (1) can be seen to be hyperbolic in the sense that the characteristic matrix, $[\vec{\mathbf{A}} \cdot \vec{n}]$, has real eigenvalues $0, \pm c$ and a complete set of linearly independent right eigenvectors.

Let $\mathbf{R}_{\vec{n}}$ be defined as the matrix of the right column eigenvectors of $[\vec{\mathbf{A}} \cdot \vec{n}]$ and let $\Lambda_{\vec{n}} = \text{diag}(0, c, -c) \equiv \text{diag}(\lambda_1, \lambda_2, \lambda_3)$ be the corresponding diagonal matrix of eigenvalues. A simple calculation shows that

$$\mathbf{R}_{\vec{n}} = \begin{bmatrix} 0 & c & -c \\ \cos \theta & -\sin \theta & -\sin \theta \\ \sin \theta & \cos \theta & \cos \theta \end{bmatrix} = [\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3],$$

where \mathbf{r}_κ , $\kappa = 1, 2, 3$, are the right column eigenvectors. The characteristic variable \mathbf{W} is defined as

$$\partial \mathbf{W} = \mathbf{R}_{\vec{n}}^{-1} \partial \mathbf{U} \quad \Rightarrow \quad \partial \mathbf{U} = \mathbf{R}_{\vec{n}} \partial \mathbf{W} = \sum_{\kappa} \mathbf{r}_\kappa \partial w_\kappa.$$

Here, ∂ denotes a derivative with respect to any independent variable. Left-multiplying (1) by $\mathbf{R}_{\vec{n}}^{-1}$, and using the characteristic variable, yields

$$\frac{\partial \mathbf{W}}{\partial t} + \Lambda_{\vec{n}}(\vec{n} \cdot \vec{\nabla}) \mathbf{W} = -\mathbf{R}_{\vec{n}}^{-1} [\vec{\mathbf{A}} \cdot \vec{s}] (\vec{s} \cdot \vec{\nabla}) \mathbf{U}, \quad (2)$$

where $\vec{s} := (\sin \theta, -\cos \theta)$. For Maxwell's equations the left-hand side of (2) defines the equations along the bicharacteristic directions; the right-hand side defines the coupling of these equations.

2.1. BICHARACTERISTIC-BASED FORMULATION

In expanded form, the system (2) can be written as

$$\begin{aligned}\frac{\partial w_1}{\partial t} &= -(\vec{s} \cdot \vec{\nabla}) H, \\ \frac{\partial w_{2,3}}{\partial t} \pm c \vec{n} \cdot \vec{\nabla} w_{2,3} &= \mp \frac{c}{2} \vec{n} \cdot (\vec{s} \cdot \vec{\nabla}) \vec{E},\end{aligned}\quad (3)$$

where $\mathbf{W} = [\vec{n} \cdot \vec{E}, -(c \vec{s} \cdot \vec{E} - H)/2c, -(c \vec{s} \cdot \vec{E} + H)/2c]^T$. The bicharacteristics $\vec{X}_\kappa(\vec{x}, \tau; t)$, $\kappa = 1, 2, 3$, are defined as the solutions of the ODEs

$$\frac{d\vec{X}_\kappa}{dt} = \vec{n} \lambda_\kappa, \quad \text{s.t. } \vec{X}_\kappa(\vec{x}, \tau; \tau) = \vec{x}, \quad \kappa = 1, 2, 3;$$

thus, $\vec{X}_1(\vec{x}, \tau; t) = \vec{x}$ and $\vec{X}_{2,3}(\vec{x}, \tau; t) = \vec{x} \mp c \vec{n}(\tau - t)$. Multiplying the system (3), *along the bicharacteristics*, by $\mathbf{R}_{\vec{n}}$, and accounting for the infinitely many directions of propagation by integrating w.r.t. $\theta \in [0, 2\pi]$, we ascertain

$$\frac{d}{dt} \left\{ \frac{1}{2\pi} \int_0^{2\pi} \mathbf{R}_{\vec{n}} \begin{bmatrix} w_1(\vec{x}, t; \vec{n}) \\ w_2(\vec{X}_2(t), t; \vec{n}) \\ w_3(\vec{X}_3(t), t; \vec{n}) \end{bmatrix} d\theta \right\} = \frac{1}{2\pi} \int_0^{2\pi} \mathbf{R}_{\vec{n}} \begin{bmatrix} -(\vec{s} \cdot \vec{\nabla}_x) H(\vec{x}, t) \\ -\frac{c}{2} \vec{n} \cdot (\vec{s} \cdot \vec{\nabla}_{X_2}) \vec{E}(\vec{X}_2(t), t) \\ \frac{c}{2} \vec{n} \cdot (\vec{s} \cdot \vec{\nabla}_{X_3}) \vec{E}(\vec{X}_3(t), t) \end{bmatrix} d\theta, \quad (4)$$

where $\vec{\nabla}_{X_\kappa} := (\partial/\partial X_\kappa, \partial/\partial Y_\kappa)$, $\kappa = 1, 2, 3$, and $\vec{X}_\kappa(t) \equiv \vec{X}_\kappa(\vec{x}, \tau; t)$, $\kappa = 2, 3$.

Let $\mathcal{L}(\mathbf{U}, \vec{x}, t)$ denote the right-hand side of (4). Then, $\mathcal{L}(\mathbf{U}, \vec{x}, t)$ can be simplified by considering the case $t = \tau$ and $t \neq \tau$ separately. For $t \neq \tau$, we observe the relation $\vec{s} \cdot \vec{\nabla}_{X_\kappa} = 1/(\lambda_\kappa(\tau - t)) d/d\theta$, $\kappa = 2, 3$. Furthermore, we shall make use of the following lemma, whose proof is rather simple and is therefore omitted.

Lemma 1 *Let $V(\vec{x}(\theta)) \in C[0, 2\pi]$. Let $\vec{X}^+ := \vec{x}_o + a \vec{n}$ and $\vec{X}^- := \vec{x}_o - a \vec{n}$, where $\vec{n} = (\cos \theta, \sin \theta)$, $\vec{x}_o \in \mathbb{R}^2$ a constant vector and a is a constant. Then*

$$\int_0^{2\pi} \sin^k \theta \cos^l \theta V(\vec{X}^+) d\theta = (-1)^{k+l} \int_0^{2\pi} \sin^k \theta \cos^l \theta V(\vec{X}^-) d\theta,$$

where $k, l \in \mathbb{N} \cup \{0\}$. □

Now, recalling the definition of $\mathbf{R}_{\vec{n}}$ and integrating by parts on the right-hand side of (4), after some algebra we find that

$$\mathcal{L}(\mathbf{U}, \vec{x}, t) = \frac{1}{2} \begin{cases} \begin{bmatrix} 0 \\ \frac{\partial H}{\partial y} \\ -\frac{\partial H}{\partial x} \end{bmatrix}(\vec{x}, t) - \frac{1}{\pi(\tau - t)} \int_0^{2\pi} (\mathbf{r}_1 \vec{n} + \mathbf{r}_3 \vec{s}) \cdot \vec{E}(\vec{X}_3(t), t) d\theta, & \text{if } t \neq \tau \\ - \left(\mathbf{A}_1 \frac{\partial \mathbf{U}}{\partial x} + \mathbf{A}_2 \frac{\partial \mathbf{U}}{\partial y} \right)(\vec{x}, \tau) & \text{if } t = \tau. \end{cases} \quad (5)$$

2.2. TEMPORAL DISCRETISATION

Equation (4) is a finite set of ordinary differential equations. We shall discretise these using a class of *second*-order *explicit* Runge-Kutta methods with obvious extension to higher order, if required. Consider solving

$$\frac{d}{dt}Z(t) = \mathcal{L}(Z, t),$$

based on the second order scheme:

$$\begin{aligned} Z^{n+\alpha} &= Z^n + \alpha \Delta t \mathcal{L}(Z^n, t^n), \\ Z^{n+1} &= Z^n + \left(1 - \frac{1}{2\alpha}\right) \Delta t \mathcal{L}(Z^n, t^n) + \frac{1}{2\alpha} \Delta t \mathcal{L}(Z^{n+\alpha}, t^n + \alpha \Delta t), \end{aligned} \quad (6)$$

where the parameter α is chosen so that $0 < \alpha \leq 1$. Taking $\alpha = 1/2$, we obtain the midpoint method; the value $\alpha = 1$ gives the classical Heun's method.

2.2.1. Heun's Discretisation in Time

Firstly, we shall state the semi-discrete scheme based on the forward Euler time discretisation.

Scheme 1 (Semi-discrete) *Applying the forward Euler discretisation in time for $t \in [t^n, t^{n+\alpha}]$, $\alpha > 0$, to (4) whose right-hand side has been rewritten using (5), we obtain the first-order time discretisation for Maxwell's equations:*

$$\begin{aligned} \begin{bmatrix} H \\ E_1 \\ E_2 \end{bmatrix}^{n+\alpha}(\vec{x}) &= \frac{1}{2} \begin{bmatrix} 0 \\ E_1 \\ E_2 \end{bmatrix}^n(\vec{x}) + \frac{1}{2\pi c} \int_0^{2\pi} \begin{bmatrix} c \\ \sin \theta \\ -\cos \theta \end{bmatrix} H^n(\vec{\mathcal{X}}_3^n) d\theta \\ &+ \frac{1}{2} \alpha \Delta t \begin{bmatrix} 0 \\ \partial/\partial y \\ -\partial/\partial x \end{bmatrix} H^n(\vec{x}) + \frac{1}{4\pi} \int_0^{2\pi} \begin{bmatrix} 4c\bar{s} \\ (1 - 3\cos 2\theta, -3\sin 2\theta) \\ (-3\sin 2\theta, 1 + 3\cos 2\theta) \end{bmatrix} \cdot \vec{E}^n(\vec{\mathcal{X}}_3^n) d\theta; \end{aligned} \quad (7)$$

we write this in compact form as

$$\mathbf{U}^{n+\alpha}(\vec{x}) := \mathbf{B}_E^\alpha(\mathbf{U}^n, \vec{\mathcal{X}}_3^n, \vec{x}),$$

where \mathbf{B}_E^α is defined by the right-hand side of (7), and $\vec{\mathcal{X}}_3(\vec{x}, t^{n+\alpha}; t) = \vec{x} + c(t^{n+\alpha} - t)\vec{n}$.

Applying the trapezoidal rule to the time integral of (4) over $[t^n, t^{n+1}]$, we obtain the following implicit second-order time discretisation of Maxwell's

equations:

$$\begin{aligned} \begin{bmatrix} H \\ E_1 \\ E_2 \end{bmatrix}^{n+1}(\vec{x}) &= \frac{1}{2} \begin{bmatrix} 0 \\ E_1 \\ E_2 \end{bmatrix}^n(\vec{x}) + \frac{1}{4\pi} \int_0^{2\pi} \begin{bmatrix} 3c\vec{s} \\ (1 - 2\cos 2\theta, -2\sin 2\theta) \\ (-2\sin 2\theta, 1 + 2\cos 2\theta) \end{bmatrix} \cdot \vec{E}^n(\vec{X}_3^n) d\theta \\ &\quad + \frac{1}{2\pi c} \int_0^{2\pi} \begin{bmatrix} c \\ \sin \theta \\ -\cos \theta \end{bmatrix} H^n(\vec{X}_3^n) d\theta + \frac{\Delta t}{4} \begin{bmatrix} 0 \\ \partial/\partial y \\ -\partial/\partial x \end{bmatrix} H^n(\vec{x}) - \frac{\Delta t}{4} [\vec{A} \cdot \vec{\nabla}] U^{n+1}(\vec{x}). \end{aligned} \quad (8)$$

Scheme 2 (Semi-discrete) Using Scheme 1 and Equation (8), we arrive at Heun's discretisation in time:

$$\begin{aligned} \mathbf{U}^{n+1}(\vec{x}) &:= \mathbf{B}_E^1(U^n, \vec{X}_3^n, \vec{x}), \\ U^{n+1}(\vec{x}) &:= \mathbf{B}_T(U^n, \mathbf{U}^{n+1}, \vec{X}_3^n, \vec{x}), \end{aligned}$$

where $\mathbf{U} = [\mathcal{H}, \mathcal{E}_1, \mathcal{E}_2]^T$, and \mathbf{B}_T is defined by the right-hand side of (8).

2.2.2. Explicit Second-order Discretisation in Time

Scheme 3 (Semi-discrete) For $0 < \alpha < 1$, we have the following second-order semi-discrete scheme,

$$\begin{aligned} \mathbf{U}^{n+\alpha}(\vec{x}) &:= \mathbf{B}_E^\alpha(U^n, \vec{X}_3^n, \vec{x}), \\ U^{n+1}(\vec{x}) &:= \mathbf{B}_{SE}^\alpha(U^n, \mathbf{U}^{n+\alpha}, \vec{X}_3, \vec{x}), \end{aligned}$$

where $\vec{X}_3^n = \vec{x} + c\alpha\Delta t \vec{n}$, $\vec{X}_3(\vec{x}, t^{n+1}; t) = \vec{x} + c(t^{n+1} - t)\vec{n}$, and

$$\begin{aligned} \mathbf{B}_{SE}^\alpha(U^n, \mathbf{U}^{n+\alpha}, \vec{X}_3, \vec{x}) &:= \frac{1}{2} \begin{bmatrix} 0 \\ E_1 \\ E_2 \end{bmatrix}^n(\vec{x}) + \frac{1}{2\pi c} \int_0^{2\pi} \begin{bmatrix} c \\ \sin \theta \\ -\cos \theta \end{bmatrix} H^n(\vec{X}_3^n) d\theta \\ &\quad + \frac{1}{2}\Delta t(1-\gamma) \begin{bmatrix} 0 \\ \partial/\partial y \\ -\partial/\partial x \end{bmatrix} H^n(\vec{x}) + \frac{1}{4\pi} \int_0^{2\pi} \begin{bmatrix} (1+\sigma)c\vec{s} \\ (1-\sigma\cos 2\theta, -\sigma\sin 2\theta) \\ (-\sigma\sin 2\theta, 1+\sigma\cos 2\theta) \end{bmatrix} \cdot \vec{E}^n(\vec{X}_3^n) d\theta \\ &\quad + \frac{1}{2}\gamma\Delta t \begin{bmatrix} 0 \\ \partial/\partial y \\ -\partial/\partial x \end{bmatrix} H^{n+\alpha}(\vec{x}) + \frac{1}{2\pi} \int_0^{2\pi} \left(\frac{\gamma}{1-\alpha} \right) \begin{bmatrix} c\vec{s} \\ (-\cos 2\theta, -\sin 2\theta) \\ (-\sin 2\theta, \cos 2\theta) \end{bmatrix} \cdot \vec{E}^{n+\alpha}(\vec{X}_3^{n+\alpha}) d\theta, \end{aligned}$$

where $\sigma := 3 - 1/\alpha$ and $\gamma := 1/2\alpha$.

Remark 1 To implement the above scheme, we have two options:

1. Calculate $\mathbf{U}^{n+\alpha}$ and then compute \mathbf{U}^{n+1} . Practically, this expands the stencil of the scheme, see Figure 1(a).
2. Substitute $\mathbf{U}^{n+\alpha}$ directly into the formula for \mathbf{U}^{n+1} . This gives a more compact scheme - information is obtained from within the characteristic cone. See Figure 1(b).

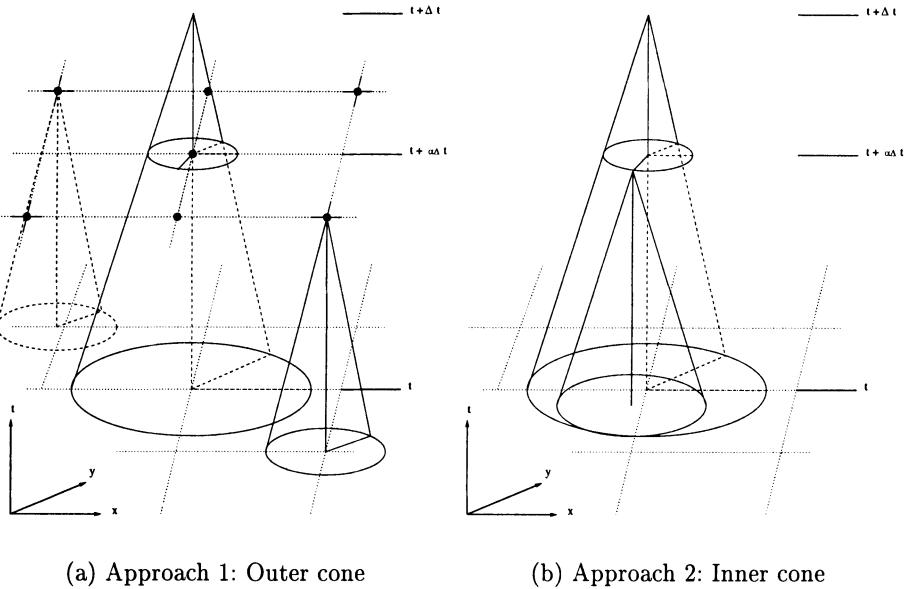


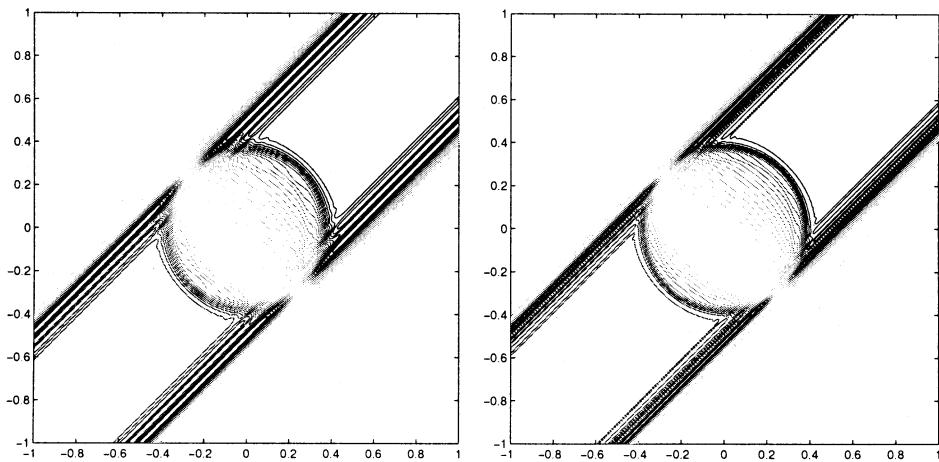
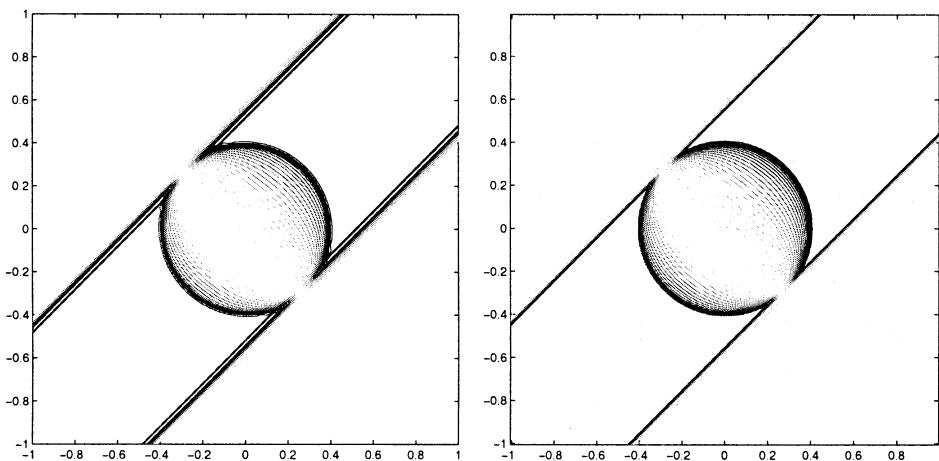
Figure 1. Characteristic cones for the explicit second-order Runge-Kutta time discretisation.

2.3. SPATIAL DISCRETISATION

The discretisation in space for the various schemes can be accomplished in a number of ways. The simplest is to *interpolate* locally by using continuous piecewise bilinear functions over a square mesh in the (x, y) -plane, leading to a first-order scheme, or continuous piecewise biquadratic functions, leading to a spatially second-order scheme. Another spatial discretisation is in the context of Finite Element (FE) methods. We shall assume a piecewise constant or discontinuous piecewise linear representation of the solution over a square mesh in the (x, y) -plane. This leads to a discontinuous Galerkin FE scheme of order one and two, respectively. The advantage in using discontinuous piecewise representation is that it provides better resolution of potential discontinuities in the solution and furnishes a more localised scheme.

Lemma 2 *Using the forward Euler time discretisation and interpolating using piecewise linears in space, or piecewise constants in the FE approach, the resulting first-order difference scheme is conditionally stable.*

Lemma 3 *Using the explicit two-stage (second-order) time discretisation, interpolating over the base of the characteristic cone using quadratic functions, and applying approach 2 (Remark 1), reduces the scheme to a Lax-Wendroff scheme.*

(a) Lax-Wendroff on square meshes with elements 200×200 , 400×400 (b) DEG_{P1}^H on square meshes with elements 200×200 , 400×400 *Figure 2.* Discontinuous problem: $\nu = 0.4$, $T = 0.4$, domain $= [-1, 1]^2$.

3. Numerical Results and Conclusion

For the numerical test problem, we shall consider the following discontinuous initial data:

$$H(\vec{x}, 0) = 0, \quad E_1(\vec{x}, 0) = E_2(\vec{x}, 0) = \frac{1}{\sqrt{2}} \begin{cases} 1, & |y| < |x|, \\ -1, & \text{elsewhere.} \end{cases}$$

The computational domain $[-1, 1] \times [-1, 1]$ is partitioned uniformly into 200×200 and 400×400 elements respectively, and the final time set to be $T = 0.4$. The Courant number is fixed at $\nu = 0.4$.

Figure 2 shows the contours of H using the Lax-Wendroff scheme (LW) and Heun's method (approach 1) with discontinuous piecewise linear (FE) spatial discretisation (DEG_{P1}^H). As we can see, DEG_{P1}^H gives a much better resolution of the discontinuities. Further analysis is needed in order to determine the stability of the scheme. Numerical experiments seem to suggest that DEG_{P1}^H is stable for a wide range of Courant numbers in $[0, 1]$.

We have shown that the application of approach 1 gives satisfactory results. Our present work concerns the implementation based on approach 2 (inner-cone); the outcome of that research will be published elsewhere.

Acknowledgements

The first author's work was supported by an EPSRC research studentship and by a CASE award studentship with DERA.

References

- Butler D S (1960). The Numerical Solution of Hyperbolic Systems of Partial Differential Equations in Three Independent Variables. *Proceeding of the Royal Society of London, Series A. Mathematical and Physical Sciences*, **255**.
- Chushkin P I (1968). Numerical Method of Characteristics for Three-Dimensional Supersonic Flows, *Progress in Aeronautical Science*, **9**.
- Cline M C and Hoffman J D (1972). Comparison of Characteristic Schemes for Three-dimensional, Steady, Isentropic Flow. *AIAA Journal*, **10**.
- Holt M (1956). The Methods of Characteristics for Steady Supersonic Rotational Flow in Three Dimensions, *Journal of Fluid Mechanics*, **1**.
- Lukáčová-Medvid'ová M, Morton K W and Warnecke G (2000). Evolution Galerkin Methods for Hyperbolic Systems in Two Space Dimensions, *Mathematics of Computation* (to appear).
- Ostkamp S (1995). Multidimensional Characteristic Galerkin Schemes and Evolution Operators for Hyperbolic Systems. *Ph.D. thesis, Universität Hannover*.
- Reddy A S, Tikekar V G and Prasad P (1982). Numerical Solution of Hyperbolic Equations by Methods of Bicharacteristic. *Journal of Mathematical and Physical Sciences*, **6**.
- Roe P (1998). Linear Bicharacteristic Schemes without Dissipation. *Siam Journal on Scientific Computing*, **19**.

AN EXACT RIEMANN SOLVER FOR MULTIDIMENSIONAL SPECIAL RELATIVISTIC HYDRODYNAMICS

J. A. PONS AND J. M. MARTI

*Departament d'Astronomia i Astrofísica,
Universitat de València, 46100 Burjassot, Spain
e-mail: jose.a.pons@uv.es*

AND

E. MUELLER

*MPI für Astrophysik,
Karl-Schwarzschild-Str. 1, 85748 Garching, Germany*

Abstract.

We have generalised the exact solution of the Riemann problem in special relativistic hydrodynamics (Martí and Müller, 1994) for arbitrary tangential flow velocities. The solution is obtained by solving the jump conditions across shocks plus an ordinary differential equation arising from the self-similarity condition along rarefaction waves, in a similar way as in purely normal flow. This solution has been used to build up an exact Riemann solver implemented in a multidimensional relativistic (Godunov-type) hydro-code.

1. Introduction

The decay of a discontinuity separating two constant initial states (*Riemann problem*) has played a very important role in the development of numerical hydrodynamic codes in classical (Newtonian) hydrodynamics after the pioneering work of Godunov (Godunov, 1959). Nowadays, most modern high-resolution shock-capturing methods (Leveque, 1992) are based on the exact or approximate solution of Riemann problems between adjacent numerical cells and the development of efficient Riemann solvers has become a research field in numerical analysis in its own (Toro, 1999).

Riemann solvers began to be introduced in numerical relativistic hydrodynamics at the beginning of the nineties and, presently, the use of high-resolution shock-capturing methods based on Riemann solvers is considered as the best strategy to solve the equations of relativistic hydrodynamics in nuclear physics (heavy ion collisions) and astrophysics (stellar core collapse, supernova explosions, extragalactic jets, gamma-ray bursts). This fact has caused a rapid development of Riemann solvers for both special and general relativistic hydrodynamics (Ibáñez and Martí, 1999; Martí and Müller, 1996).

The main idea behind the solution of a Riemann problem (defined by two constant initial states, L and R , at left and right of their common contact surface) is that the self-similarity of the flow through rarefaction waves and the Rankine-Hugoniot relations across shocks allow one to connect the intermediate states I_* ($I = L, R$) with their corresponding initial states, I . The analytical solution of the Riemann problem in classical hydrodynamics (Courant and Friedrichs, 1948) rests on the fact that the normal velocity in the intermediate states, $v_{I_*}^n$, can be written as a function of the pressure p_{I_*} in that state (and the flow conditions in state I). Thus, once p_{I_*} is known, $v_{I_*}^n$ and all other unknown state quantities of I_* can be calculated.

In the case of relativistic hydrodynamics the same procedure can be followed (Martí and Müller, 1994; Pons, Martí and Müller, 2000), the major difference with classical hydrodynamics stemming from the role of tangential velocities. While in the classical case the decay of the initial discontinuity does not depend on the tangential velocity (which is constant across shock waves and rarefactions), in relativistic calculations the components of the flow velocity are coupled through the presence of the Lorentz factor. An iterative nonlinear relativistic Riemann solver that takes into account the effects of nonvanishing tangential velocity components has already been implemented (Falle and Komissarov, 1996) for Riemann problems involving two strong rarefactions. Other authors (Balsara, 1994) have developed a Riemann solver based on the two-shock approximation where the rarefaction wave is treated as a shock wave. In this work we build a Riemann solver based on the exact solution of a general Riemann problem in Minkowski spacetime with arbitrary tangential velocities.

2. The equations of relativistic hydrodynamics

The equations of relativistic hydrodynamics admit the following conservative formulation

$$\partial_t \mathbf{U} + \partial_i \mathbf{F}^{(i)} = 0 \quad (1)$$

where \mathbf{U} and $\mathbf{F}^{(i)}(\mathbf{U})$ ($i = 1, 2, 3$) are, respectively, the vectors of conserved variables and fluxes

$$\mathbf{U} = (D, S^1, S^2, S^3, \tau)^T \quad (2)$$

$$\mathbf{F}^{(i)} = (Dv^i, S^1 v^i + p\delta^{1i}, S^2 v^i + p\delta^{2i}, S^3 v^i + p\delta^{3i}, S^i - Dv^i)^T. \quad (3)$$

The conserved variables (the rest-mass density, D , the momentum density, S^i , and the energy density τ) are defined in terms of the *primitive variables* (proper rest-mass density, ρ , velocity components, v^i , and specific internal energy, ε), according to

$$D = \rho W, \quad S^i = \rho h W^2 v^i, \quad \tau = \rho h W^2 - p - D \quad (4)$$

where p is the pressure, $W = (1 - v^2)^{-1/2}$ is the Lorentz factor and $h = 1 + \varepsilon + p/\rho$ the specific enthalpy. We use units in which the speed of light is set to unity.

3. Relation between the normal flow velocity and pressure behind relativistic rarefaction waves

Choosing the surface of discontinuity to be normal to the x -axis, rarefaction waves are self-similar solutions of the flow equations depending only on the combination $\xi = x/t$. Getting rid of all the terms with y and z derivatives in equations (1) and substituting the derivatives of x and t in terms of the derivatives of ξ , the system of equations can be reduced to just one ordinary differential equation (ODE) and two algebraic conditions

$$\rho h W^2 (v^x - \xi) dv^x + (1 - \xi v^x) dp = 0 \quad (5)$$

$$h W v^{y,z} = \text{constant}, \quad (6)$$

with ξ constrained by

$$\xi = \frac{v^x (1 - c_s^2) \pm c_s \sqrt{(1 - v^2)[1 - v^2 c_s^2 - (v^x)^2 (1 - c_s^2)]}}{1 - v^2 c_s^2}, \quad (7)$$

because non-trivial similarity solutions exist only if the Wronskian of the original system vanish. We have denoted by c_s the speed of sound, provided by the equation of state. The plus and minus sign correspond to rarefaction waves propagating to the right (\mathcal{R}_\rightarrow) and left (\mathcal{R}_\leftarrow), respectively.

From equations (6) it follows that $v^y/v^z = \text{constant}$, i.e., the tangential velocity does not change direction across rarefaction waves. Notice that, in a kinematical sense, the Newtonian limit ($v^i \ll 1$) leads to $W = 1$, but equations (6) do not reduce to the classical limit $v^{y,z} = \text{constant}$, because the specific enthalpy still couples the tangential velocities. Thus, even for

slow flows, the Riemann solution presented in this paper must be employed for thermodynamically relativistic situations ($h \gg 1$).

Using (7) and the definition of the sound speed, the ODE (5) can be written as

$$\frac{dv^x}{dp} = \pm \frac{1}{\rho h W^2 c_s} \frac{1}{\sqrt{1 + g(\xi_{\pm}, v^x, v^t)}} \quad (8)$$

where $v^t = \sqrt{(v^y)^2 + (v^z)^2}$ is the absolute value of the tangential velocity and

$$g(\xi_{\pm}, v^x, v^t) = \frac{(v^t)^2 (\xi_{\pm}^2 - 1)}{(1 - \xi_{\pm} v^x)^2}. \quad (9)$$

Considering that in a Riemann problem the state ahead of the rarefaction wave is known, equation (8) can be integrated with the constraint $hWv^t = \text{constant}$, allowing us to connect the states ahead (*a*) and behind (*b*) the rarefaction wave. Thus the solution is only a function of p_b .

4. Relation between post-shock flow velocities and pressure for relativistic shock waves.

The Rankine-Hugoniot conditions relate the states on both sides of a shock and are based on the continuity of the mass flux and the energy-momentum flux across shocks (Taub, 1948; Taub, 1978; Königl, 1980). Considering the surface of discontinuity as normal to the x -axis, the invariant mass flux across the shock can be written as

$$j \equiv W_s D_a (V_s - v_a^x) = W_s D_b (V_s - v_b^x). \quad (10)$$

where V_s is the coordinate velocity of the hyper-surface that defines the position of the shock wave and W_s is the corresponding Lorentz factor, $W_s = (1 - V_s^2)^{1/2}$. According to our definition, j is positive for shocks propagating to the right. In terms of the mass flux, j , the Rankine-Hugoniot conditions are

$$\begin{aligned} [v^x] &= -\frac{j}{W_s} \left[\frac{1}{D} \right], & [p] &= \frac{j}{W_s} \left[\frac{S^x}{D} \right], \\ [hWv^{y,z}] &= 0, & [v^x p] &= \frac{j}{W_s} \left[\frac{\tau}{D} \right]. \end{aligned} \quad (11)$$

which implies that the quantities $hWv^{y,z}$ are constant across a shock wave and, hence, that the orientation of the tangential velocity does not change. The same result was obtained for rarefaction waves (see §3). Equations (11) can be manipulated to obtain v_b^x as a function of p_b , j and V_s . Using the

relation $S^x = (\tau + p + D)v^x$ and after some algebra, one finds

$$v_b^x = \left(h_a W_a v_a^x + \frac{W_s(p_b - p_a)}{j} \right) \left(h_a W_a + (p_b - p_a) \left(\frac{W_s v_a^x}{j} + \frac{1}{\rho_a W_a} \right) \right)^{-1}. \quad (12)$$

The final step is to express j and V_s as a function of the post-shock pressure. First, from the definition of the mass flux we obtain

$$V_s^\pm = \frac{\rho_a^2 W_a^2 v_a^x \pm |j| \sqrt{j^2 + \rho_a^2 W_a^2 (1 - v_a^x)^2}}{\rho_a^2 W_a^2 + j^2} \quad (13)$$

where V_s^+ (V_s^-) corresponds to shocks propagating to the right (left).

Second, from the Rankine-Hugoniot relations and the physical solution of h_b obtained from the Taub adiabat (Thorne, 1973) (the relativistic version of the Hugoniot adiabat), that relates only thermodynamic quantities on both sides of the shock, the square of the mass flux j^2 can be obtained as a function of p_b . Using the positive (negative) root of j^2 for shock waves propagating towards the right (left), the desired relation between the post-shock normal velocity v_b^x and the post-shock pressure p_b is obtained (Martí and Müller, 1994; Pons, Martí and Müller, 2000).

5. The Riemann Solver based on the exact solution.

The time evolution of a Riemann problem can be represented as:

$$I \rightarrow L \mathcal{W}_\leftarrow L_* \mathcal{C} R_* \mathcal{W}_\rightarrow R \quad (14)$$

where \mathcal{W} denotes a simple wave (shock or rarefaction), moving towards the initial left (\leftarrow) or right (\rightarrow) states. Between them, two new states appear, namely L_* and R_* , separated from each other through the third wave \mathcal{C} , which is a contact discontinuity.

As in the Newtonian case, the compressive character of shock waves (density and pressure rise across the shock) allows us to discriminate between shocks (\mathcal{S}) and rarefaction waves (\mathcal{R}):

$$\mathcal{W}_{\leftarrow(\rightarrow)} = \begin{cases} \mathcal{R}_{\leftarrow(\rightarrow)}, & p_b \leq p_a \\ \mathcal{S}_{\leftarrow(\rightarrow)}, & p_b > p_a \end{cases} \quad (15)$$

where p is the pressure and subscripts a and b denote quantities ahead and behind the wave. For the Riemann problem $a \equiv L(R)$ and $b \equiv L_*(R_*)$ for \mathcal{W}_\leftarrow and \mathcal{W}_\rightarrow , respectively.

The solution of the Riemann problem consists in finding the positions of the waves separating the four states and the intermediate states, L_* and

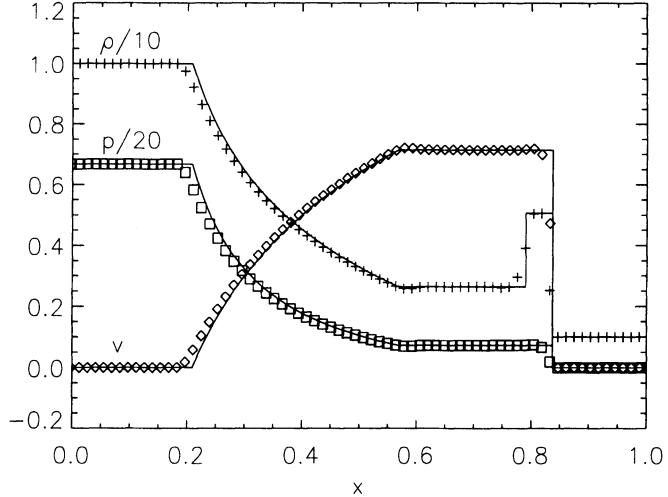


Figure 1. Exact (solid lines) and numerical profiles along the diagonal of pressure (squares), density (crosses) and normal velocity (diamonds) for the 2-dimensional relativistic shock tube discussed in the text.

R_* . The functions \mathcal{W}_\rightarrow and \mathcal{W}_\leftarrow allow one to determine the functions $v_{R*}^x(p)$ and $v_{L*}^x(p)$, respectively. The pressure p_* and the flow velocity v_*^x in the intermediate states are then given by the condition

$$v_{R*}^x(p_*) = v_{L*}^x(p_*) = v_*^x. \quad (16)$$

which is an implicit algebraic equation in p_* and can be solved by means of an iterative method. When p_* and v_*^x have been obtained, the equation of state gives the specific internal energy and the remaining state variables of the intermediate state I_* can be calculated using the relations between I_* and the respective initial state I given through the corresponding wave.

The influence of tangential velocities on the solution of a Riemann problem has been studied in a recent work (Pons, Martí and Müller, 2000), where it was found that the structure of the solution remains unchanged for different tangential velocities, although the values in the constant state may change by a large amount. Notice that the solution of the Riemann problem depends on the modulus, but not on the direction, of the tangential velocity.

A first interesting application of this solution is that it can be used to test the different approximate relativistic Riemann solvers and the multi-dimensional hydrodynamic codes based on directional splitting. Complementarily, it can be used to construct Riemann Solvers. As an example, we have simulated a relativistic tube (Sod, 1978), in a 100×100 Cartesian

grid, where the initial discontinuity was located in a main diagonal. The initial states were $\rho_L = 10$, $\rho_R = 1$, $p_L = 13.3$, $p_R = 0.66 \times 10^{-3}$, $v_L = 0$, $v_R = 0$, and an ideal gas equation of state with adiabatic index $\gamma = 5/3$ was employed. Spatial order of accuracy was set to second order by means of a monotonic piecewise linear reconstruction procedure and second order in time is obtained by using a Runge-Kutta method for time advancing. The exact solution of the Riemann problem is used at every interface to calculate the numerical fluxes. The results are shown in Figure 1. Profiles of all variables are stable and discontinuities are well resolved without excessive smearing.

Acknowledgements

This work has been partially supported by DGICYT PB97-1432. J.A.P. acknowledges a fellowship from the Ministerio de Educación y Cultura.

References

- Balsara D S (1994). Riemann Solver for Relativistic Hydrodynamics. *J. Comput. Phys.*, **114**, 284.
- Courant R and Friedrichs K O (1948). Supersonic Flows and Shock Waves. Interscience.
- Falle S A E G and Komissarov S S (1996). An Upwind Numerical Scheme for Relativistic Hydrodynamics with a General Equation of State. *Mon. Not. Royal Astron. Soc.* **278**, 586.
- Godunov S K (1959). A Finite Difference Method for the Computation of Discontinuous Solutions of the Equations of Fluid Dynamics. *Mat. Sb.* **47**, pp 357-393.
- Ibáñez J M and Martí J M (1999). Riemann Solvers in Relativistic Astrophysics. *J. Comput. Appl. Math.*, in press.
- Königl A (1980). Relativistic Gas Dynamics in Two Dimensions. *Phys. Fluids.*, **23**, 1083.
- LeVeque R J (1992). Numerical Methods for Conservation Laws (Second Edition) . Birkhäuser.
- Martí J M and Müller E (1994). The Analytical Solution of the Riemann Problem in Relativistic Hydrodynamics. *J. Fluid Mech.*, **258**, 317.
- Martí J M and Müller E (1996). Extension of the Piecewise Parabolic Method to One-Dimensional Relativistic Hydrodynamics. *J. Comput. Phys.* **123**, 1.
- Martí J M and Müller E (1999). Numerical Hydrodynamics in Special Relativity. *Living Reviews in Relativity*, Vol. 2; at <http://www.livingreviews.org/Articles/Volume2>
- Pons J A, Martí J M and Müller E (2000). The Exact Solution of the Riemann Problem with non-zero tangential velocities in Relativistic Hydrodynamics. *J. Fluid Mech.*, in press.
- Sod G A (1978). A Survey of Several Finite Difference Methods for Systems of Nonlinear Hyperbolic Conservation Laws. *J. Comput. Phys.*, **27**, 1.
- Taub A H (1948). Relativistic Rankine-Hugoniot Relations. *Phys. Rev.*, **74**, 328.
- Taub A H (1978). Relativistic Fluid Mechanics. *Ann. Rev. Fluid Mech.*, **10**, 32.
- Thorne K S (1973). Relativistic Shocks: The Taub Adiabat. *Astrophys. J.*, **179**, 897.
- Toro E F (1999). Riemann Solvers and Numerical Methods for Fluid Dynamics. Second Edition, Springer-Verlag.

EXPERIENCE WITH THE OSHER SCHEME FOR APPLIED AERODYNAMICS

N. QIN

*Centre for Computational Aerodynamics,
College of Aeronautics, Cranfield University,
Bedford MK43 0AL, England, UK.
Email: n.qin@cranfield.ac.uk*

Abstract.

This paper presents some of our experience in using Osher's approximate Riemann solver in solving the Navier-Stokes equations for high speed aerodynamic problems. Grid convergence in boundary layers, grid alignment with flow features in multi-dimensional applications and some anomalies in numerical simulations are discussed.

1. Introduction

Good shock capturing does not always imply good boundary layer capturing. When applying the shock capturing schemes to the Navier-Stokes solutions, care needs to be exercised to ensure accurate capturing of the viscous layers. Van Leer et al.(Van Leer, 1987) studied various flux formulae regarding their capabilities for viscous flows and demonstrated the importance of the choice of shock capturing methods for viscous flows.

2. Grid convergence for Navier-Stokes solutions

The capability in capturing the boundary layers or other viscous layers is reflected in the number of grid points required to resolve these layers accurately. Computationally, the smaller the number is, the more efficient the numerical simulation will be.

A numerical study has been carried out to illustrate the importance of the flux scheme on the grid convergence property of the solution. The test case is for a hypersonic laminar flow around a sharp cone at zero angle of

attack. The conditions for the test case follow an early experimental work with a 7-degree half cone angle, an incoming flow Mach number of 9.16, a free stream temperature of 59.8K, and a Reynolds number of 55 million per meter. The computation is carried out at station 0.044m from the cone tip. The wall is isothermal at 290K as in the experiment. Note that the case is similar to the case studied in (Van Leer, 1987), where an adiabatic wall boundary condition was specified.

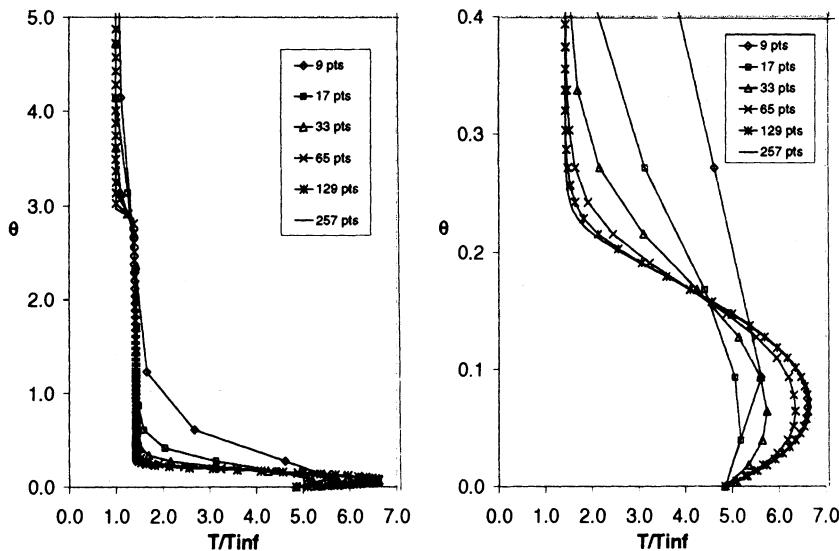


Figure 1. Grid convergence of van Leer flux vector splitting scheme: temperature profiles-overall(left) and in the boundary layer (right)

The inviscid flux terms are discretised using one of the two shock capturing schemes, namely, the van Leer flux vector splitting scheme (Van Leer, 1982) or the Osher approximate Riemann solver (Osher and Solomon, 1982). Van Leer's MUSCL approach is used to achieve a formal 3rd order accuracy for the inviscid discretisation.

Fig.1 shows the grid convergence for the flux vector splitting scheme regarding the temperature distributions in the flowfield. It can be seen that the shock wave is sharply resolved (with only one intermediate point).

On the other hand, the high sensitivity of the viscous feature of the solution to the grid density is also clearly illustrated. The coarse grid solutions diffuse the physical boundary layer, smoothing out both the temperature peak in the boundary layer and the knee at the edge of the thermal boundary layer. However, the *convergence property* of the numerical formulation guarantees that, as the grid is refined, it will approach a grid converged

result representing the solution of the governing equations. In the present case, a grid converged result has been achieved when the grid is refined to reach 257 points. As the grid is refined, the numerical diffusion is reduced. One can only obtain reliable viscous parameters such as skin friction coefficients and heat transfer rates after such a study has been carried out. This attributes to one of the major difficulties in the Navier-Stokes solutions in addition to the turbulence modelling issues.

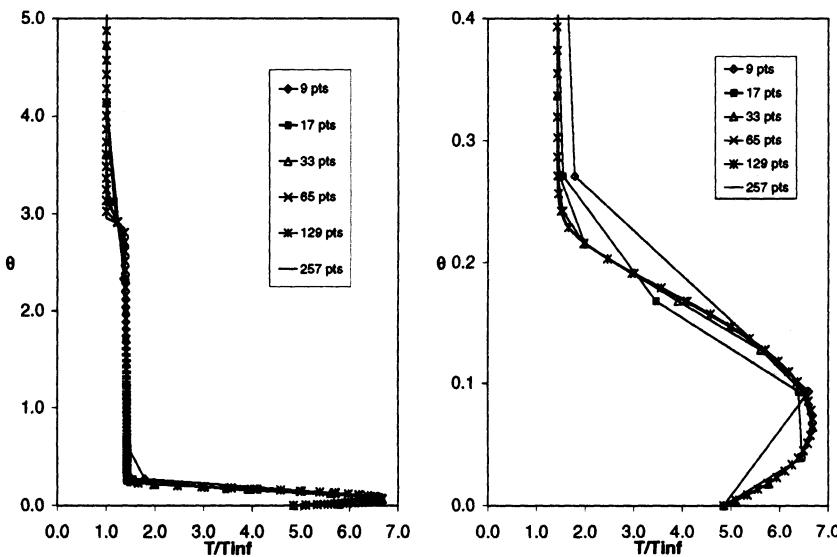


Figure 2. Grid convergence of Osher scheme: temperature profiles- overall(left) and in the boundary layer(right)

The picture shown in Fig.2 is fundamentally different from Fig.1 regarding the viscous features. Here the Osher approximate Riemann solver has been employed for the inviscid fluxes. The shock capturing property is very similar to the flux vector splitting but the boundary layer resolution illustrate some interesting features. Generally, the temperature boundary layer profiles become much less sensitive to the grid density as compared with the previous case. Amazingly, on the coarsest grid (9 points), the only one point inside the boundary layer is dotted nearly on the final converged boundary layer profile. (Note that all the marks in the figures show the solution at that grid point.) The results brighten up the gloomy picture drew from the previous method for Navier-Stokes solutions. If a proper numerical scheme is used, only a few grid points may be required to resolve the boundary layers or free shear layers accurately. The number of points

required to resolve the boundary layer needs only to be enough to describe the boundary layer profile.

3. Grid alignment and shock capturing in multidimension

Most Riemann solver based schemes to date is strictly one dimensional only from their derivation. One-dimensional approximate Riemann solvers are naturally extended to multidimensional applications by directional splitting or the finite volume approach, where the Riemann problems are solved locally normal to the cell interfaces.

Flow features are mostly direction-related, such as shock waves and shear layers. The weakness of the directional splitting approach is primarily due to the difficulties in resolving flow features oblique to the cell interfaces. If the grid used in the computational simulation can be aligned with the flow features, the high resolution features of the one dimensional Riemann solver based methods can be maintained and used as an efficient approach for multidimensional problems.

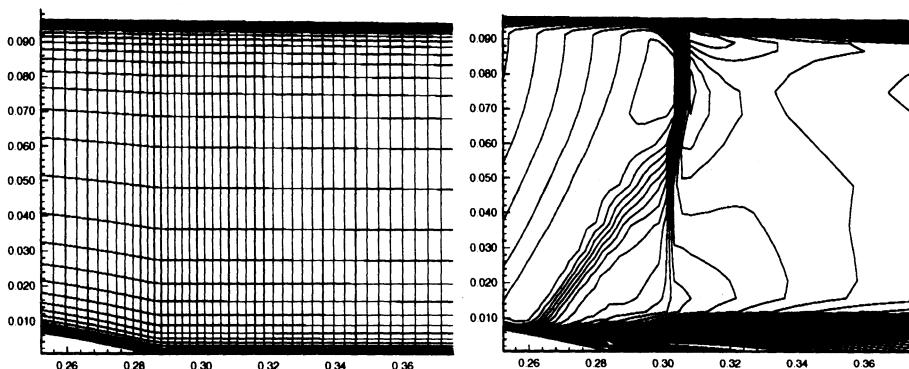


Figure 3. Regular grid and Mach number contours

Fig.3 shows a time-averaged Navier-Stokes solution (Qin and Zhu, 1999) for the Delery transonic bump problem on a pre-generated regular 65×85 grid. The SST two-equation turbulence model is coupled with the mean flow equations, which are solved using a slightly extended version of the Osher approximate Riemann solver. Capturing the λ -shock structure due to the strong shock/boundary layer interaction puts a severe test on the capability of both the turbulence model and the numerical method. It can be seen in Fig.3 that, although the general structure of the λ -shock has been captured, the oblique leg of the shock structure is poorly resolved due to the misalignment of the regular grid with the shock and inability of the one-dimensional Riemann solver to resolve oblique features with high

resolution. To improve the resolution of the solution without extra grid points, the grid is adapted using the solution obtained on the regular grid.

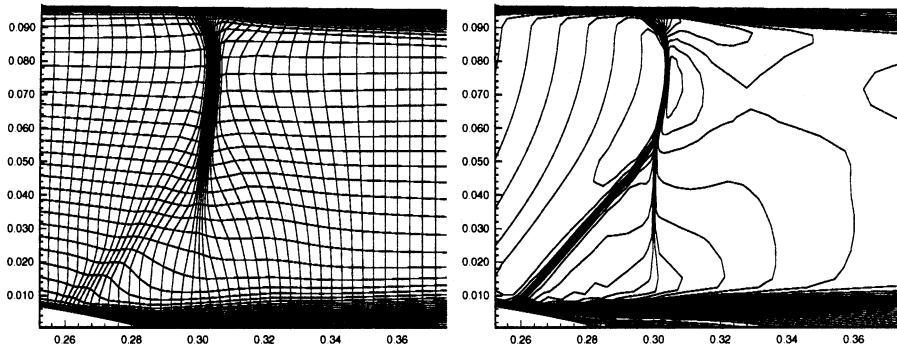


Figure 4. Adapted grid and Mach number contours

The resulting grid is shown in Fig.4. Two advantages of the grid can be observed: (1) the clustering of the grid points in the high flow gradient regions normal to the gradient directions, and (2) the alignment of the grid lines with the flow features. A good resolution of the flowfield is thus achieved on a relatively coarse grid for a complicated flow interaction problem.

4. The carbuncle problem and its cure

We have seen some good features of the Osher approximate Riemann solver in the previous two sections for applied aerodynamic applications. However, in a more recent study, we came across a problem which seems to be related to the carbuncle phenomenon identified by some previous researchers. The problem occurred when we tried to solve a hypersonic flow passing around a swept cylinder for the study of attachment line transition (Qin and Ludlow, 1999). The numerical analysis was carried out for laminar hypersonic viscous flows around a highly swept cylinder. The cylinder sweep angle is 66.5 deg, the freestream Mach number and temperature are 10.5 and 53K and the Reynolds number based on the cylinder diameter is 6.1×10^5 . An isothermal no-slip wall boundary condition at 294K is specified. The cylinder is cut flat along the vertical streamwise direction such that a sharp tip is formed.

In the numerical simulation of the above problem, a phenomenon similar to the carbuncle problem was shown in our Navier-Stokes solution (Qin and Ludlow, 1999). In the present computation, the phenomenon was only observed at certain downstream stations away from the tip region. The

anomaly spoils the solution along the attachment line, crucial for the transition study, as shown in the density and pressure contours on the left in Figs.5 and 6.

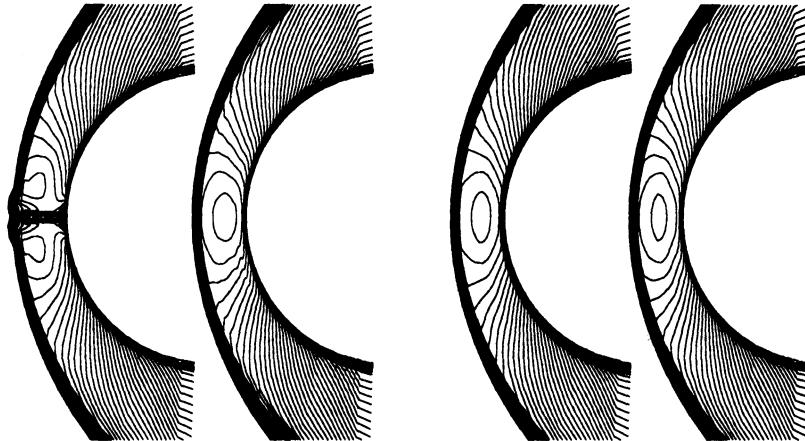


Figure 5. Density contours for Osher, AUSM+, AUSM+W and Hybrid schemes 5 diameters from the tip.

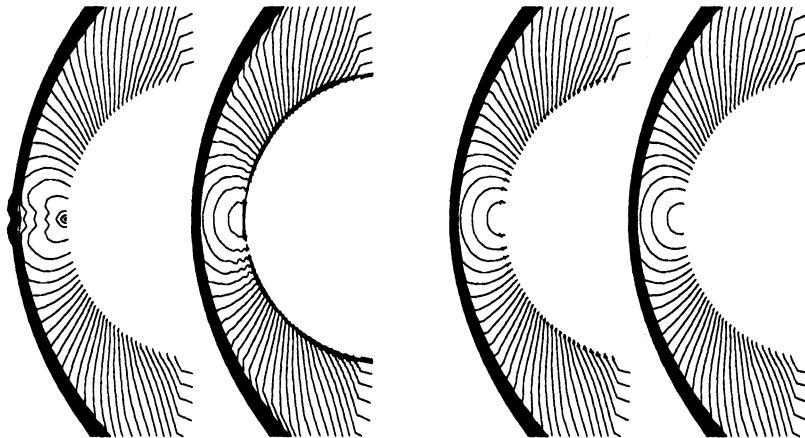


Figure 6. Pressure contours for Osher, AUSM+, AUSM+W and Hybrid schemes 5 diameters from the tip.

The carbuncle phenomenon was notoriously associated with the popular Roe approximate Riemann solver near the symmetry line for 2D blunt noses in Euler solutions (Quirk, 1994). A slight trace of such problem is also reported by (Wada and Liou, 1994) for the Osher solver for the same flow problem. The latter authors also tested both solvers for Navier-Stokes solutions for the 2D cylinder flow problem and no carbuncle problem was experienced. Gressier and Moschetta (Gressier and Moschetta, 1998) presented

recently an interesting analysis of the relation between exact capturing the contact discontinuities and the failings of some upwind schemes in Euler solutions. The present results, though disappointing, revealed an anomaly of the Osher solver for a particular class of flow problems, indicating that neither (1) the physical viscosity nor (2) the strong flow in the third dimension can prevent the carbuncle problem from occurring. The results also indicate that the carbuncle phenomenon is not just associated with 2D or 3D stagnation point but can occur for a wider class of hypersonic viscous flow problems when swept blunt leading edges exist.

Recently, Liou (Liou, 1995) proposed a new flux formula, AUSM+ motivated by the desire to combine the numerical efficiency and robustness of flux vector splitting with the accuracy in contact discontinuity resolution from approximate Riemann solvers. In addition to exact resolution of 1D contact and shock discontinuities, it is positivity preserving, free from oscillations for slowly moving shocks and free from the carbuncle phenomenon. Obviously the AUSM+ scheme became our subsequent choice for the hypersonic swept cylinder problem after the failure of the Osher solver. As promised, the AUSM+ flux formula overcomes the carbuncle problem in the present solution (Figs.5 and 6). However, strong pressure oscillations at the edge of the boundary layer were observed as shown in the pressure contours in Fig.6. The pressure oscillations are also shown in Fig.7 for the AUSM+ scheme. These oscillations spoil the whole boundary layer resolution including that along the attachment line. Also noticed in the pressure profile is the overshoot at the shock wave in the solution.

AUSM+W was proposed by Liou(Liou, 1995) along with AUSM+ to cure the overshoot problem of the latter for the collision of strong shocks. This scheme was tested in the present study to investigate its effectiveness on the shock overshoot and the boundary layer edge oscillation problems identified earlier with the AUSM+ solver. The overshoot of the pressure at the shock wave was cured by the AUSM+W scheme. It is also clear that the oscillations at the edge of the boundary layer were reduced but not fully cured(Figs.6 and 7). However inspection of the temperature profile in Fig.7 and the density contours in Fig.5 indicates that a sacrifice in the boundary layer resolution has been made for the gains. The boundary layer becomes thicker with an excessive numerical dissipation as compared with AUSM+. (Note that the Osher temperature profile is strongly influenced by the carbuncle problem along the symmetry line.)

The above failures led us to compose a hybrid scheme combining the good features of both Osher and AUSM+ schemes or, in other words, avoiding the anomalies exhibited in both schemes. The resulting scheme uses Osher numerical fluxes in the wall normal and spanwise direction and AUSM+ numerical fluxes in the direction wrapping around the cylinder.

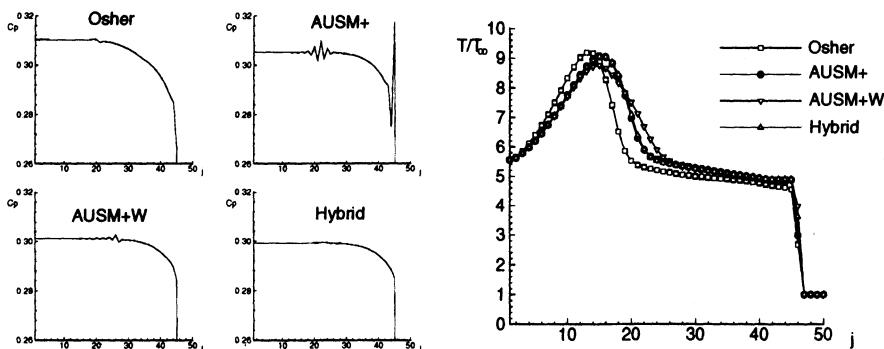


Figure 7. Pressure and temperature profiles along the symmetry line for Osher (top left), AUSM+ (top right), AUSM+W (bottom left) and hybrid (bottom right)

The results are shown on the right of Figs 5 and 6. The pressure and temperature profiles are plotted in Fig.7 in comparison with the other schemes. Both the anomalies, the carbuncle phenomenon and the boundary layer edge oscillations are cured without sacrificing the shock and boundary layer resolution.

5. Conclusions

Our experience with the Osher approximate Riemann solver has generally been satisfactory for various applications using Navier-Stokes equations. The scheme's excellent grid convergence feature has been revealed for boundary layers in viscous solutions, which makes it possible to use a small number of grid points to resolve viscous layers in N-S solutions. Adaptation of the grid lines to the flow features can offer a cost effective way for multi-dimensional applications of the one-dimensional Riemann solvers. However, for a wide range of engineering applications, practical problems may still arise, such as the carbuncle problem and the boundary layer edge pressure oscillation encountered in the swept cylinder case, and a good understanding of the features of the numerical schemes will help to cure them.

6. Acknowledgements

I would like to thank research staff and students in the Centre for Computational Aerodynamics, who contributed tremendously to the buildup of practical experience of numerical schemes for various applied aerodynamic problems.

References

- Van Leer B, Thomas J L, Roe P and Newsome R W (1987). Numerical flux formulas for the Euler and Navier-Stokes equations. AIAA Paper 87-1104.
- Van Leer B (1982). Flux vector splitting for the Euler equations. *Lect Notes in Phys*, **170**, p507.
- Osher S and Solomon F(1982). Upwind difference schemes for hyperbolic conservation laws. *Math Comp*, **38**, p339.
- Qin N and Zhu Y (1999). Grid adaption for shock/turbulent boundary layer interaction. *AIAA J*, **37**, p1129.
- Qin N and Ludlow D K (1999). A cure for anomalies of Osher and AUSM+ schemes for hypersonic viscous flows around swept cylinders. *22nd Inter Symp on Shock Waves*, Paper 1690.
- Quirk J J (1994). A contribution to the great Riemann solver debate. *Int J Numer Methods Fluids*, **18**, p555.
- Wada Y and Liou M-S (1994). A flux splitting scheme with high resolution and robustness for discontinuities. AIAA Paper 94-0083.
- Gressier J and Moschetta J-M (1998). On the pathological behaviour of upwind schemes. AIAA Paper 98-0110.
- Liou M-S (1995) A sequel to AUSM: AUSM+. *J Comp Phys*, **129**, p364.
- Liou M-S (1995). Progress towards an improved CFD method: AUSM+. AIAA Paper 95-1701.

A HIGH-RESOLUTION GODUNOV METHOD FOR MODELING ANOMALOUS FLUID BEHAVIOUR

WILLIAM J. RIDER AND JASON W. BATES

Hydrodynamic Methods Group (X-3)

Applied Physics Division

Los Alamos National Laboratory

Los Alamos, New Mexico 87545, U.S.A.

Email: wjr@lanl.gov, batesj@lanl.gov

Abstract. A standard assumption made when solving problems in compressible hydrodynamics is that the adiabatic compressibility of a fluid decreases with increasing pressure, or equivalently, that isentropes have a convex shape (downward) in the plane of specific volume versus pressure. This property is characteristic of all ideal gases and underlies classical hydrodynamic phenomenology. For real materials, however, isentropes may be locally concave near a phase transition. This can give rise to “anomalous” structures like smooth compressive waves and rarefactive shocks as the physical solutions. Here, we describe the construction of high-resolution Godunov schemes for modeling anomalous fluids. Particular attention is paid to the development of an appropriate Riemann solver to treat non-convex isentropes. Our approach is tested on a van der Waals gas with a distinct anomalous region.

1. Introduction

A standard assumption made when solving problems in compressible hydrodynamics is that isentropes are convex downward everywhere in the plane of specific volume V versus pressure p . Equivalently, this condition can be expressed by requiring the fundamental derivative $G \equiv \frac{1}{2}V^3c^{-2}(\partial^2 p/\partial V^2)_s = c^{-1}(\partial pc/\partial \rho)_s$, where c is the speed of sound and $\rho = 1/V$ is the density, to be strictly positive (Thompson, 1971). The derivatives here are calculated with the specific entropy s held fixed. Although it is not a thermodynamic requirement, the inequality $G > 0$ holds for ideal gases and is deeply en-

sconced in classical hydrodynamic phenomenology (*i.e.*, compressive shock waves and expansive rarefactions). It has long been known (Bethe, 1942; Zel'dovich and Raizer, 1966), however, that nonconvex regions where $G < 0$ can exist for fluids near a phase transition. In these cases, “anomalous” structures such as smooth compressive waves and rarefactive shocks can be the physical solutions (Thompson and Lambrakis, 1973). In this paper, we discuss the construction of high-resolution Godunov methods (Godunov, 1959; Toro, 1999) for modeling anomalous fluid properties.

A high-resolution Godunov algorithm is typically comprised of several steps: the reconstruction, time centering, a Riemann solver, and a conservative update (Toro, 1999; LeVeque, 1992). The Riemann solver, which is used to compute the updates of the hydrodynamic fluxes, most acutely feels the effects of a nonconvex isentrope. To account for anomalous structures in a Riemann solution, the standard assumptions regarding shocks, rarefactions and entropy-satisfying solutions must be modified (Menikoff and Plohr, 1989). This is accomplished by using the fundamental derivative G to aid in the construction of the Riemann solution. We shall describe a Riemann solver that explicitly accounts for the curvature of isentropes. We begin in Sect. 2 by considering a large-heat-capacity van der Waals gas with nonconvex isentropes in the (V, p) plane, which we use as a paradigm of anomalous fluid behaviour. In Sect. 3, we summarize the formulation of our Godunov scheme and the appropriate Riemann solver. In Sect. 4, we test our method on one-dimensional simulations of shock tubes filled with an anomalous van der Waals gas.

2. The Anomalous van der Waals Gas

The equation of state for a van der Waals gas is $(p + a/V^2)(V - b) = NkT$, where N is the number of molecules per unit mass, k is Boltzmann’s constant, T is the temperature and a and b are constants. Supplementing this equation is the expression for the entropy increment: $T ds = d\varepsilon + p dV = c_V dT + (p + a/V^2) dV$, where ε is the specific internal energy per unit mass. For simplicity, we assume that the specific heat per unit mass at constant volume, c_V , is equal to a constant. In dimensionless form, the pressure and specific internal energy are given by

$$\frac{p}{p_0} = \frac{8(T/T_0)}{3(V/V_0) - 1} - \frac{3}{(V/V_0)^2}, \quad (1)$$

$$\frac{\varepsilon}{\varepsilon_0} = \frac{\mu c_V}{R} (T/T_0) - \frac{9}{8(V/V_0)}, \quad (2)$$

where the normalization values $p_0 \equiv a/27b^2$, $V_0 \equiv 3b$, $T_0 \equiv 8\mu a/(27Rb)$, and $\varepsilon_0 = RT_0/\mu$ are chosen so that the critical point occurs at $p/p_0 = 1$

and $V/V_0 = 1$. Here, μ is the molecular weight of the gas and R is the universal gas constant. For a large value of the specific heat, van der Waals gases possess an anomalous region just to the right of the critical point in the (V, p) plane (Bates and Montgomery, 1999). An example is shown in Fig. 1 for the case $c_V = 80R/\mu$. Physically, such a large value of $\mu c_V/R$ corresponds to a gas with many internal molecular degrees of freedom.

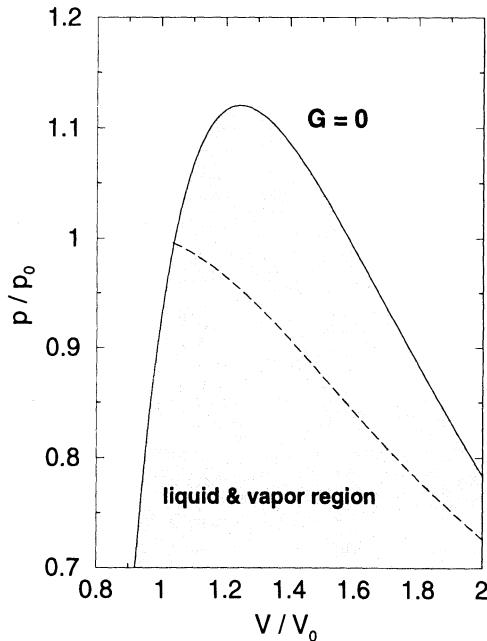


Fig. 1 Anomalous region (shaded) for a van der Waals gas with $\mu c_V/R = 80$. The curve labeled $G = 0$ is the locus of inflection points of the isentropes; below this curve, the parameter G is negative. The dotted line demarks the boundary between the purely gaseous regime and the two-phase co-existence region.

3. An Adaptive Riemann Solver

The equations we wish to solve are those of Eulerian hydrodynamics in a one-dimensional, fixed coordinate system. Working in terms of dimensionless variables and letting x denote the Cartesian coordinate and t the time, Euler's equations can be expressed in the vector form as

$$\mathbf{U}_t + \mathbf{F}_x = 0, \quad (3)$$

where $\mathbf{U} = (\rho, \rho v, E)^T$, $\mathbf{F} = [\rho v, \rho v^2 + p, v(E + p)]^T$, and T denotes the transpose. The symbol v represents the fluid velocity, and the total energy

density E is related to the specific internal energy according to $E = \rho e + \rho v^2/2$. To discretize our computational grid, we let $x_j = j\Delta x$ be a uniform mesh with $x_{j+1/2}$ the boundary between the j th and $j+1$ cells, and t^n be the time level. Here, Δx is the mesh spacing, Δt is the time step, and j is a positive integer.

Our Godunov method is based on a two-step Hancock scheme (Toro, 1999; Huynh, 1995) with the predictor stage formulated in terms of primitive variables $\mathbf{W} = (\rho, v, p)^T$. The solution procedure begins with a calculation of the gradient \mathbf{S} of \mathbf{W} in each cell:

$$\mathbf{S}_j^n = \frac{\mathbf{W}_{j+1}^n - \mathbf{W}_{j-1}^n}{2\Delta x}, \quad (4)$$

which is limited to ensure a locally monotonic profile. The next step is to interpolate time-centered values of the primitive variables to the cell edges. This is done by making a first-order Taylor expansion in space and time:

$$\mathbf{W}_{j+1/2,L}^{n+1/2} = \mathbf{W}_j^n + \frac{\Delta x}{2} \mathbf{S}_j^n + \frac{\Delta t}{2} \frac{\partial \mathbf{W}_j}{\partial t}, \quad (5)$$

$$\mathbf{W}_{j+1/2,R}^{n+1/2} = \mathbf{W}_{j+1}^n - \frac{\Delta x}{2} \mathbf{S}_{j+1}^n + \frac{\Delta t}{2} \frac{\partial \mathbf{W}_{j+1}}{\partial t}. \quad (6)$$

The time derivatives on the right hand sides of the above equations are estimated from the Euler equations expressed in terms of primitive variables. The subscripts L and R in Eqs. (5) and (6) denote “left” and “right” states, respectively. On either side of each cell interface, we now have time-centered expressions for the primitive variables. The resulting sequence of Riemann problems is solved to give

$$\mathbf{W}_{j+1/2}^{n+1/2} = \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} \mathbf{W}(x_{j+1/2}, t) dt = \mathcal{R}(\mathbf{W}_{j+1/2,L}^{n+1/2}, \mathbf{W}_{j+1/2,R}^{n+1/2}), \quad (7)$$

where \mathcal{R} stands for “Riemann solver.” Finally, these quantities are used to compute the fluxes of conserved variables and update the Euler equations from time level n to $n+1$ according to

$$\mathbf{U}_j^{n+1} = \mathbf{U}_j^n - \frac{\Delta t}{\Delta x} [\mathbf{F}(\mathbf{W}_{j+1/2}^{n+1/2}) - \mathbf{F}(\mathbf{W}_{j-1/2}^{n+1/2})]. \quad (8)$$

We now turn to a description of the Riemann solver that we have developed to simulate the motion of anomalous fluids.

The basis of the Riemann solver employed here is the adaptive two-shock method (Rider, 1999). Its foundation is the second-order continuity of wave curves at their centering point. We use a quadratic curve defined by

$$p_* = p_o + \rho_o \left(c_o + \frac{1}{2} G U_p \right) U_p, \quad (9)$$

where the pre-wave and post-wave states bear the subscripts “ o ” and “ $*$,” respectively. The symbol U_p represents the particle velocity defined as $v_* - v_o$ for the right wave, and $v_o - v_*$ for the left wave.

Many aspects of the adaptive two-shock solver remain unchanged from the version discussed by Rider (1999). Some of the salient features are: (i) for weak waves, a quadratic approximation to the wave curves is solved exactly; (ii) for stronger waves, an iterative approach is used for the solution with an adaptive choice for the quadratic term in the wave curves; (iii) the adaptive quadratic terms are defined to recover the correct asymptotic behaviour of the wave curves; (iv) for shocks, the Rankine-Hugoniot relations (Courant and Friedrichs, 1948) are used to define post-shock conditions; and (v) continuous smooth waves are reconstructed based upon isentropic relations. This Riemann solver was used to solve several problems involving anomalous van der Waals gases. Relatively few changes are needed to deal with the nonconvex portions of isentropes. One change deals with the selection of whether a wave is a shock or adiabatic. This can be effectively accomplished through testing the sign of the product GU_p ; if this is positive, the wave is a shock, otherwise it is a continuous isentropic transition. The other change in the method is the assumed asymptotic behaviour of the quadratic terms in the wave curves as waves become strong. Because all isentropes are asymptotically convex, we reproduce behaviour adaptively and as waves become stronger, the quadratic term becomes positive. One can reasonably estimate that particle Mach numbers of the nonconvex portion do not greatly exceed one.

4. Results

We now test our Godunov scheme by performing one-dimensional simulations of shock tubes (Zel'dovich and Raizer, 1966). At time $t = 0$, a static initial discontinuity in the middle of the tube separates a region on the left of higher pressure and density from a region of lower pressure and density on the right. For $t > 0$, the discontinuity is permitted to evolve in time. The result of this sort of simulation for an ideal gas is familiar and shall not be presented here.

Figure 2 shows a simulation of a shock tube filled with an anomalous van der Waals gas. In an ideal gas, a shock would travel to the right compressing the lower density gas there, while a rarefactive wave would spread to the left. Here, the opposite is true; a non-steepening wave compresses the lower density gas on the right, while a *rarefactive shock* propagates into the higher density gas on the left. The fundamental derivative G in this case is everywhere negative.

A second run was performed in the case that G changes sign. This is

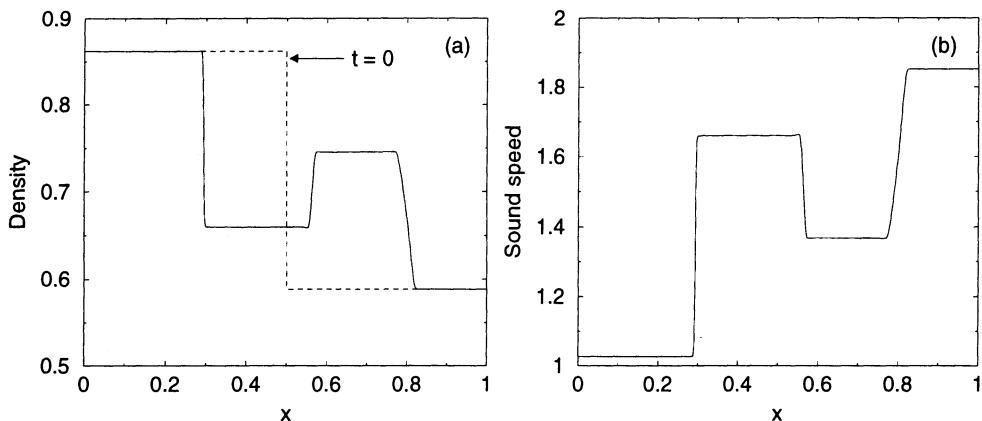


Fig. 2 Evolution of an initially static discontinuity (shock tube problem) in an anomalous fluid with $G < 0$. Profiles of (a) density, and (b) the sound speed are shown at $t = 0.4$. A rarefactive shock wave propagates leftward at $x \sim 0.3$ while a smooth compressive wave moves to the right at $x \sim 0.8$; a contact discontinuity lies between.

shown in Figure 3, where a complicated composite wave structure develops. From left to right, a rarefactive wave ($0.03 < x < 0.4$), a rarefactive shock ($x \sim 0.4$), a contact discontinuity ($x \sim 0.68$), a compressive shock ($x \sim 0.84$), a compressive wave ($0.84 < x < 0.88$), and another compressive shock ($x \sim 0.88$) can be seen. This is a case in which the left and right states lie outside of the anomalous region.

In conclusion, we have developed a high-order Godunov scheme and an adaptive Riemann solver to simulate the flow of anomalous fluids. A central feature of this formulation is the computation of the fundamental derivative G to estimate wave speeds. This is quite useful because when the inequality $G < 0$ is satisfied, one knows to expect anomalous wave structures such as rarefactive shocks and isentropic compressive waves in the physical solution. Another advantage of our solver over more conventional schemes such as Roe's method (Roe, 1981) is its adaptive formulation. This feature ameliorates pathological behaviours that are common to high-resolution methods (Quirk, 1994), without having to resort to excessive dissipation. Our solver also captures contact discontinuities well, and is robust for strong shock waves. Furthermore, the method allows for a more accurate estimate of wave speeds in multi-shock problems, which is often a crucial issue for time step control.

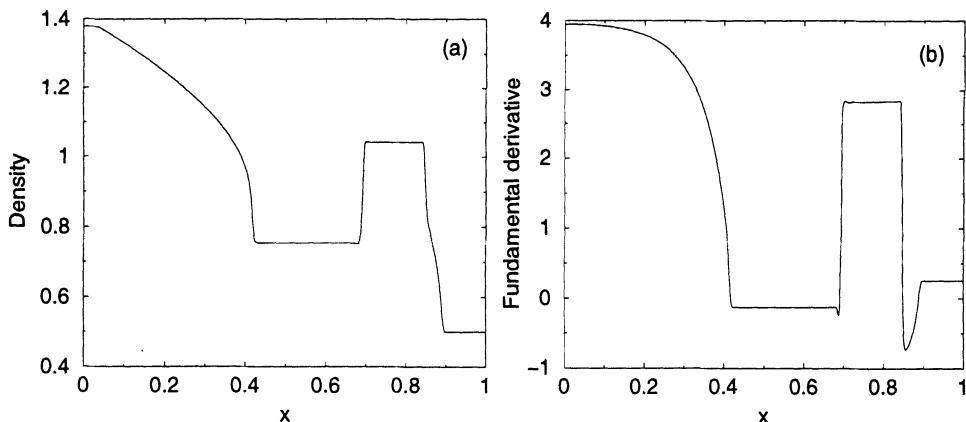


Fig. 3 Evolution of an initially static discontinuity in an anomalous fluid when the fundamental derivative changes sign. Profiles of (a) density, and (b) the fundamental derivative are shown at $t = 0.4$. In (a), a complicated composite wave structure is visible.

References

- Bates J W and Montgomery D C (1999). Some Numerical Studies of Exotic Shock Wave Behavior. *Phys. Fluids* **11**, pp 462-475.
- Bethe H A (1942). The Theory of Shock Waves for an Arbitrary Equation of State. Clearinghouse for Federal Scientific and Technical Information, US Department of Commerce, Washington D.C. Report No PB-32189.
- Courant R and Friedrichs K O (1948). Supersonic Flow and Shock Waves. Interscience.
- Godunov S K (1959). A Finite Difference Method for the Computation of Discontinuous Solutions of the Equations of Fluid Dynamics. *Mat. Sb.* **47**, pp 357-393.
- Huynh H T (1995). Accurate Upwind Methods for the Euler Equations. *SIAM J. Numer. Anal.* **32**, pp 1565-1619.
- LeVeque R J (1992). Numerical Methods for Conservation Laws. Birkhäuser Verlag.
- Menikoff R and Plohr B J (1989). The Riemann Problem for Fluid Flow of Real Materials. *Rev. Mod. Phys.* **61**, pp 75-130.
- Quirk J J (1994). A Contribution to the Great Riemann Solver Debate. *Inter. J. Num. Meth. Fluids* **18**, pp 555-574.
- Rider W J (1999). An Adaptive Riemann Solver using a Two-Shock Approximation. *Comp. Fluids* **28**, pp 741-777.
- Roe P L (1981). Approximate Riemann Solvers, Parameter Vectors, and Difference Schemes. *J. Comp. Phys.* **43**, pp 357-372.
- Thompson P A (1971). A Fundamental Derivative of Gas Dynamics. *Phys. Fluids* **14**, pp 1843-1849.
- Thompson P A and Lambrakis K C (1973). Negative Shock Waves. *J. Fluid Mech.* **60**, pp 187-208.
- Toro E F (1999). Riemann Solvers and Numerical Methods for Fluid Dynamics, Second Edition. Springer-Verlag.
- Zel'dovich Y B and Raizer Y P (1966). Shock Waves and High-Temperature Hydrodynamic Phenomena. Academic.

TOWARD GODUNOV-TYPE METHODS FOR HYPERBOLIC CONSERVATION LAWS WITH STIFF RELAXATION

PHILIP L. ROE AND JEFFREY A. F. HITTINGER

W. M. Keck Foundation

Laboratory for Computational Fluid Dynamics

Department of Aerospace Engineering

The University of Michigan

philroe@engin.umich.edu, jhitt@engin.umich.edu,

Abstract. We explore the issues involved in creating a Godonov-type method for flows with stiff source terms. We briefly survey the theory of such flows with particular emphasis on different asymptotic regimes and the solution of the Riemann problem. We then discuss practical issues, such as the choice of working variables, with particular reference to a set of eleven equations describing a non-equilibrium diatomic gas. We propose a uniformly-valid discretisation technique incorporating an approximate Riemann solver, but the details of the solver remain to be determined.

1. Introduction

We are concerned here with fluid flows that in some sense depart from an equilibrium state. For example, in a liquid containing bubbles or a gas containing dust particles, there are two different velocities involved. If we try to treat each phase as a continuum, these velocities coexist at every point. In general they will not be the same, but the existence of a drag force tends to bring them into equality. Similarly, if the molecules of a gas do not collide often enough to maintain thermodynamic equilibrium, then it is no longer possible to define scalar quantities called pressure or temperature. If the gas is not too ‘dilute’, these quantities can be redefined as tensors, and the collisions can be thought of as driving these tensors toward multiples of the unit tensor.

We are also motivated by the fact that there is an equivalence, for long waves, between hyperbolic-relaxation systems and hyperbolic-diffusion sys-

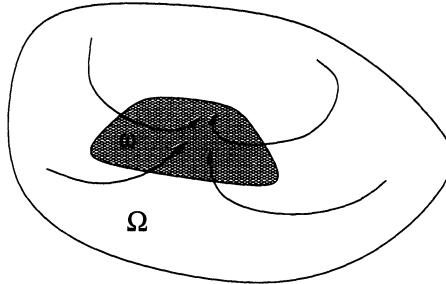


Figure 1. Nonequilibrium initial data are driven onto the equilibrium manifold.

tems such as the Navier-Stokes equations. Although relaxation systems are less explored from a computational viewpoint, they have potential advantages. One can take larger explicit time-steps, there is no global coupling to complicate parallelization, and the absence of any need to evaluate second derivatives should put less onus on generating smooth grids.

The general form of the fluid flow equations with relaxation is

$$\partial_t \mathbf{u} + \operatorname{div} \mathbf{f}(\mathbf{u}) + \mathbf{A}_i \partial_{x_i} \mathbf{u} = \mathbf{s}(\mathbf{u}). \quad (1)$$

By comparison with the more usual situation, complications arise here from two causes. One is that the spatial derivatives may not be entirely in divergence form (2nd term) but may also appear in quasilinear form (3rd term). This may happen either because there are effects that are inherently nonconservative (the added mass terms in bubbly liquids), or else because we choose not to use conserved variables as unknowns, even though we could (we consider this issue later). The second cause of complications is the *source term* $\mathbf{s}(\mathbf{u})$, representing drag forces, or collisions, or whatever phenomena bring the situation into equilibrium.

The set of ordinary differential equations

$$\partial_t \mathbf{u} = \mathbf{s}(\mathbf{u}) \quad (2)$$

describes what happens to spatially uniform initial data. We have a relaxation system if all the nonzero eigenvalues of this problem are negative. Then an arbitrary initial condition somewhere within the state space Ω will eventually tend to a state for which $\mathbf{s}(\mathbf{u}) = 0$, that is to say, an *equilibrium state*. The set ω of all equilibrium states is called the *equilibrium manifold*, and every initial state will be drawn onto it, as in Figure 1, unless prevented by events in the flow. An example from aeronautics is that of a re-entering space vehicle (Figure 2). At high altitudes, the bow shock produces a non-equilibrium flow which returns to equilibrium over the length of the plane. However, encountering a second shock wave, caused perhaps by a deflected control surface, could postpone the attainment of equilibrium.

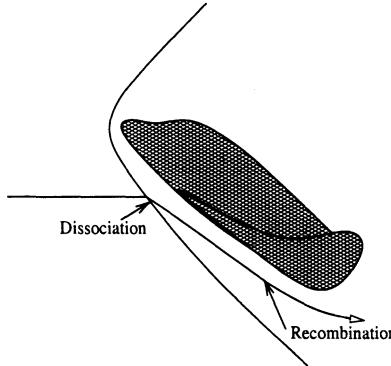


Figure 2. High temperatures behind the shockwave of a re-entry vehicle cause dissociation of ions and electrons, which recombine in the subsequent flow.

The numerical challenge is greatest when the source term is stiff. Stiffness, here, is defined to be the situation in which pure relaxation (2) would take place on time scales much smaller than the residence time of an acoustic wave inside a computational cell. The latter is the time step that we could take in the absence of source terms, but if we still take such a time step we will miss all the details of the relaxation process. We then have two options. The first is to reduce the time step until the relaxation is in fact resolved. This is unavoidable if the detailed process is of interest, but we have in mind situations where it may not be. For example, the relaxation process may be significant only in certain parts of the flow. In that case, we would like to use an adaptive strategy that permits large time steps elsewhere, secure in knowing that omitting the detail has had no untoward effect. Also, using the relaxation problem as a surrogate for a dissipative problem is only attractive if we do not have to resolve the small time scales, but only those longer scales on which the two problems are equivalent. For this purpose, we may never need to resolve the rapid transients.

It is not clear how to apply Godunov-type methods to relaxing flows, because the solution to the Riemann problem is no longer self-similar and is, in fact, extremely complicated, even for linear problems. For instance, the wave speeds are generally quite different at early and late times. This paper will, nevertheless, explore the possibilities.

2. Linear Analysis

2.1. EQUILIBRIUM AND RELAXATION VARIABLES

Consider the linear problem in one space dimension,

$$\partial_t \mathbf{u} + \mathbf{A} \partial_x \mathbf{u} = \mathbf{Q} \mathbf{u}, \quad (3)$$

where \mathbf{u} is a vector of dimension m and \mathbf{A}, \mathbf{Q} are constant $m \times m$ matrices. The matrix \mathbf{A} is of full rank and diagonalisable with real eigenvalues; \mathbf{Q} is of rank $m - n$, ($n > 0$), with negative semidefinite eigenvalues. The source term vanishes whenever \mathbf{u} lies in the nullspace of \mathbf{Q} , which is thereby identified with the equilibrium manifold and which is here a linear subspace of dimension n .

We take all of the variables in (3) to be dimensionless with respect to internal scales. Specifically, we consider times relative to a characteristic relaxation time; in the dimensional case, the eigenvalues of \mathbf{Q} , with dimensions of inverse time, characterize the time it takes to drive uniform initial data to equilibrium. The eigenvalues of \mathbf{A} would have the dimensions of speed, and so a characteristic length scale can be defined as the product of a characteristic velocity from \mathbf{A} and a characteristic relaxation time from \mathbf{Q} . With these loosely-defined dimensionless variables, we present here an informal treatment of the early- and late-time behaviour; the conclusions can be verified by formal asymptotics.

For initial data having high wavenumbers, that is, small wavelengths relative to the characteristic relaxation length scale, the differentiated term dominates the undifferentiated term. The balance therefore must be that

$$\partial_t \mathbf{u} + \mathbf{A} \partial_x \mathbf{u} \approx 0. \quad (4)$$

This is the *frozen limit* in which changes act so rapidly relative to the characteristic relaxation time that, to leading order, the relaxation is irrelevant. In this limit, the wavespeeds are the eigenvalues of \mathbf{A} , and these are the speeds with which high-wavenumber modes propagate. However, the high-wavenumber modes do not persist for very long; a dispersion analysis shows that these modes decay rapidly (on the relaxation time scale).

On the other hand, consider initial data very close to equilibrium; the solution will remain so if it is gently perturbed by long waves. So we can assume that \mathbf{u} remains close to the nullspace of \mathbf{Q} , which is spanned by the right nullvectors of \mathbf{Q} and contained in the $n \times m$ matrix \mathbf{R}_0 . It is illuminating to rewrite the system (3) to highlight this feature.

Consider the diagonalisation $\mathbf{Q} = \mathbf{R}\Lambda\mathbf{L}$ where \mathbf{R} and \mathbf{L} are the matrices of right and left eigenvectors of \mathbf{Q} , respectively, and Λ is the diagonal matrix of the eigenvalues of \mathbf{Q} . Partition \mathbf{R} and \mathbf{L} as follows, so that the nullvectors in each are blocked thus:

$$\mathbf{L} = \begin{pmatrix} \mathbf{L}_0 \\ \mathbf{L}_1 \end{pmatrix}_{m \times m} \quad \text{and} \quad \mathbf{R} = \begin{pmatrix} \mathbf{R}_0 & \mathbf{R}_1 \end{pmatrix}_{n \times m}. \quad (5)$$

The normalization $\mathbf{L} = \mathbf{R}^{-1}$ implies that

$$\mathbf{L}_0 \mathbf{R}_0 = \mathbf{I}_n, \quad \mathbf{L}_1 \mathbf{R}_1 = \mathbf{I}_{m-n}, \quad \text{and} \quad \mathbf{L}_0 \mathbf{R}_1 = \mathbf{L}_1 \mathbf{R}_0 = 0. \quad (6)$$

Now set $\mathbf{u} = \mathbf{R}_0 \mathbf{u}_e + \mathbf{R}_1 \mathbf{u}_r$. The significance of this decomposition is that \mathbf{u}_e represents the n variables that do not vanish in equilibrium, whereas \mathbf{u}_r represents the remainder that do. Now we rewrite (3) as

$$\partial_t(\mathbf{R}_0 \mathbf{u}_e + \mathbf{R}_1 \mathbf{u}_r) + \mathbf{A} \partial_x(\mathbf{R}_0 \mathbf{u}_e + \mathbf{R}_1 \mathbf{u}_r) = \mathbf{Q}(\mathbf{R}_0 \mathbf{u}_e + \mathbf{R}_1 \mathbf{u}_r). \quad (7)$$

Multiplying by \mathbf{L}_0 and \mathbf{L}_1 , respectively, we diagonalise \mathbf{Q} :

$$\partial_t \mathbf{u}_e + \mathbf{L}_0 \mathbf{A} \mathbf{R}_0 \partial_x \mathbf{u}_e + \mathbf{L}_0 \mathbf{A} \mathbf{R}_1 \partial_x \mathbf{u}_r = 0, \quad (8a)$$

$$\partial_t \mathbf{u}_r + \mathbf{L}_1 \mathbf{A} \mathbf{R}_0 \partial_x \mathbf{u}_e + \mathbf{L}_1 \mathbf{A} \mathbf{R}_1 \partial_x \mathbf{u}_r = \mathbf{L}_1 \mathbf{Q} \mathbf{R}_1 \mathbf{u}_r = \Lambda_1 \mathbf{u}_r, \quad (8b)$$

where Λ_1 contains the $m - n$ negative definite eigenvalues of \mathbf{Q} .

For a flow that is completely in equilibrium, we can set $\mathbf{u}_r \equiv 0$ and then we obtain the *equilibrium limit*

$$\partial_t \mathbf{u}_e + (\mathbf{L}_0 \mathbf{A} \mathbf{R}_0) \partial_x \mathbf{u}_e = 0, \quad (9)$$

showing that the properties of equilibrium advection depend on the $n \times n$ matrix $\mathbf{L}_0 \mathbf{A} \mathbf{R}_0$. In general, the eigenvalues of this Jacobian matrix will be different from those of the original Jacobian \mathbf{A} , and this demonstrates one way in which the advection and relaxation couple.

To the next approximation, we neglect, for wavelengths large compared to the relaxation length scale, the derivatives of \mathbf{u}_r compared with \mathbf{u}_r itself and obtain from (8b)

$$\Lambda_1 \mathbf{u}_r \approx (\mathbf{L}_1 \mathbf{A} \mathbf{R}_0) \partial_x \mathbf{u}_e. \quad (10)$$

Substituting this into (8a) gives the *near-equilibrium limit*

$$\partial_t \mathbf{u}_e + (\mathbf{L}_0 \mathbf{A} \mathbf{R}_0) \partial_x \mathbf{u}_e \approx -(\mathbf{L}_0 \mathbf{A} \mathbf{R}_1) \Lambda_1^{-1} (\mathbf{L}_1 \mathbf{A} \mathbf{R}_0) \partial_{xx} \mathbf{u}_e. \quad (11)$$

Close to equilibrium, therefore, the equilibrium variables are governed by advection-diffusion equations. The approach to equilibrium is stable if the diffusion coefficient matrix $-(\mathbf{L}_0 \mathbf{A} \mathbf{R}_1) \Lambda_1^{-1} (\mathbf{L}_1 \mathbf{A} \mathbf{R}_0)$, appearing on the right of (11), is positive definite. Conditions on \mathbf{A} and \mathbf{Q} sufficient to guarantee this are at present only partly established (Zeng, 1999).

2.2. THE LINEAR RIEMANN PROBLEM

The structure of the solution to the Riemann problem has only recently become clear, even for the linear case (Zeng, 1999), and certain details are still not known. However, some general results can be found.

Consider Riemann initial data,

$$\mathbf{u}(x < 0, 0) = \mathbf{u}_L \quad \text{and} \quad \mathbf{u}(x \geq 0, 0) = \mathbf{u}_R. \quad (12)$$

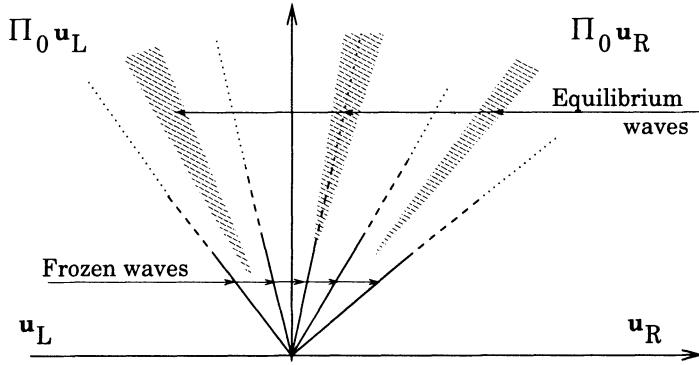


Figure 3. Structure of the solution to a Riemann problem with relaxation.

Since the homogeneous Riemann problem does not introduce a length scale, the solution will be self-similar in the variable $\xi = x/t$. The solution to the nonhomogeneous problem will be self-similar in both the frozen and equilibrium limits (Figure 3). It therefore makes sense to take $\mathbf{u} = \mathbf{u}(\xi, t)$, and to consider the problem in the form

$$(\mathbf{A} - \xi \mathbf{I}) \partial_\xi \mathbf{u} + t(\partial_t \mathbf{u} - \mathbf{Q} \mathbf{u}) = 0. \quad (13)$$

2.2.1. Early Time Behaviour

For $t \rightarrow 0$, we take as our *ansatz* the power series coordinate expansion (useful also in the nonlinear case (Le Floch and Raviart, 1987; Bourgeade, Le Floch and Raviart, 1989))

$$\mathbf{u}(x, t) = \mathbf{u}_0(\xi) + t\mathbf{u}_1(\xi) + \frac{t^2}{2}\mathbf{u}_2(\xi) + \dots \quad (14)$$

This succeeds if

$$\begin{aligned} & (\mathbf{A} - \xi \mathbf{I}) \partial_\xi \mathbf{u}_0 + t[(\mathbf{A} - \xi \mathbf{I}) \partial_\xi \mathbf{u}_1 + \mathbf{u}_1 - \mathbf{Q} \mathbf{u}_0)] \\ & + \frac{t^2}{2}[(\mathbf{A} - \xi \mathbf{I}) \partial_\xi \mathbf{u}_2 + 2\mathbf{u}_2 - 2\mathbf{Q} \mathbf{u}_1] + \dots = 0, \end{aligned} \quad (15)$$

provided that the \mathbf{u}_j are piecewise continuous with at most m jumps.¹ Each order must independently vanish; enforcing this gives a recurrence relationship for the \mathbf{u}_i ,

$$(\mathbf{A} - \xi \mathbf{I}) \partial_\xi \mathbf{u}_j + j\mathbf{u}_j = j\mathbf{Q} \mathbf{u}_{j-1}, \quad (j = 0, 1, 2, \dots), \quad (16)$$

where we define $\mathbf{u}_{-1} = 0$.

¹In fact, it can be shown that each \mathbf{u}_j is piecewise polynomial with degree j .

Clearly \mathbf{u}_0 is the regular Riemann solution for the frozen limit (4). So waves initially propagate as discontinuities with speeds $\{\lambda_k^A; k = 1, \dots, m\}$, the eigenvalues of \mathbf{A} . These are, in fact, the only speeds with which abrupt discontinuities can ever travel; in effect, at a discontinuity, the source term can be neglected.² The jump across the k th discontinuity must be of the form

$$\alpha_k(t) \mathbf{r}_k^A = \left(\sum_{j=0}^{\infty} \frac{\alpha_{k,j} t^j}{j!} \right) \mathbf{r}_k^A, \quad (k = 1, \dots, m), \quad (17)$$

where \mathbf{r}_k^A is the k th right eigenvector of \mathbf{A} .

Now multiply (16) by ℓ_k^A , the k th left eigenvector of \mathbf{A} :

$$(\lambda_k^A - \xi) \partial_\xi (\ell_k^A \cdot \mathbf{u}_j) + j(\ell_k^A \cdot \mathbf{u}_j) = j \ell_k^A \mathbf{Q} \mathbf{u}_{j-1}. \quad (18)$$

Evaluating this expression at $\xi = \lambda_k^A \pm \epsilon$ for $\epsilon \rightarrow 0$, we obtain

$$\alpha_{k,j} = (\ell_k^A \mathbf{Q} \mathbf{r}_k^A) \alpha_{k,j-1}. \quad (19)$$

Solving this recurrence relation and substituting it back into (17), we find, remarkably, that the series solution provides the discontinuity strengths for all time:

$$\alpha_k(t) = \exp(\ell_k^A \mathbf{Q} \mathbf{r}_k^A t) \alpha_k(0), \quad (k = 1, \dots, m). \quad (20)$$

So the initial discontinuities decay exponentially, unless the eigenvector \mathbf{r}_k^A happens to lie in the nullspace of \mathbf{Q} . For such cases the wave does not decay and $\alpha_k(t) \equiv \alpha_k(0)$. In the examples shown later, the factor $\ell_k^A \mathbf{Q} \mathbf{r}_k^A$ turns out to be in the range $[0.01, 0.5]$ and the nonlinear computations in Section 5 broadly confirm the linear estimates (20).

However, at large times, we expect that the solution will obey the equilibrium equation (9) which has its own set of jump conditions, the eigenvectors of $\mathbf{L}_0 \mathbf{A} \mathbf{R}_0$. Where then do these equilibrium discontinuities come from? This is the part that is poorly understood, so we will go straight to the large-time limit.

2.2.2. Late Time Behaviour

It is clear that outside the domain of influence of the origin, the two initially constant states \mathbf{u}_L and \mathbf{u}_R will decay like

$$\mathbf{u}_L(t) = \exp(\mathbf{Q}t) \mathbf{u}_L(0) \quad \text{and} \quad \mathbf{u}_R(t) = \exp(\mathbf{Q}t) \mathbf{u}_R(0). \quad (21)$$

²Later we will see how, in the near-equilibrium limit, ‘discontinuities’ arise that appear to contradict this statement.

It is easy to show that

$$\lim_{t \rightarrow \infty} \exp \mathbf{Q}t = \mathbf{\Pi}_0(\mathbf{Q}), \quad (22)$$

which is the projector into the nullspace of \mathbf{Q} . Therefore, at sufficiently large time, we have a Riemann problem for the matrix $\mathbf{L}_0 \mathbf{A} \mathbf{R}_0$ with the specific data $\mathbf{\Pi}_0 \mathbf{u}_L$ and $\mathbf{\Pi}_0 \mathbf{u}_R$. In the coordinate $\xi = x/t$, we expect to see n equilibrium discontinuities whose speeds will be the eigenvalues of $\mathbf{L}_0 \mathbf{A} \mathbf{R}_0$ rather than those of \mathbf{A} . However, only long waves survive into the near equilibrium limit, so we expect that these ‘discontinuities’ will actually have smooth, continuous transitions.

To expedite asymptotic analysis of this limit, it is useful to introduce a second length scale. While artificial, this new equilibrium length scale could be related, for instance, to characteristic lengths of the long waves remaining in the near-equilibrium solution at a particular large value of time. The introduction of such a length scale also introduces a small parameter, effectively the inverse of a Reynolds number, in which the solution can be expanded. Justification for this process comes from the fact that the leading-order asymptotic solution is independent of the both the small parameter and the equilibrium length scale.

Defining \hat{x} and \hat{t} to be length and time variables made dimensionless with this new equilibrium length scale and the associated time scale, the general system (8) in equilibrium variables becomes

$$\partial_{\hat{t}} \mathbf{u}_e + \mathbf{L}_0 \mathbf{A} \mathbf{R}_0 \partial_{\hat{x}} \mathbf{u}_e + \mathbf{L}_0 \mathbf{A} \mathbf{R}_1 \partial_{\hat{x}} \mathbf{u}_r = 0, \quad (23a)$$

$$\partial_{\hat{t}} \mathbf{u}_r + \mathbf{L}_1 \mathbf{A} \mathbf{R}_0 \partial_{\hat{x}} \mathbf{u}_e + \mathbf{L}_1 \mathbf{A} \mathbf{R}_1 \partial_{\hat{x}} \mathbf{u}_r = \frac{1}{\epsilon} \mathbf{\Lambda}_1 \mathbf{u}_r, \quad (23b)$$

where the parameter $\epsilon \ll 1$ is the ratio of the characteristic relaxation time to the representative equilibrium time.

We now take as an *ansatz* the power series parameter expansion

$$\begin{pmatrix} \mathbf{u}_e \\ \mathbf{u}_r \end{pmatrix} = \sum_{j=0}^{\infty} \epsilon^j \begin{pmatrix} \phi_j(\hat{x}, \hat{t}) \\ \psi_j(\hat{x}, \hat{t}) \end{pmatrix} \quad (24)$$

for $\epsilon \rightarrow 0$. Substituting this into (23) and equating terms of like order, we find the constraints

$$(\partial_{\hat{t}} + \mathbf{L}_0 \mathbf{A} \mathbf{R}_0 \partial_{\hat{x}}) \phi_j = -\mathbf{L}_0 \mathbf{A} \mathbf{R}_1 \partial_{\hat{x}} \psi_j, \quad (25a)$$

$$\mathbf{\Lambda}_1 \psi_j = \mathbf{L}_1 \mathbf{A} \mathbf{R}_0 \partial_{\hat{x}} \phi_{j-1} + (\partial_{\hat{t}} + \mathbf{L}_1 \mathbf{A} \mathbf{R}_1 \partial_{\hat{x}}) \psi_{j-1}, \quad (25b)$$

for $j = 0, 1, 2, \dots$ and where we define $\phi_{-1} = \psi_{-1} = 0$. This implies that $\psi_0 \equiv 0$, which is expected near equilibrium. Using this fact in (25b), we

find that, to $O(\epsilon^2)$, the system (23) reduces to

$$(\partial_t + \mathbf{L}_0 \mathbf{A} \mathbf{R}_0 \partial_{\hat{x}}) \mathbf{u}_e \approx -\epsilon \mathbf{L}_0 \mathbf{A} \mathbf{R}_1 \Lambda_1^{-1} \mathbf{L}_1 \mathbf{A} \mathbf{R}_0 \partial_{\hat{x}\hat{x}} \mathbf{u}_e, \quad (26a)$$

$$\Lambda_1 \mathbf{u}_r \approx \mathbf{L}_1 \mathbf{A} \mathbf{R}_0 \partial_{\hat{x}} \mathbf{u}_e. \quad (26b)$$

Introducing the equilibrium characteristic variables $\mathbf{c}_e = \mathbf{L}_e \mathbf{u}_e$, where \mathbf{L}_e is the $n \times n$ matrix of left eigenvectors of $\mathbf{L}_0 \mathbf{A} \mathbf{R}_0$, the left side of (26a) can be diagonalised,

$$(\partial_t + \Lambda_e \partial_{\hat{x}} + -\epsilon \mathbf{L}_e \mathbf{L}_0 \mathbf{A} \mathbf{R}_1 \Lambda_1^{-1} \mathbf{L}_1 \mathbf{A} \mathbf{R}_0 \mathbf{R}_e \partial_{\hat{x}\hat{x}}) \mathbf{c}_e \approx 0, \quad (27)$$

where $\mathbf{R}_e = \mathbf{L}_e^{-1}$ and Λ_e is the diagonal matrix of eigenvalues of $\mathbf{L}_0 \mathbf{A} \mathbf{R}_0$.

In general, the coefficient matrix multiplying the diffusion term will *not* be diagonal. However, to leading order, each characteristic variable $c_{e,k}$ is piecewise constant with a single jump at the k th wave ($k = 1, \dots, n$). Thus, for Riemann initial data and at long times, the equations (27) decouple, and we are left with n advection-diffusion equations³

$$(\partial_t + \lambda_{e,k} \partial_{\hat{x}} - \epsilon \nu_k \partial_{\hat{x}\hat{x}}) c_{e,k} \approx 0, \quad k = 1, \dots, n, \quad (28)$$

where ν_k is the k th diagonal element of $\{\mathbf{L}_e \mathbf{L}_0 \mathbf{A} \mathbf{R}_1 \Lambda_1^{-1} \mathbf{L}_1 \mathbf{A} \mathbf{R}_0 \mathbf{R}_e\}$. The solutions for $\epsilon \rightarrow 0$ are error functions propagating with the equilibrium wavespeeds

$$c_{e,k}(x, t) \approx \beta_{e,k} + \alpha_{e,k} \left(1 + \operatorname{erf} \left\{ \sqrt{\frac{t}{4\nu_k}} (\xi - \lambda_{e,k}) \right\} \right) \quad (29)$$

where $\beta_{e,k}$ is the value to the left of the k th wave and $\alpha_{e,k}$ is the jump across the k th wave. These are the waves that dominate the equilibrium solution. Because their widths are proportional to \sqrt{t} , each assumes, for large enough t , the appearance of a discontinuity in the similarity coordinate x/t .

The corresponding solutions for the relaxation variables from (26b) are therefore decaying Gaussians also propagating with the equilibrium wave speeds:

$$\mathbf{u}_r(x, t) \approx -\Lambda_1^{-1} \mathbf{L}_1 \mathbf{A} \mathbf{R}_0 \sum_{k=1}^n \frac{\alpha_{e,k}}{\sqrt{\pi \nu_k t}} \exp \left\{ -\frac{t(\xi - \lambda_{e,k})^2}{4\nu_k} \right\} \mathbf{r}_{e,k}. \quad (30)$$

It is worth remarking that the integral with respect to x of \mathbf{u}_r approaches a finite limit; the relaxation variables never disappear completely.

³As in (20), an exception occurs if some eigenvector of \mathbf{A} lies in the nullspace of \mathbf{Q} . By returning to (3), it is clear that a disturbance proportional to such a vector propagates for all time with no attenuation and no dissipation. One of the ν_k will then be zero.

The features of the solution to the Riemann problem are sketched in Figure 3. At early times, the initial discontinuity is carried along the frozen waves which decay exponentially. Somehow, the nonuniform flow that develops at small times between the decaying discontinuities ‘focuses’ itself into the smooth equilibrium waves. This *intermediate regime* is the period that is hardest to analyse, although numerical simulations (Section 5) seem to show that it happens rather smoothly. We are not deterred by our inability to solve the linear Riemann problem precisely. Instead, we are consoled by the thought that several viable Godunov-type schemes rely on ‘Riemann-solvers’ that are quite approximate, for example those of HLLE-type (Harten, Lax and van Leer, 1983).

3. The Physical Problem

Many relaxation systems arising in physics are quite large. We wanted to study a problem typical of realistic applications, but not so large that the algebra would be overwhelming. The analysis in Section 2 shows that much of the physics derives from interactions between the matrices \mathbf{A} and \mathbf{Q} ,⁴ and the matrix products that express this could be very complicated. However, physically-motivated problems often exhibit rather simple structures, and we hoped to be able to identify these. The system that we chose comprised eleven equations in eleven unknowns, and represents a nonequilibrium diatomic gas through the first ten translational velocity moments of its velocity distribution $\mathcal{G}(\mathbf{v}, \boldsymbol{\omega}; \mathbf{x}, t)$ and one rotational energy moment. Here \mathbf{v} and $\boldsymbol{\omega}$ are translational and rotational molecular velocities, respectively. The moments are defined as

$$\begin{aligned} m_0(\mathbf{x}, t) &= \mathcal{M} \int \mathcal{G}(\mathbf{v}, \boldsymbol{\omega}; \mathbf{x}, t) d\mathbf{v} d\boldsymbol{\omega}, \\ m_i(\mathbf{x}, t) &= \mathcal{M} \int \mathcal{G}(\mathbf{v}, \boldsymbol{\omega}; \mathbf{x}, t) v_i d\mathbf{v} d\boldsymbol{\omega}, \\ m_{ij}(\mathbf{x}, t) &= \mathcal{M} \int \mathcal{G}(\mathbf{v}, \boldsymbol{\omega}; \mathbf{x}, t) v_i v_j d\mathbf{v} d\boldsymbol{\omega} \\ m_{\omega^2}(\mathbf{x}, t) &= \mathcal{I} \int \mathcal{G}(\mathbf{v}, \boldsymbol{\omega}; \mathbf{x}, t) \omega^2 d\mathbf{v} d\boldsymbol{\omega} \end{aligned} \quad (31)$$

where \mathcal{M} is the molecular mass, \mathcal{I} is the moment of inertia about the two meaningful axes of rotation, and the integrals are over the whole of the five-dimensional velocity space. The zeroth moment m_0 can be identified with the fluid density ρ , the first moments m_i with the momentum in the

⁴Such interactions are ignored in simple operator-splitting methods. For a critique of operator-splitting in this context see (Pember, 1993). The wish to acknowledge these interactions dictates the rather elaborate splitting in (Cafisch, Jin and Russo, 1997).

i -direction, and the second moments m_{ij} with the flux of i -momentum in the j -direction. We define the bulk velocities $u_i = m_i/\rho$ and the random velocities $c_i = v_i - u_i$. Then we can define a pressure tensor

$$P_{ij} = \mathcal{M} \int \mathcal{G}(\mathbf{c}, \boldsymbol{\omega}; \mathbf{x}, t) c_i c_j d\mathbf{c} d\boldsymbol{\omega}. \quad (32)$$

Extending the treatment of monatomic gases in (Levermore, 1996), we assume that the distribution of random velocities has the Gaussian form

$$\mathcal{G}(\mathbf{c}, \boldsymbol{\omega}; \mathbf{x}, t) = \frac{\rho \mathcal{I}}{\mathcal{M}(2\pi)^{5/2} \Delta^{1/2} \mathfrak{K} T_{\text{rot}}} \exp \left(-\frac{1}{2} \left[\Theta_{ij}^{-1} c_i c_j + \frac{I\omega^2}{\mathfrak{K} T_{\text{rot}}} \right] \right) \quad (33)$$

where \mathfrak{K} is Boltzmann's constant; T_{rot} is a rotational temperature; and $\Delta = \det \boldsymbol{\Theta}$ where $\boldsymbol{\Theta}$ is a ‘temperature tensor’ defined by $\Theta_{ij} = P_{ij}/\rho$.

The pressure tensor allows us to represent anisotropic pressures (viscous stresses) but not heat transfer, which requires the diagonal elements of the third-order moments. We excluded these out of a wish to keep the system as small as possible whilst remaining physically significant.⁵ However, we also included an additional scalar quantity, the energy in the rotational modes, which physically introduces the possibility of a non-monatomic gas, and mathematically introduces a second relaxation time.

The equations can be written in conservation form for the moments but, as usual, non-conservative ‘primitive’ variables simplify the algebra. A compromise is a ‘semiconservative’ formulation, in which the unknowns are

$$\mathbf{u} = (\rho, \rho u_i, P_{ij}, E_{\text{rot}})^T \quad (34)$$

In these variables, defining $\Sigma_{ijk} = u_i \Theta_{jk} + u_j \Theta_{ik} + u_k \Theta_{ij}$, we have

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -u_1^2 & 2u_1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -u_1 u_2 & u_2 & u_1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ -u_1 u_3 & u_3 & 0 & u_1 & 0 & 0 & 1 & 0 & 0 & 0 \\ -\Sigma_{111} & 3\Theta_{11} & 0 & 0 & u_1 & 0 & 0 & 0 & 0 & 0 \\ -\Sigma_{112} & 2\Theta_{12} & \Theta_{11} & 0 & 0 & u_1 & 0 & 0 & 0 & 0 \\ -\Sigma_{113} & 2\Theta_{13} & 0 & \Theta_{11} & 0 & 0 & u_1 & 0 & 0 & 0 \\ -\Sigma_{122} & \Theta_{22} & 2\Theta_{12} & 0 & 0 & 0 & 0 & u_1 & 0 & 0 \\ -\Sigma_{123} & \Theta_{23} & \Theta_{13} & \Theta_{12} & 0 & 0 & 0 & 0 & u_1 & 0 \\ -\Sigma_{133} & \Theta_{33} & 0 & 2\Theta_{13} & 0 & 0 & 0 & 0 & 0 & u_1 \\ -u_1 E_{\text{rot}} & E_{\text{rot}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & u_1 \end{bmatrix}, \quad (35)$$

⁵We believe there are, in fact, situations such as aeolian sound generation where heat transfer can be neglected.

and, defining $\phi = 1/3\tau_t - 2/15\tau_r$,

$$\mathbf{Q} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \phi - \frac{1}{\tau_t} & 0 & 0 & \phi & 0 & \phi \frac{2}{5\tau_r} \\ 0 & 0 & 0 & 0 & 0 & -\frac{1}{\tau_t} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{\tau_t} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \phi & 0 & 0 & \phi - \frac{1}{\tau_t} & 0 & \phi \frac{2}{5\tau_r} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{\tau_t} & 0 \\ 0 & 0 & 0 & 0 & \phi & 0 & 0 & \phi & 0 & \phi \frac{2}{5\tau_r} \\ 0 & 0 & 0 & 0 & \frac{1}{5\tau_r} & 0 & 0 & \frac{1}{5\tau_r} & 0 & -\frac{1}{5\tau_r} \end{bmatrix}, \quad (36)$$

where τ_t and τ_r are, respectively, the relaxation times for the translational and rotational energies. A nice feature of using these variables is that \mathbf{Q} is ‘almost constant’. In fact, the relaxation times depend on the fluid state, and, to achieve a match with ‘long wave’ solutions of the Navier-Stokes equations, they must be chosen as

$$\tau_t = \frac{\mu}{p} \quad \text{and} \quad \tau_r = \frac{15}{4} \frac{\mu_B}{p}, \quad (37)$$

where μ and μ_B are, respectively, the coefficients of shear and bulk viscosity and where p is the thermodynamic pressure defined below (39).

Eigenanalysis of the Jacobian \mathbf{A} identifies five distinct frozen wavespeeds. The two fastest are genuinely nonlinear acoustic waves travelling with speeds $u_1 \pm \sqrt{3\Theta_{11}}$. The remaining nine waves are all linearly degenerate. There are two pairs of ‘pseudo-acoustic’ shear waves which travel with speeds $u_1 \pm \sqrt{\Theta_{11}}$. Five waves are convected in the x_1 direction with the bulk velocity u_1 ; they propagate entropy and deviations from thermodynamic equilibrium. In contrast, in the equilibrium (Euler) limit, there are of course five waves: two acoustic waves with speeds $u_1 \pm \sqrt{1.4p/\rho}$, two convected shear waves, and a convected entropy wave. Note that $\Theta_{11} \rightarrow p/\rho$ as the gas approaches equilibrium.

Now we ask if there could be a more revealing ‘canonical form’ for the equations. A good first move is to diagonalise $\mathbf{Q} = \mathbf{R}\Lambda\mathbf{L}$ with

$$\Lambda = \text{diag}(0, 0, 0, 0, 0, \tau_t^{-1}, \tau_t^{-1}, \tau_t^{-1}, \tau_t^{-1}, \tau_t^{-1}, \tau_r^{-1})^T. \quad (38)$$

Diagonalisation automatically reveals the nullspace, here spanned by the density, momentum, and thermodynamic pressure:

$$p = \frac{1}{5} \text{Tr } P_{ij} + \frac{2}{5} E_{\text{rot}}. \quad (39)$$

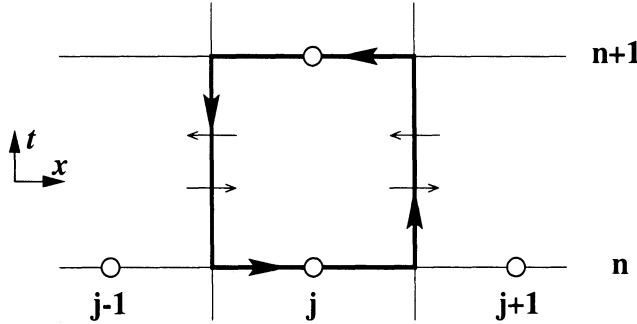


Figure 4. The path of integration around a computational cell.

The diagonalising variables are not unique because various bases can be chosen for the subspaces associated with each of the three eigenvalues $\{0, \tau_t^{-1}, \tau_r^{-1}\}$, but the choice of partly-conservative variables

$$\mathbf{u}^d = (\rho, \rho u_1, \rho u_2, \rho u_3, p, P_{12}, P_{13}, P_{23}, P_{11}-P_{22}, P_{22}-P_{33}, E_{\text{rot}}-p)^T \quad (40)$$

is reasonable. Useful simplicity comes from the variability of \mathbf{Q} only appearing in \mathbf{A} . Because of this, the matrices \mathbf{R} and \mathbf{L} are constant, and, hence, so are \mathbf{R}_0 , \mathbf{R}_1 , \mathbf{L}_0 , and \mathbf{L}_1 . This justifies the faith expressed earlier that a physically-motivated system would have at least some simple properties.

4. A Useful Transformation

We now return to the linear problem

$$\partial_t \mathbf{u} + \mathbf{A} \partial_x \mathbf{u} = \mathbf{Q} \mathbf{u}, \quad (41)$$

which we write under the change of variable $\mathbf{u} = \exp\{\mathbf{Q}t\}\mathbf{w}$ as

$$\partial_t \mathbf{w} + [\exp\{-\mathbf{Q}t\} \mathbf{A} \exp\{\mathbf{Q}t\}] \partial_x \mathbf{w} = 0. \quad (42)$$

Note that the origin for t is arbitrary. The equation now ‘looks homogeneous’, in that the source term has been eliminated, but the matrix is no longer a constant. Near to $t = 0$ we have the frozen limit, and for very large t the equilibrium limit. Integrate this equation along the path shown in Figure 4, denoting the average value of \mathbf{w} in $[x_{j-1/2}, x_{j+1/2}]$ by $\bar{\mathbf{w}}_j$ and writing $\Delta \mathbf{w}_j(t) = \mathbf{w}_{j+1/2}(t) - \mathbf{w}_{j-1/2}(t)$,

$$\bar{\mathbf{w}}_j^{n+1} - \bar{\mathbf{w}}_j^n = -\frac{1}{\Delta x} \int_{t^n}^{t^{n+1}} [\exp\{-\mathbf{Q}t\} \mathbf{A} \exp\{\mathbf{Q}t\}] \Delta \mathbf{w}_j(t) dt. \quad (43)$$

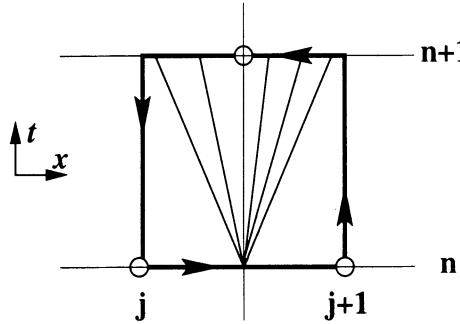


Figure 5. The path of integration around a computational cell on a staggered mesh.

Transforming back to the original variables, this becomes

$$\bar{\mathbf{u}}_j^{n+1} = \exp\{\mathbf{Q}\Delta t\}\bar{\mathbf{u}}_j^n - \frac{1}{\Delta x} \int_{t^n}^{t^{n+1}} \exp\{\mathbf{Q}(t^{n+1} - t)\} \mathbf{A} [\mathbf{u}_{j+1/2}(t) - \mathbf{u}_{j-1/2}(t)] dt. \quad (44)$$

This formula is the very familiar basis of regular finite-volume schemes for the case $\mathbf{Q} = 0$, where the matrix exponentials reduce to identity matrices. With \mathbf{Q} non-zero, it shows two effects. The cell average decays ‘internally’ according to the ODE and also changes due to the events on its boundaries. We have the ingredients for a Godunov-type solver if we have some reasonable approximation to those events.

4.0.3. A Staggered Mesh Scheme

Bereux and Sainsaulieu evade any detailed discussion of the Riemann problem by employing a staggered mesh (Figure 5). This gives the great simplification that the fluxes can be evaluated in a region where only the ordinary differential equation (2) holds, so that $\mathbf{u}(t) = \exp\{\mathbf{Q}(t - t^n)\}\mathbf{u}(t^n)$ and the first-order scheme becomes:

$$\bar{\mathbf{u}}_j^{n+1} = \frac{1}{2} \exp\{\mathbf{Q}\Delta t\} [\bar{\mathbf{u}}_{j-1}^n + \bar{\mathbf{u}}_{j+1}^n] - \frac{1}{\Delta x} \int_{t^n}^{t^{n+1}} \exp\{\mathbf{Q}(t^{n+1} - t)\} \mathbf{A} \exp\{\mathbf{Q}(t - t^n)\} \Delta \mathbf{u}_j dt, \quad (45)$$

where $\Delta \mathbf{u}_j = \mathbf{u}_{j+1/2} - \mathbf{u}_{j-1/2}$. If the time step is very small compared to the relaxation times, this is the Lax-Friedrichs scheme for the frozen problem, provided that \mathbf{A} is chosen such that

$$\tilde{\mathbf{A}}(\mathbf{u}_{j+1/2}^n - \mathbf{u}_{j-1/2}^n) = \mathbf{F}_{j+1/2}^n - \mathbf{F}_{j-1/2}^n. \quad (46)$$

In the opposite limit, the matrix in (45) becomes

$$\lim_{t \rightarrow \infty} (\exp\{-\mathbf{Q}t\} \mathbf{A} \exp\{\mathbf{Q}t\}), \quad (47)$$

and we can show that the first n rows and columns of this are given by $\mathbf{L}_0 \mathbf{A} \mathbf{R}_0$. Since we wish to obtain solutions to the equilibrium problem, we need to choose a linearisation of \mathbf{Q} such that this block represents the equilibrium Jacobian. In our problem, \mathbf{L}_0 and \mathbf{R}_0 are actually constants, so that this takes place automatically.

(Bereux and Sainsaulieu, 1997) upgrade this scheme to second-order accuracy by making the initial states piecewise linear rather than piecewise constant (Nessyahu and Tadmor, 1990). Our objective is (eventually) to improve on this by developing a non-staggered version, since that should improve the resolution of slowly-moving waves. These are the waves that carry the anisotropic stresses in our model, and we want these to be as accurate as possible. Our intention is to return to (44), inserting into that formula ‘reasonable’ approximations for the interface solutions $\mathbf{u}_{j \pm 1/2}(t)$. Since an accurate solution even of the linearized Riemann problem is not feasible, we examine in Section 5 the solutions to some representative Riemann problems to see which features are typical and must be represented.

4.1. CHOOSING THE LINEARISATIONS

To apply the idea at all, however, we must first choose a linear Riemann problem to approximate, and here, that means choosing linearisations of both \mathbf{A} and \mathbf{Q} . It seems natural to choose $\tilde{\mathbf{A}}$, the linearisation of \mathbf{A} , by satisfying the usual condition (Roe, 1981),

$$\tilde{\mathbf{A}}(\mathbf{u}_L, \mathbf{u}_R)[\mathbf{u}_R - \mathbf{u}_L] = \mathbf{f}_R - \mathbf{f}_L, \quad (48)$$

so that the near-frozen solution is accurate, at least for those components that we wish to treat conservatively. (Bereux and Sainsaulieu, 1997) suggest that $\tilde{\mathbf{Q}}$ be chosen in the following way. Let $\mathbf{u}_\infty(\mathbf{u})$ be the limit as $t \rightarrow \infty$ of solution to the ODE (2) with \mathbf{u} as initial data. Then we should choose $\tilde{\mathbf{Q}}$ in such a way that

$$\lim_{t \rightarrow \infty} \exp(\tilde{\mathbf{Q}}t)\mathbf{u} = \mathbf{u}_\infty(\mathbf{u}), \quad (49)$$

that is, the linearized relaxation problem drives the initial data to the correct equilibrium solution. This ensures that spatially uniform data will be correctly solved at large times and that our equilibrium Riemann problem will be supplied with the proper data. For the Bereux-Sainsaulieu approach, which ‘hides’ the equilibrium waves, this is sufficient. In our approach, we require such a $\tilde{\mathbf{Q}}$ in each cell, to secure the proper decay of the

piecewise constant states, and also at each interface, to obtain the proper interactions.

In the cells, we need to know everything about $\tilde{\mathbf{Q}}$. However, for the problem on which we focus, this information is simply the two relaxation times τ_t and τ_r . We have a choice whether to evaluate these from (37) at $t = t^n$ or $t = t^{n+1}$ or in some average sense. There seems no overwhelming reason for any choice. All choices will drive the data to the correct limit, but at different rates.

At the interfaces, we need only to know enough about $\tilde{\mathbf{Q}}$ to be able to form the product $\mathbf{L}(\tilde{\mathbf{Q}})\tilde{\mathbf{A}}\mathbf{R}(\tilde{\mathbf{Q}})$. But since in our problem \mathbf{L}, \mathbf{R} are constant matrices, any linearisation of \mathbf{Q} will serve; in fact there is no need to construct any linearisation at all.

5. Typical Riemann Solutions

We now turn to a key issue: what kinds of behaviour should our approximate Riemann solver be able to represent? Probably there is no absolute answer to this question. It will depend on the purposes to which the code is to be put and, particularly, whether there is near-frozen behaviour that needs to be resolved.

If the relaxing gas model is to be used to replace the Navier-Stokes model under ‘normal’, continuum conditions, then the relaxation times obtained from (37) are tiny, of the order 10^{-10} s with corresponding length scales for wave propagation around 10^{-8} m. Events having a macroscopic influence on shear or boundary layers are on a much larger scale, so that an approximate Riemann solver needs to be valid only in the near-equilibrium limit.

On the other hand, there are instances where the details of the relaxation may be of great interest. For very high altitude flight, the gas is rarefied, and so the relaxation scales are several orders of magnitude larger than in the continuum case. In the study of microscale devices, the characteristic scales of the devices are minuscule and, therefore, more comparable to the relaxation scales.

We present here two representative solutions to the diatomic ten moment system to provide some insight into the types of features which might arise in the transition between the frozen and the equilibrium limits. Each was generated with a high-resolution MUSCL scheme using Roe’s linearisation and the double minmod slope limiter. A very fine uniform grid of 10^4 cells was used to properly resolve the frozen behavior, but the cell size was rescaled and the solution projected onto the resulting coarser grid whenever the propagating waves approached the limits of the computational domain. In both simulations, the gas was taken to be air with a molecular mass of 28.96kg/kmol. State variables are normalized with respect to the

equilibrium state $(\rho_0, p_0) = (1 \text{ kg/m}^3, 10^5 \text{ N/m}^2)$, with the characteristic relaxation time scale taken to be the translational relaxation time, τ_t , and the characteristic velocity scale taken to be $\sqrt{p_0/\rho_0}$.

5.1. EXAMPLE I: INITIAL SHEAR PROBLEM

If the relaxation system is to be used as a replacement for the Navier-Stokes system, it will typically need to resolve shear layers. We take as initial conditions $\rho(x, 0) = E_{\text{rot}}(x, 0) = 1$, $P_{ij}(x, 0) = \delta_{ij}$, and $\mathbf{u} = (0, -0.1 \text{sgn}(x), 0)^T$. This represents a shear in the u_2 velocity. Six time slices of the evolution of u_2 and P_{12} are presented in Figure 6.

In fact, this problem is mathematically equivalent to the impulsively stopped plate initial conditions for the incompressible Navier-Stokes equations. Consider the half-plane $x > 0$ with the plate $x = 0$. For $t < 0$, the plate and the fluid above it would be moving with a uniform velocity of $u_2^0 = -0.1$, but the plate would impulsively stop at time $t = 0$. In this case, the Navier-Stokes equations reduce to a heat equation for the velocity u_2 : $\partial_t u_2 = \partial_{xx} u_2$. The solution is

$$u_2(x, t) = -u_2^0 \operatorname{erf} \left\{ \frac{x}{2\sqrt{t}} \right\}, \quad (50)$$

which is also plotted in Figure 6. For this type of Riemann problem, it should be possible to find very efficient approximations.

5.2. EXAMPLE II: A NON-EQUILIBRIUM SHOCK TUBE

It is useful to consider more general initial conditions in order to gauge just how complicated the solution of the Riemann problem could become. Here we consider a generalization of Sod's commonly-used test problem (Sod, 1978), where the initial data, presented in Table 1, are set to somewhat random non-equilibrium conditions that will excite the various frozen waves. In Figure 7, the solutions are presented at three different times. It can be seen that the solution is much more complicated than the previous example, but that by a dimensionless time $t = 100$ the solution is very similar to the well-known Euler solution.

	ρ	u_1	u_2	u_3	P_{11}	P_{12}	P_{13}	P_{22}	P_{23}	P_{33}	E_{rot}
L	1	0	-0.1	0.2	0.8	0.03	0.02	1.2	0.01	1.1	0.95
R	0.125	0	0.1	-0.15	0.1	0.001	0.002	0.3	0.003	0.05	0.025

TABLE 1. Initial data for non-equilibrium Sod problem.

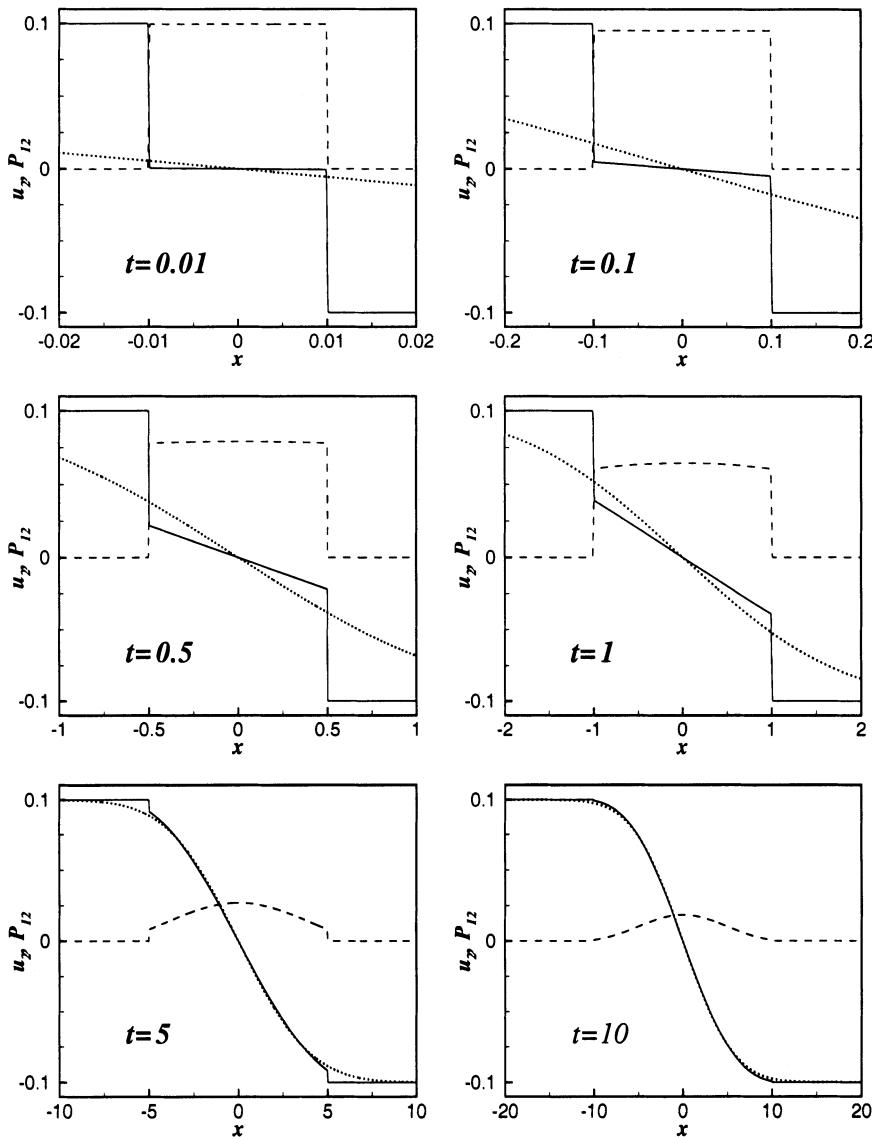


Figure 6. Evolution of the velocity u_2 (—) and the shear pressure P_{12} (---) for the initial shear layer problem. Note that the spatial scale varies for each plot. The analytic solution to the impulsively stopped plate problem (50) is represented, for u_2 only, by the dotted line (···). At early times, the initial discontinuity causes frozen quasi-acoustic waves to propagate away from the origin with a normalized speed of unity. These initial discontinuities decay, and structure forms between the frozen waves. At time $t = 10$, the solution is near equilibrium, but vestiges of the frozen waves are still apparent. The error function and Gaussian structures predicted in (29) and (30) for the linear problem clearly arise.

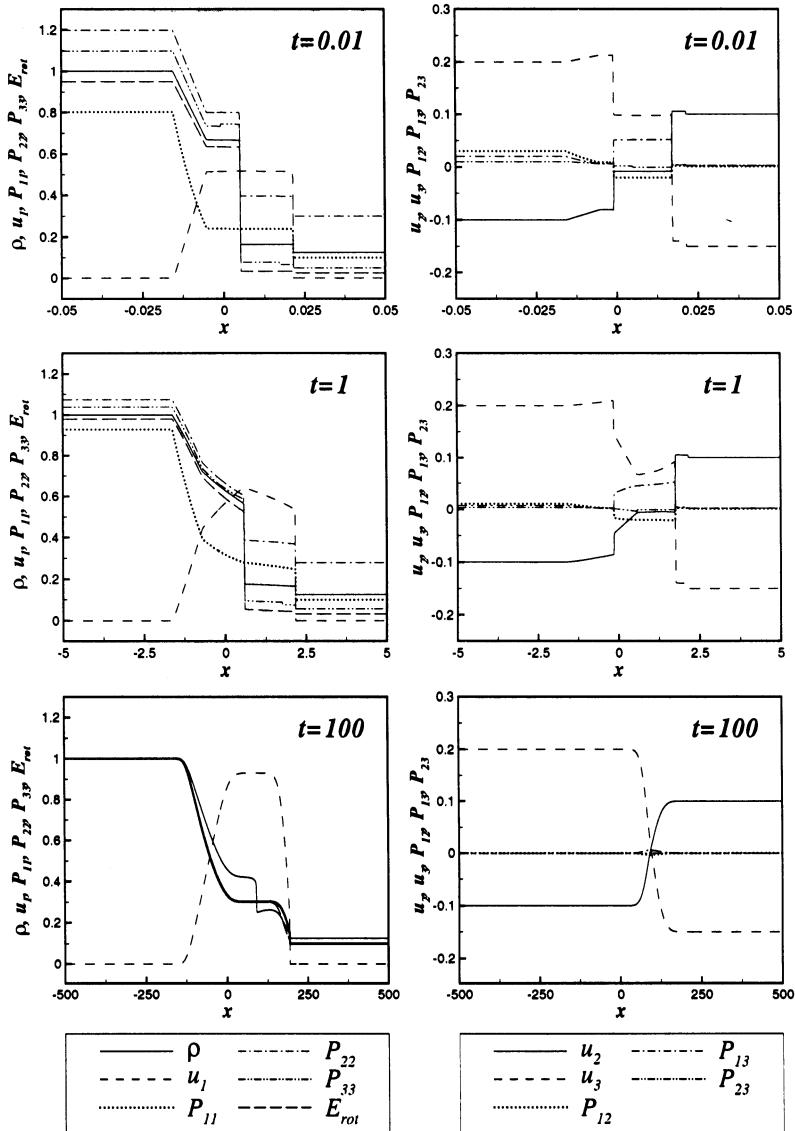


Figure 7. Solutions at three different times to the non-equilibrium Sod problem. Each row corresponds to a particular time level, and the eleven state variables are divided between the two columns. At $t = 0.01$, the solution is effectively still in the frozen limit, and waves with all five frozen speeds are clearly evident. The solution is in transition at $t = 1$; note the structure in the velocity profiles and the decay of the shear waves. Also, outside the domain of influence, the components of P_{ij} are approaching their equilibrium values. By time $t = 100$, the solution is approaching equilibrium. The components of internal energy have nearly equilibrated, although a small remnant of the frozen shock still remains. This is predicted by the linear analysis (20). This solution is nearing that for the Euler equations except that the waves are slightly smoothed by the relaxation processes. They are well enough resolved that none of this can be attributed to numerics.

Physically, $t = 100$ corresponds to a time on the order of 10^{-8} s, which is significantly less than, say, the period of sound waves at the uppermost limit of human hearing ($5 \cdot 10^{-5}$ s). Therefore, even for acoustic calculations, it seems that a Riemann solver would only need to be accurate in the near-equilibrium limit and could possibly be based upon simple modifications to the of the Riemann problem for the Euler equations.

6. Conclusions

We have tried to survey those analytical properties of hyperbolic relaxation systems most relevant to computation. The general case, even when linearised, is complex, but we feel that some common physical systems may prove to be quite tractable. We have examined one typical system in detail and find that it is less formidable than one might fear, especially if the relaxation variables are expressed in non-conservative form.

We have proposed, but not yet implemented, a finite-volume Godunov-type method that requires an estimate of the interface flux from a non-self-similar Riemann problem. On the strength of some highly resolved numerical solutions to typical Riemann problems, and some analysis of the decay of the early time solution, we conjecture that in many practical situations the flux estimate should be obtainable from an efficient approximation.

References

- Bereux, F., Sainsaulieu, L., A Roe-type Riemann solver for Hyperbolic Systems with relaxation based on time-dependent wave decomposition, *Num. Math.*, **77**, p 1433, 1997.
- Bourgeade A., Le Floch, P., Raviart, P-A., An asymptotic expansion for the solution of the generalised Riemann Problem, Part 2: Application to the system of gas dynamics, *Ann. Inst. Henri Poincaré*, **6** p 437, 1989.
- Cafisch, R., Jin, S., Russo, G., Uniformly accurate schemes for hyperbolic systems with relaxation, *SIAM Numer. Anal.*, **34**, p 246, 1997.
- Harten, A., Lax, P.D., van Leer, B., On upstream differencing and Godunov-type schemes for hyperbolic conservation laws, *SIAM Review*, **25**, pp 35-61, 1983.
- Le Floch, P., Raviart, P-A., An asymptotic expansion for the solution of the generalised Riemann Problem, Part 1: General theory, *Ann. Inst. Henri Poincaré*, **5** p 179, 1987.
- Levermore, C.D., Moment closure hierarchies for kinetic theories, *J. Stat. Phys.*, **83**, p 1021, 1996.
- Pember, R., Numerical methods for hyperbolic conservation laws with stiff relaxation, II higher-order Godunov methods, *SIAM J. Sci. Stat. Comput.*, **14**, no 4, p 824, 1993.
- Nessyahu, H., Tadmor, E., Non-oscillatory central differencing for hyperbolic conservation laws, *J. Comput. Phys.* **87**, p 408, 1990.
- Roe, P. L., Approximate Riemann solvers, parameter vectors and difference schemes, *J. Comput. Phys.* **43**, p 357, 1981.
- Sod, G., A survey of several finite difference methods for systems of nonlinear hyperbolic conservation laws, *J. Comput. Phys.* **27**, p 41, 1978.
- Zeng, Y., Gasdynamics in thermal nonequilibrium and general hyperbolic systems with relaxation, preprint, University of Alabama at Birmingham, 1999.

THERMODYNAMICS AND HYPERBOLIC SYSTEMS OF BALANCE LAWS IN CONTINUUM MECHANICS

E. I. ROMENSKY

Sobolev Institute of Mathematics,

Novosibirsk, 630090, Russia

Email: evrom@math.nsc.ru

Abstract.

The class of thermodynamically compatible systems of balance laws with source terms is considered. Every system of this class is hyperbolic and generated by only one thermodynamic potential. Besides, each equation of such system has a conservative form. For instance, equations of motion of elastic condutors and multiphase media are considered.

1. Introduction

The goal of this article is to formulate the thermodynamics laws and clarify the connection between these laws and the well-posedness of systems of equations of mechanics of continuous media.

In the processes without dissipation, the equations of continuum mechanics can be written in the form of the conservation laws

$$\frac{\partial \Phi_i^0(q_1, \dots, q_n)}{\partial t} + \frac{\partial \Phi_i^k(q_1, \dots, q_n)}{\partial x_k} = 0, \quad i = 1, \dots, N, \quad (1)$$

where $\Phi_i^\alpha(q_1, \dots, q_n)$ ($\alpha = 0, 1, 2, 3$) are known functions of parameters q_1, \dots, q_n characterizing the states of a medium.

In many situations, the number of conservation laws N turns out to be greater than the number of unknowns n ($N > n$). One of the reasons is the fact that every closed system of equations of mechanics of continuous media admits an additional conservation laws, e.g., the conservation law for energy. Another reason is the fact that there are additional stationary conservation laws that are satisfied by solutions of nonstationary equations. Such a situation can occur in many cases.

The following question arises: How to formulate the thermodynamics laws for systems of the form (1)? As is known, the local thermodynamics deals with thermodynamic potentials related with parameters of states of a medium by the first principle of thermodynamics which has, for example, in the case of a gas, the form

$$dE = -pdV + TdS,$$

where E is the internal energy, p is the pressure, V is the specific volume, T is the temperature, and S is the entropy.

To formulate the thermodynamic laws for the system (1), it is necessary to clarify the structure of the functions Φ^α .

S. K. Godunov (Godunov, 1961) was the first researcher who begins to study deep connections between thermodynamics laws and well-posedness of governing differential equations of continuum mechanics. In (Godunov, 1961), Godunov proposed to consider a class of symmetric hyperbolic systems of conservation laws

$$\frac{\partial L_{q_i}^0}{\partial t} + \frac{\partial L_{q_i}^k}{\partial x_k} = 0, \quad i = 1, \dots, n, \quad (2)$$

where $L^\alpha(q_1, \dots, q_n)$ ($\alpha = 0, 1, 2, 3$) are functions (potentials) of parameters q_1, \dots, q_n of the state of a medium.

Systems of the form (2) possess two remarkable properties. First, they admit the following additional conservation law:

$$\frac{\partial(q_i L_{q_i}^0 - L^0)}{\partial t} + \frac{\partial(q_i L_{q_i}^k - L^k)}{\partial x_k} = 0. \quad (3)$$

To verify this fact, it is necessary to sum all the equations in (2) multiplied by the corresponding factors q_i . After some transformations, we obtain equation (3). Second, the system (2) can be written in the form of a symmetric system as follows:

$$L_{q_i q_j}^0 \frac{\partial q_j}{\partial t} + L_{q_i q_j}^k \frac{\partial q_j}{\partial x_k} = 0. \quad (4)$$

Moreover, if the potential L^0 is convex with respect to the variables q_1, \dots, q_n (which is equivalent to the positive definiteness of the matrix $L_{q_i q_j}^0$), then the system (4) is hyperbolic.

At present, there are many publications devoted to thermodynamic properties of equations of mechanics and physics (cf., for example, (Godunov, 1986), (Godunov and Romensky, 1998), (Ruggeri and Müller, 1999) and the references there).

Although the class of equations (2) includes a lot of equations of mechanics, there are many systems that do not belong to this class. For example, this class does not contain systems including additional stationary conservation laws. Generalizations of the classification (2) in this case (cf., for example, (Godunov, 1972), (Friedrichs, 1978)) cannot be explained from the thermodynamic point of view. Furthermore, too many generating potentials L^α are used in the case of the class (2), whereas all the properties of some object can be characterized by only one potential (e.g., the internal energy) within the framework of the classic thermodynamics. Therefore, we tried to find a class of thermodynamically compatible systems of conservation laws (Godunov and Romensky, 1995), (Romensky, 1998). It should be noted that the results of (Godunov and Romensky, 1995), (Romensky, 1998) were obtained after careful analysis of a large number of concrete systems of continuum mechanics. The formulation of a class of thermodynamically compatible systems in (Godunov and Romensky, 1995), (Romensky, 1998) allows us to introduce new versions of governing equations of various media with complicated properties.

We emphasize that every system of the class of thermodynamically compatible systems of conservation laws is generated by only one potential and can be reduced to a symmetric hyperbolic system,

One of possible generalizations of this class to the case of systems invariant relative to rotations is considered in (Godunov, Mikhailova and Romensky, 1996), (Mikhailova, 1997).

Here, we generalize the class of thermodynamically compatible systems to the case of balance laws with source terms.

2. Thermodynamically compatible systems of conservation laws

Some systems of equations of continuum media describing physical processes without dissipation can be joined to the class of systems of conservation laws of the form

$$(5.1) \quad \frac{\partial L_{q_i}}{\partial t} + \frac{\partial (u_k L)_{q_i}}{\partial x_k} = 0,$$

$$(5.2) \quad \frac{\partial L_{u_i}}{\partial t} + \frac{\partial [(u_k L)_{u_i} - r_{i\alpha} L_{r_{k\alpha}} - b_i L_{b_k} - d_i L_{d_k} + j_k L_{j_i} - \delta_{ik} j_\alpha L_{j_\alpha}]}{\partial x_k} = 0,$$

$$(5.3) \quad \frac{\partial L_{r_{il}}}{\partial t} + \frac{\partial [u_k L_{r_{il}} - u_i L_{r_{kl}}]}{\partial x_k} = 0,$$

$$(5.4) \quad \frac{\partial L_{d_i}}{\partial t} + \frac{\partial [u_k L_{d_i} - u_i L_{d_k} - e_{ikl} b_l]}{\partial x_k} = 0, \quad (5)$$

$$(5.5) \quad \frac{\partial L_{b_i}}{\partial t} + \frac{\partial [u_k L_{b_i} - u_i L_{b_k} + e_{ikl} d_l]}{\partial x_k} = 0,$$

$$(5.6) \quad \frac{\partial L_n}{\partial t} + \frac{\partial [u_k L_n + j_k]}{\partial x_k} = 0,$$

$$(5.7) \quad \frac{\partial L_{j_k}}{\partial t} + \frac{\partial [u_\alpha L_{j_\alpha} + n]}{\partial x_k} = 0,$$

where e_{ikl} is the unit pseudoscalar. Every term of equations in (5) is determined by only one generating potential L which depends on variables

$$q_i \ (i = 0, 1, 2, \dots); \ u_i, \ r_{il} \ (i, l = 1, 2, 3);$$

$$d_i, \ b_i \ (i = 1, 2, 3); \ n, \ j_k \ (k = 1, 2, 3).$$

We divided variables into groups in order to emphasize the couple structure of equations of the system (5). Note that the number of variables q_i , as well as the number of pairs of variables d_i, b_i and n, j_k can as large as desired. In this case, we must consider the same number of pairs of equations in the system (5). A pair of variables u_i, r_{il} is connected with the velocity and deformation. Hence equations (5.2) and (5.3) do not admit any repetition.

The system (5) is closed. However, there are many additional stationary conservation laws compatible with (5). The form of these conservation laws is defined by the structure of flux terms formed by the operators **grad**, **div**, **rot**:

$$(6.1) \quad \frac{\partial L_{r_{il}}}{\partial x_i} = 0, \quad l = 1, 2, 3,$$

$$(6.2) \quad \frac{\partial L_{d_i}}{\partial x_i} = 0,$$

$$(6.3) \quad \frac{\partial L_{b_i}}{\partial x_i} = 0, \tag{6}$$

$$(6.4) \quad \frac{\partial L_{j_k}}{\partial x_i} - \frac{\partial L_{j_i}}{\partial x_k} = 0, \quad i \neq k.$$

It is easy to prove that the systems (5) and (6) are compatible. For example, to verify that equation (6.1) is satisfied by solutions to the system (5), it suffices to differentiate equation (5.3) with respect to x_i . We find

$$\frac{\partial}{\partial t} \frac{\partial L_{r_{il}}}{\partial x_i} = 0.$$

If $\partial(L_{r_{il}})/\partial x_i = 0$ holds for $t = 0$, then the same equality must hold for $t > 0$.

Note that the system (5), together with the system (6), admits an additional nonstationary conservation law which is similar to the first principle

of thermodynamics for the considered system of thermodynamically compatible conservation laws. To deduce this conservation law, we consider the following linear combination of equations of the systems (5) and (6):

$$\begin{aligned} q_i \cdot (5.1) + u_i \cdot (5.2) + r_{il} \cdot (5.3) + d_i \cdot (5.4) + b_i \cdot (5.5) + n \cdot (5.6) + j_k \cdot (5.7) \\ + u_i r_{il} \cdot (6.1) + u_i d_i \cdot (6.2) + u_i b_i \cdot (6.3) + 2u_i j_k \cdot (6.4). \end{aligned}$$

After simple but cumbersome transformations we obtain the equation

$$\frac{\partial \Pi_0}{\partial t} + \frac{\partial \Pi_k}{\partial x_k} = 0, \quad (7)$$

where

$$\begin{aligned} \Pi_0 &= q_i L_{q_i} + u_i L_{u_i} + r_{il} L_{r_{il}} + d_i L_{d_i} + b_i L_{b_i} + n L_n + j_k L_{j_k} - L, \\ \Pi_k &= u_k (q_i L_{q_i} + u_i L_{u_i} + r_{il} L_{r_{il}} + d_i L_{d_i} + b_i L_{b_i} + n L_n) \\ &\quad + u_l (j_k L_{j_l} - r_{l\beta} L_{r_{k\beta}} - d_l L_{d_k} - b_l L_{b_k}) + j_k n + e_{k\alpha\beta} d_\alpha b_\beta. \end{aligned}$$

Equation (7) is the energy conservation law for the system (5) complemented with the system (6).

The class of equations of the systems (5) and (6) contains a number of examples of systems of equations of continuum mechanics such as the gas-dynamics equations, magnetohydrodynamics equations, equations of nonlinear elasticity, and equations of electrodynamics of moving dielectrics, etc.

It is important to note that, using (6), the system (5) can be written in the form of equivalent symmetric system which is a hyperbolic one provided that the generating potential L is convex. To perform the required transformation, it is necessary to add to (5.2) the following combinations of stationary equations of the system (6):

$$r_{i\alpha} \frac{\partial L_{r_{k\alpha}}}{\partial x_k} + d_i \frac{\partial L_{d_k}}{\partial x_k} + b_i \frac{\partial L_{b_k}}{\partial x_k} + j_k \left(\frac{\partial L_{j_k}}{\partial x_i} - \frac{\partial L_{j_i}}{\partial x_k} \right) = 0.$$

To equations (5.3), (5.4), (5.5), and (5.7), we respectively add the equalities

$$u_i \frac{\partial L_{r_{kl}}}{\partial x_k} = 0, \quad u_i \frac{\partial L_{d_k}}{\partial x_k} = 0, \quad u_i \frac{\partial L_{b_k}}{\partial x_k} = 0, \quad u_i \left(\frac{\partial L_{j_k}}{\partial x_i} - \frac{\partial L_{j_i}}{\partial x_k} \right) = 0.$$

We obtain a system whose quasilinear form is symmetric:

$$\frac{\partial L_{q_i}}{\partial t} + \frac{\partial (u_k L)_{q_i}}{\partial x_k} = 0,$$

$$\begin{aligned}
& \frac{\partial L_{u_i}}{\partial t} + \frac{\partial(u_k L)_{u_i}}{\partial x_k} - L_{r_{k\alpha}} \frac{\partial r_{i\alpha}}{\partial x_k} - L_{d_k} \frac{\partial d_i}{\partial x_k} - L_{b_k} \frac{\partial b_i}{\partial x_k} + L_{j_i} \frac{\partial j_k}{\partial x_k} - L_{j_\alpha} \frac{\partial j_\alpha}{\partial x_i} = 0, \\
& \frac{\partial L_{r_{il}}}{\partial t} + \frac{\partial(u_k L)_{r_{il}}}{\partial x_k} - L_{r_{kl}} \frac{\partial u_i}{\partial x_k} = 0, \\
& \frac{\partial L_{d_i}}{\partial t} + \frac{\partial(u_k L)_{d_i}}{\partial x_k} - L_{d_k} \frac{\partial u_i}{\partial x_k} - e_{ikl} \frac{\partial b_l}{\partial x_k} = 0, \\
& \frac{\partial L_{b_i}}{\partial t} + \frac{\partial(u_k L)_{b_i}}{\partial x_k} - L_{b_k} \frac{\partial u_i}{\partial x_k} + e_{ikl} \frac{\partial d_l}{\partial x_k} = 0, \\
& \frac{\partial L_n}{\partial t} + \frac{\partial(u_k L)_n}{\partial x_k} + \frac{\partial j_k}{\partial x_k} = 0, \\
& \frac{\partial L_{j_l}}{\partial t} + \frac{\partial(u_k L)_{j_l}}{\partial x_k} + L_{j_\alpha} \frac{\partial u_\alpha}{\partial x_l} - L_{j_l} \frac{\partial u_k}{\partial x_k} + \frac{\partial n}{\partial x_l} = 0.
\end{aligned} \tag{8}$$

This system is hyperbolic if the matrix of second-order derivatives of L is positive definite, i.e., if the generating potential L is convex.

Taking more pairs of variables d_i , b_i , we can easily obtain analogs of the energy conservation law (7) and the symmetric system (8) for such generalized system.

3. Thermodynamically compatible systems of balance laws

The class of thermodynamically compatible systems of conservation laws (cf. Sec. 2) includes only equations for processes without dissipation, whereas the phenomenon of dissipation is of particular interest. We consider the formalization of systems of balance laws with lower source terms that describe processes with dissipation. The question is how to add lower terms to equations of the systems (5) and (6) in such a way that the conservative structure of equations remain without change and the thermodynamics laws hold. We discuss only the method of introducing source terms in equations of the system (5) that are associated with stationary laws of the system (6). After that, it is easy to introduce source terms in the remaining conservation laws.

Instead of (5) and (6), we consider the system

$$(9.1) \quad \frac{\partial L_{q_\omega}}{\partial t} + \frac{\partial(u_k L)_{q_\omega}}{\partial x_k} = \frac{1}{q_\omega} (r_{ij} \varphi_{ij} + d_i J_i + b_i K_i + j_k \pi_k),$$

$$(9.2) \quad \frac{\partial L_{q_i}}{\partial t} + \frac{\partial(u_k L)_{q_i}}{\partial x_k} = 0, \quad i = 0, 1, 2, \dots,$$

$$(9.3) \quad \frac{\partial L_{u_i}}{\partial t} + \frac{\partial[(u_k L)_{u_i} - r_{i\alpha} L_{r_{k\alpha}} - b_i L_{b_k} - d_i L_{d_k} + j_k L_{j_i} - \delta_{ik} j_\alpha L_{j_\alpha}]}{\partial x_k} = 0,$$

$$(9.4) \quad \frac{\partial L_{r_{il}}}{\partial t} + \frac{\partial[u_k L_{r_{il}} - u_i L_{r_{kl}}]}{\partial x_k} = -(u_i \beta_l + \varphi_{il}),$$

$$(9.5) \quad \frac{\partial L_{d_i}}{\partial t} + \frac{\partial [u_k L_{d_i} - u_i L_{d_k} - e_{ikl} b_l]}{\partial x_k} = -(u_i R + J_i),$$

$$(9.6) \quad \frac{\partial L_{b_i}}{\partial t} + \frac{\partial [u_k L_{b_i} - u_i L_{b_k} + e_{ikl} d_l]}{\partial x_k} = -(u_i Q + K_i), \quad (9)$$

$$(9.7) \quad \frac{\partial L_n}{\partial t} + \frac{\partial (u_k L_n + j_k)}{\partial x_k} = 0,$$

$$(9.8) \quad \frac{\partial L_{j_k}}{\partial t} + \frac{\partial (u_\alpha L_{j_\alpha} + n)}{\partial x_k} = -(e_{k\alpha\beta} u_\alpha \omega_\beta + \pi_k),$$

$$(9.9) \quad \frac{\partial L_{r_{kl}}}{\partial x_k} = \beta_l,$$

$$(9.10) \quad \frac{\partial L_{d_i}}{\partial x_i} = R,$$

$$(9.11) \quad \frac{\partial L_{b_i}}{\partial x_i} = Q,$$

$$(9.12) \quad \frac{\partial L_{j_k}}{\partial x_\alpha} - \frac{\partial L_{j_\alpha}}{\partial x_k} = -e_{k\alpha\beta} \omega_\beta.$$

Here, we introduced new variables β_l , φ_{il} , R , J_i , Q , K_i , ω_b , and π_k . We specify the choice of these variables below. Note that the generating potential depends on the “old” variables.

First of all, we verify that the right-hand sides introduced in equations of the systems (5) and (6) do not contradict the energy conservation law. Indeed, reasoning in the same way as in Sec. 2, we can consider the linear combination

$$q_\omega \cdot (9.1) + q_i \cdot (9.2) + u_i \cdot (9.3) + r_{il} \cdot (9.4) + d_i \cdot (9.5) + b_i \cdot (9.6) + n \cdot (9.7) + j_k \cdot (9.8)$$

$$+ u_i r_{il} \cdot (9.9) + u_i d_i \cdot (9.10) + u_i b_i \cdot (9.11) + 2u_i j_k \cdot (9.12)$$

and check that the right-hand sides disappear. As a result, we obtain the same energy conservation law (7):

$$\frac{\partial}{\partial t} (q_i L_{q_i} + u_i L_{u_i} + r_{il} L_{r_{il}} + d_i L_{d_i} + b_i L_{b_i} + n L_n + j_k L_{j_k} - L)$$

$$+ \frac{\partial}{\partial x_k} [u_k (q_i L_{q_i} + u_i L_{u_i} + r_{il} L_{r_{il}} + d_i L_{d_i} + b_i L_{b_i} + n L_n)]$$

$$+ u_l (j_k L_{j_l} - r_{l\beta} L_{r_{k\beta}} - d_l L_{d_k} - b_l L_{b_k}) + j_k n + e_{k\alpha\beta} d_\alpha b_\beta] = 0.$$

As was noted, in comparison with (5) and (6), the system (9) contains new variables. To close the system (9), it suffices to indicate the dependence of the functions φ_{il} , J_i , K_i , and π_k on the variables (parameters of the state

of a medium) q_ω , q_i , u_i , r_{il} , d_i , b_i , n , and j_k . Then we study the following (equivalent) system:

$$(10.1) \quad \frac{\partial L_{q_\omega}}{\partial t} + \frac{\partial(u_k L)_{q_\omega}}{\partial x_k} = \frac{1}{q_\omega} (r_{ij} \varphi_{ij} + d_i J_i + b_i K_i + j_k \pi_k),$$

$$(10.2) \quad \frac{\partial L_{q_i}}{\partial t} + \frac{\partial(u_k L)_{q_i}}{\partial x_k} = 0, \quad i = 0, 1, 2, \dots,$$

$$(10.3) \quad \frac{\partial L_{u_i}}{\partial t} + \frac{\partial[(u_k L)_{u_i} - r_{i\alpha} L_{r_{k\alpha}} - b_i L_{b_k} - d_i L_{d_k} + j_k L_{j_i} - \delta_{ik} j_\alpha L_{j_\alpha}]}{\partial x_k} = 0,$$

$$(10.4) \quad \frac{\partial L_{r_{il}}}{\partial t} + \frac{\partial(u_k L)_{r_{il}}}{\partial x_k} - L_{r_{kl}} \frac{\partial u_i}{\partial x_k} = -\varphi_{ij},$$

$$(10.5) \quad \frac{\partial L_{d_i}}{\partial t} + \frac{\partial[(u_k L)_{d_i} - e_{ikl} b_l]}{\partial x_k} - L_{d_k} \frac{\partial u_i}{\partial x_k} = -J_i, \quad (10)$$

$$(10.6) \quad \frac{\partial L_{b_i}}{\partial t} + \frac{\partial[(u_k L)_{b_i} + e_{ikl} d_l]}{\partial x_k} - L_{b_k} \frac{\partial u_i}{\partial x_k} = -K_i,$$

$$(10.7) \quad \frac{\partial L_n}{\partial t} + \frac{\partial[(u_k L)_n + j_k]}{\partial x_k} = 0,$$

$$(10.8) \quad \frac{\partial L_{j_k}}{\partial t} + u_\alpha \frac{\partial L_{j_k}}{\partial x_\alpha} + L_{j_\alpha} \frac{\partial u_\alpha}{\partial x_k} + \frac{\partial n}{\partial x_k} = -\pi_k,$$

$$(10.9) \quad \frac{\partial L_{r_{kl}}}{\partial x_k} = \beta_l,$$

$$(10.10) \quad \frac{\partial L_{d_i}}{\partial x_i} = R,$$

$$(10.11) \quad \frac{\partial L_{b_i}}{\partial x_i} = Q,$$

$$(10.12) \quad \frac{\partial L_{j_k}}{\partial x_\alpha} - \frac{\partial L_{j_\alpha}}{\partial x_k} = -e_{k\alpha\beta} \omega_\beta.$$

The system (10) differs from the system (9) by (equivalent) form of equations (10.4), (10.5), (10.6), and (10.8). For a given generating potential

$$L(q_\omega, q_i, u_i, r_{il}, d_i, b_i, n, j_k),$$

equations (10.1)-(10.8) form a closed system. Stationary balance laws (10.9)-(10.12) can be used for computing β_l , Q , R , and ω_β .

It is important to note that the systems (10.1)-(10.8) and (10.9)-(10.12) must be compatible. This requirement leads to differential connections between two groups of functions β_l , Q , R , ω_β and φ_{il} , J_i , K_i , π_k . The connections can be expressed in the form

$$\frac{\partial \beta_l}{\partial t} + \frac{\partial(u_i \beta_l + \varphi_{il})}{\partial x_i} = 0,$$

$$\begin{aligned}\frac{\partial R}{\partial t} + \frac{\partial(u_i R + J_i)}{\partial x_i} &= 0, \\ \frac{\partial Q}{\partial t} + \frac{\partial(u_i Q + K_i)}{\partial x_i} &= 0, \\ \frac{\partial \omega_\gamma}{\partial t} + \frac{\partial(u_i \omega_\gamma - u_\gamma \omega_i + e_{\gamma i \mu} \pi_\mu)}{\partial x_i} &= 0.\end{aligned}\tag{11}$$

It is easy to prove that the expressions (11) provide the compatibility of the system (9) and the system (10). For example, if the equality

$$\frac{\partial L_{r_{kl}}}{\partial x_k} - \beta_l = 0$$

holds for $t = 0$, then it remains valid for $t > 0$. Indeed, differentiating equation (9.4) with respect to x_i and subtracting from the result the first equation of the system (11), we find

$$\frac{\partial}{\partial t} \frac{\partial L_{r_{il}}}{\partial x_i} - \frac{\partial \beta_l}{\partial t} = \frac{\partial}{\partial t} \left(\frac{\partial L_{r_{il}}}{\partial x_i} - \beta_l \right) = 0,$$

which implies the required fact.

Thus, we have indicated the class of thermodynamically compatible systems of balance laws of continuum mechanics in the form the system (9) and the conditions (11). Later, we give some examples of equations subject to the above formalism.

We note that the system (9) cannot be treated by the symmetrization method applied to the systems of conservation laws (without lower terms) in Sec. 2. Indeed, the quasilinear form of equation (10.3) u_j contains the functions β_l , R , Q , and ω_β which can be expressed in terms of the derivatives of q_ω , q_i , u_i , r_{il} , d_i , b_i , n , and j_k . Therefore, in particular cases, we symmetrize, in fact, an extended system of equations that includes the derivatives of unknown functions (cf., for example, (Romensky, 1995)), whereas the general formal method is not still found.

4. Examples

Consider two examples of systems of equations of continuum mechanics, where dissipation is taken into account by source terms. These systems can be studied by method of the previous section.

4.1. EQUATIONS OF MOTION OF ELASTIC CONDUCTORS

The first example is connected with modeling of electrodynamics of moving conductors (Landau and Lifshitz, 1982). The description of dissipative

processes in an elastic conductor under the action of electric current is well known. The Ohm law is used in the closure of the Maxwell equations of electromagnetic field.

In order to describe the motion of moving elastic conductors by the system (9), we choose variables $q_\omega, q_0, u_1, u_2, u_3, r_{11}, r_{12}, \dots, r_{33}, d_1, d_2, d_3, b_1, b_2$, and b_3 . In accordance with this choice, we extract the corresponding subsystem of the system (9):

$$\begin{aligned}
 (12.1) \quad & \frac{\partial L_{q_\omega}}{\partial t} + \frac{\partial(u_k L)_{q_\omega}}{\partial x_k} = \frac{d_i J_i}{q_\omega}, \\
 (12.2) \quad & \frac{\partial L_{q_0}}{\partial t} + \frac{\partial(u_k L)_{q_0}}{\partial x_k} = 0, \\
 (12.3) \quad & \frac{\partial L_{u_i}}{\partial t} + \frac{\partial[(u_k L)_{u_i} - r_{i\alpha} L_{r_{i\alpha}} - b_i L_{b_k} - d_i L_{d_k}]}{\partial x_k} = 0, \\
 (12.4) \quad & \frac{\partial L_{r_{il}}}{\partial t} + \frac{\partial[u_k L_{r_{il}} - u_i L_{r_{kl}}]}{\partial x_k} = 0, \\
 (12.5) \quad & \frac{\partial L_{d_i}}{\partial t} + \frac{\partial[u_k L_{d_i} - u_i L_{d_k} - e_{ikl} b_l]}{\partial x_k} = -(u_i R + J_i), \\
 (12.6) \quad & \frac{\partial L_{b_i}}{\partial t} + \frac{\partial[u_k L_{b_i} - u_i L_{b_k} + e_{ikl} d_l]}{\partial x_k} = 0, \\
 (12.7) \quad & \frac{\partial L_{r_{kl}}}{\partial x_k} = 0, \\
 (12.8) \quad & \frac{\partial L_{d_i}}{\partial x_i} = R, \\
 (12.9) \quad & \frac{\partial L_{b_i}}{\partial x_i} = 0.
 \end{aligned} \tag{12}$$

We indicate the physical sense of the variables in the system (12). We assume that the equation of state is given in the form

$$E = E(V, S, c_{11}, c_{12}, \dots, c_{33}),$$

where E is the specific internal energy depending on the specific volume V , the entropy S , and the distortion tensor (the deformation gradient) c_{ij} . We assume that the state of a medium is also characterized by the velocity vectors u_i , the electric field d_i , and magnetic field b_i . In the system (12), for variables we take

$$q_\omega = T = E_S, \quad q_0 = E - S E_S - V E_V - c_{il} E_{c_{il}} - \frac{u_i u_i}{2},$$

$$u_1, u_2, u_3, r_{11} = E_{c_{11}}, r_{12} = E_{c_{12}}, \dots, r_{33} = E_{c_{33}}, d_1, d_2, d_3, b_1, b_2, b_3,$$

and for the generating potential we take the function

$$L = -EV + \frac{\varepsilon}{c} \frac{d_i d_i}{2} + \frac{\mu}{c} \frac{b_i b_i}{2} + \frac{\varepsilon \mu}{c^2} \begin{vmatrix} u_1 & b_1 & d_1 \\ u_2 & b_2 & d_2 \\ u_3 & b_3 & d_3 \end{vmatrix},$$

where c is the velocity of light, ε is the dielectric constant, and μ is the magnetic permeability.

Under this choice of physical parameters, equation (12.1) is the entropy balance law

$$\frac{\partial \rho S}{\partial t} + \frac{\partial \rho S u_k}{\partial x_k} = \frac{d_i J_i}{T} \quad (13)$$

and equation (12.2) is the mass conservation law

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho u_k}{\partial x_k} = 0.$$

Equation (12.3) is the momentum conservation law with ponderomotive forces, which will not be written here. We only note that this law can be easily obtained by using the formulas

$$\begin{aligned} L_{u_i} &= \rho u_i - \frac{\varepsilon \mu}{c^2} e_{i\alpha\beta} d_\alpha b_\beta, & L_{r_{il}} &= \rho c_{il}, \\ L_{d_i} &= \frac{\varepsilon}{c} d_i - \frac{\varepsilon \mu}{c^2} e_{i\alpha\beta} u_\alpha b_\beta, & L_{b_i} &= \frac{\mu}{c} b_i + \frac{\varepsilon \mu}{c^2} e_{i\alpha\beta} u_\alpha b_\beta. \end{aligned}$$

Equation (12.4) expresses the conservation law for the components of the distortion tensor

$$\frac{\partial \rho c_{il}}{\partial t} + \frac{\partial (\rho c_{il} u_k - \rho c_{kl} u_i)}{\partial x_k} = 0.$$

The pair of equations (12.5) and (12.6) is the Maxwell equations in moving medium (Landau and Lifshitz, 1982)

$$\begin{aligned} \frac{\partial D_i}{\partial t} + \frac{\partial (u_k D_i - u_i D_k - e_{ikl} b_l)}{\partial x_k} &= -(u_i R + J_i), \\ \frac{\partial B_i}{\partial t} + \frac{\partial (u_k B_i - u_i B_k + e_{ikl} d_l)}{\partial x_k} &= 0, \end{aligned}$$

where $D_i = \frac{\varepsilon}{c} \left(d_i - \frac{\mu}{c} e_{i\alpha\beta} u_\alpha b_\beta \right)$ is the electric induction vector, $B_i = \frac{\mu}{c} \left(b_i + \frac{\varepsilon}{c} e_{i\alpha\beta} u_\alpha d_\beta \right)$ is the magnetic induction vector, R is the volume electric charge, and J_i is the electric current vector.

To close this system, we can use the Ohm law

$$J_i = \sigma d_i,$$

where σ is the electric conductivity. Moreover, the stationary laws (12.7)–(12.9) take the form

$$\frac{\partial \rho c_{kl}}{\partial x_k} = 0, \quad \frac{\partial D_i}{\partial x_i} = R, \quad \frac{\partial B_i}{\partial x_i} = 0.$$

The differential connection (the second equation of the system (11)) takes the form

$$\frac{\partial R}{\partial t} + \frac{\partial (u_i R + J_i)}{\partial x_i} = 0$$

called the electric charge conservation law.

We also note that under the above choice of the form of the Ohm law, equation (13) for the entropy has the form

$$\frac{\partial \rho S}{\partial t} + \frac{\partial \rho S u_k}{\partial x_k} = \frac{\sigma d_i d_i}{T}$$

and the nonnegative ($\sigma > 0$) right-hand side is known as “Joule heat source.”

4.2. EQUATIONS OF MOTION OF MULTIPHASE MEDIA

The second example presents the conservative version of equations of motion of multiphase media with different velocities of motion of phases and interphase friction. For the sake of simplicity, we consider only two-phase case. Let the state of each phase is characterized by densities ρ_1, ρ_2 and velocities u_i^1, u_i^2 ($i = 1, 2, 3$).

For parameters of the state of two-phase medium we take the density $\rho = \rho_1 + \rho_2$, the average velocity $u_i = \frac{\rho_1}{\rho} u_i^1 + \frac{\rho_2}{\rho} u_i^2$, the concentration of the second phase $\delta = \frac{\rho_2}{\rho}$, the difference of the velocities of phases $w_i = u_i^2 - u_i^1$, and the entropy S . Assume that the equation of state has the form

$$E(V, S, \delta, w_1, w_2, w_3) = E^0(V, S, \delta) + \frac{1}{2} \delta(1 - \delta) w_i w_i$$

where E^0 is the potential energy of mixture, $V = \frac{1}{\rho}$ is the specific volume.

We describe the motion of this medium by the following subsystem of the general system (9):

$$(14.1) \quad \frac{\partial L_{q_\omega}}{\partial t} + \frac{\partial (u_k L)_{q_\omega}}{\partial x_k} = \frac{j_k \pi_k}{q_w},$$

$$(14.2) \quad \frac{\partial L_{q_0}}{\partial t} + \frac{\partial (u_k L)_{q_0}}{\partial x_k} = 0,$$

$$(14.3) \quad \frac{\partial L_{u_i}}{\partial t} + \frac{\partial [(u_k L)_{u_i} + j_k L_{j_i} - \delta_{ik} j_\alpha L_{j_\alpha}]}{\partial x_k} = 0$$

$$(14.4) \quad \frac{\partial L_n}{\partial t} + \frac{\partial(u_k L_n + j_k)}{\partial x_k} = 0, \quad (14)$$

$$(14.5) \quad \frac{\partial L_{j_k}}{\partial t} + \frac{\partial(u_\alpha L_{j_\alpha} + n)}{\partial x_k} = -(e_{k\alpha\beta} u_\alpha \omega_\beta + \pi_k),$$

$$(14.6) \quad \frac{\partial L_{j_k}}{\partial x_\alpha} - \frac{\partial L_{j_\alpha}}{\partial x_k} = -e_{k\alpha\beta} \omega_\beta$$

where for the potential L we take the function

$$L = -E_V + \rho w_k E_{w_k}$$

and for variables we take

$$q_\omega = T = E_S, \quad q_0 = E - S E_S - V E_V - \delta E_\delta - \frac{u_i u_i}{2},$$

$$u_1, u_2, u_3, \quad n = E_\delta, \quad j_k = \rho E_{w_k}.$$

The physical sense of equations of the system (14) is as follows:
equation (14.1) expresses the entropy balance law

$$\frac{\partial \rho S}{\partial t} + \frac{\partial \rho S u_k}{\partial x_k} = \frac{j_k \pi_k}{T}, \quad (15)$$

equation (14.2) is the mass conservation law

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho u_k}{\partial x_k} = 0, \quad (16)$$

equation (14.3) is the momentum conservation law for the mixture

$$\frac{\partial \rho u_i}{\partial t} + \frac{\partial(\rho u_i u_k - \delta_{ik} E_V + \rho w_i E_{w_k})}{\partial x_k} = 0, \quad (17)$$

equation (14.4) is the concentration conservation law for the second phase

$$\frac{\partial \rho \delta}{\partial t} + \frac{\partial(\rho \delta u_k + \rho E_{w_k})}{\partial x_k} = 0, \quad (18)$$

equation (14.5) is the balance law for the difference of velocities of phases

$$\frac{\partial w_k}{\partial t} + \frac{\partial(u_\alpha w_\alpha + E_\delta)}{\partial x_k} = -(e_{k\alpha\beta} u_\alpha \omega_\beta + \pi_k). \quad (19)$$

The stationary equation (14.6) is used to determine the vorticity ω_β :

$$\frac{\partial w_k}{\partial x_\alpha} - \frac{\partial w_\alpha}{\partial x_k} = -e_{k\alpha\beta} \omega_\beta. \quad (20)$$

The vorticity is subject to the following differential connection (the last equation of the system (11)):

$$\frac{\partial \omega_k}{\partial t} + \frac{\partial (u_i \omega_k - u_k \omega_i + e_{ki\mu} \pi_\mu)}{\partial x_i} = 0. \quad (21)$$

To close the system, it is necessary to indicate the dependence of the source of interphase friction π_k on the parameters of the state of a medium. In the simplest case, the dependence of on the interphase friction is expressed by the formula

$$\pi_k = \kappa j_k = \kappa \rho E_{w_k}, \quad (22)$$

where κ is the friction depending, in general, on the parameters of the state of a medium (e.g., density, temperature, etc.). Of course, the dependence of π_k on the parameters of a medium can be more complicated. We emphasize that any formula expressing such a dependence must provide the nonnegativity of entropy production in equation (15):

$$j_k \pi_k \geq 0$$

Passing to the variables ρ_1 , ρ_2 , u_i^1 , u_i^2 , and S in equations (15)–(20), we obtain the following system describing the motions of two-phase medium with interphase friction:

$$\begin{aligned} \frac{\partial \rho S}{\partial t} + \frac{\partial \rho S u_k}{\partial x_k} &= \frac{j_k \pi_k}{T}, \\ \frac{\partial (\rho_1 + \rho_2)}{\partial t} + \frac{\partial (\rho_1 u_k^1 + \rho_2 u_k^2)}{\partial x_k} &= 0, \\ \frac{\partial (\rho_1 u_i^1 + \rho_2 u_i^2)}{\partial t} + \frac{\partial (\rho_1 u_i^1 u_k^1 + \rho_2 u_i^2 u_k^2 + \delta_{ik} \rho^2 E_\rho^0)}{\partial x_k} &= 0 \quad (23) \\ \frac{\partial \rho_2}{\partial t} + \frac{\partial \rho_2 u_k^2}{\partial x_k} &= 0, \\ \frac{\partial (u_k^2 - u_k^1)}{\partial t} + \frac{\partial (u_\alpha^2 u_\alpha^2 - u_\alpha^1 u_\alpha^1 + E_\delta^0)}{\partial x_k} &= - \left(e_{k\alpha\beta} \frac{\rho_1 u_\alpha^1 + \rho_2 u_\alpha^2}{\rho_1 + \rho_2} \omega_\beta + \pi_k \right), \\ \frac{\partial (u_k^2 - u_k^1)}{\partial x_\alpha} - \frac{\partial (u_\alpha^2 - u_\alpha^1)}{\partial x_k} &= - e_{k\alpha\beta} \omega_\beta, \end{aligned}$$

where $E^0(V, \delta, S)$ is the potential energy of the mixture, $V = \frac{1}{\rho_1 + \rho_2}$ is the specific volume, $\delta = \frac{\rho_2}{\rho_1 + \rho_2}$ is the concentration of the second phase, and S is the entropy.

To close the system, we can use, for example, the interphase friction law (22)

$$\pi_k = \kappa \frac{\rho_1 \rho_2}{\rho_1 + \rho_2} (u_k^2 - u_k^1).$$

We note that the first four equations of the system (23) are well known in the mechanics of two-phase media (Landau, and Ye. M. Lifshitz, 1988).

The pair of the last equations are new governing equations with the respect to the difference of velocities of the motion of phases. Due to this choice, the equations of the constructed model of a two-phase medium form a conservative hyperbolic system. This model was successfully used in modeling shock-wave processes in gas/liquid mixtures (Resnyansky, Milton and Romensky, 1997).

A generalization of these equations to the case where the number of phases is larger than two is based on the same ideas. If the state K of the phases are characterized by densities $\rho_1, \rho_2, \dots, \rho_K$ and velocities $u_i^1, u_i^2, \dots, u_i^K$, then for variables of the state of the multiphase medium it is necessary to choose the density $\rho = \rho_1 + \rho_2 + \dots + \rho_K$, the average velocity $u_i = \frac{\rho_1}{\rho} u_i^1 + \frac{\rho_2}{\rho} u_i^2 + \dots + \frac{\rho_K}{\rho} u_i^K$, the concentrations of phases $\delta_2 = \rho_2/\rho, \dots, \delta_K = \rho_K/\rho$, the differences of velocities $w_i^2 = u_i^2 - u_i^1, \dots, w_i^K = u_i^K - u_i^1$, and the entropy S . Assume that the equation of state is given in the form

$$\begin{aligned} E(V, S, \delta_2, \dots, \delta_K, w_1^2, w_2^2, \dots, w_3^2, \dots, w_1^K, w_2^K, w_3^K) \\ = E^0(V, S, \delta_2, \dots, \delta_K) + \frac{1}{2} \delta_2 w_i^2 w_i^2 + \dots + \frac{1}{2} \delta_K w_i^K w_i^K \\ - \frac{1}{2} (\delta_2 w_i^2 + \dots + \delta_K w_i^K) (\delta_2 w_i^2 + \dots + \delta_K w_i^K). \end{aligned}$$

After that, the system of motion of the continuum K of phases can be taken in the form of the following subsystem of the system (9):

$$\begin{aligned} \frac{\partial L_{q_\omega}}{\partial t} + \frac{\partial (u_k L)_{q_\omega}}{\partial x_k} &= \frac{j_k^\gamma \pi_k^\gamma}{q_\omega}, \\ \frac{\partial L_{q_0}}{\partial t} + \frac{\partial (u_k L)_{q_0}}{\partial x_k} &= 0, \\ \frac{\partial L_{u_i}}{\partial t} + \frac{\partial [(u_k L)_{u_i} + j_k^\gamma L_{j_i^\gamma} - \delta_{ik} j_\alpha^\gamma L_{j_\alpha^\gamma}]}{\partial x_k} &= 0, \\ \frac{\partial L_{n^\gamma}}{\partial t} + \frac{\partial (u_k L_{n^\gamma} + j_k^\gamma)}{\partial x_k} &= 0, \quad \gamma = 2, \dots, K, \\ \frac{\partial L_{j_k^\gamma}}{\partial t} + \frac{\partial (u_\alpha L_{j_\alpha^\gamma} + n)}{\partial x_k} &= -(e_{k\alpha\beta} u_\alpha \omega_\beta^\gamma + \pi_k^\gamma), \quad \gamma = 2, \dots, K, \\ \frac{\partial L_{j_k^\gamma}}{\partial x_\alpha} - \frac{\partial L_{j_\alpha^\gamma}}{\partial x_k} &= -e_{k\alpha\beta} \omega_\beta^\gamma, \quad \gamma = 2, \dots, K. \end{aligned}$$

Acting as in the case of a two-phase medium, we choose the generating potential

$$L = -E_V - \rho w_k^\gamma E_{w_k^\gamma}$$

and variables

$$q_\omega = T = E_S, \quad q_0 = E - SE_S - VE_V - \delta_\gamma E_{\delta_\gamma} - \frac{u_i u_i}{2},$$

$$u_1, u_2, u_3, \quad n^\gamma = E_{\delta_\gamma}, \quad j_k^\gamma = \rho E_{w_k^\gamma},$$

where $\gamma = 2, \dots, K$ everywhere (we adopt the convention regarding summation with respect to related indices.)

Using the variables $\rho, u_i, \delta_\gamma, w_i^\gamma$, and S , we write the system of equations of motion of multi-phase medium in the same way as the system (15)–(20) but with index γ in the corresponding places:

$$\begin{aligned} \frac{\partial \rho S}{\partial t} + \frac{\partial \rho S u_k}{\partial x_k} &= \frac{\rho E_{w_k^\gamma} \pi_k^\gamma}{T}, \\ \frac{\partial \rho}{\partial t} + \frac{\partial \rho u_k}{\partial x_k} &= 0, \\ \frac{\partial \rho u_i}{\partial t} + \frac{\partial (\rho u_i u_k + \delta_{ik} \rho^2 E_\rho + \rho w_i^\gamma E_{w_k^\gamma})}{\partial x_k} &= 0, \\ \frac{\partial \rho \delta_\gamma}{\partial t} + \frac{\partial (\rho \delta_\gamma u_k + \rho E_{w_k^\gamma})}{\partial x_k} &= 0, \quad \gamma = 2, \dots, K, \\ \frac{\partial w_k^\gamma}{\partial t} + \frac{\partial (u_\alpha w_\alpha^\gamma + E_{\delta_\gamma})}{\partial x_k} &= -(e_{k\alpha\beta} u_\alpha \omega_\beta^\gamma + \pi_k^\gamma), \quad \gamma = 2, \dots, K, \\ \frac{\partial w_k^\gamma}{\partial x_\alpha} - \frac{\partial w_\alpha^\gamma}{\partial x_k} &= -e_{k\alpha\beta} \omega_\beta^\gamma, \quad \gamma = 2, \dots, K. \end{aligned} \tag{24}$$

Moreover, the vorticity vectors $\omega_1^\gamma, \omega_2^\gamma, \omega_3^\gamma$ ($\gamma = 2, \dots, K$) are subject to differential connections with sources of friction π_k :

$$\frac{\partial \omega_k^\gamma}{\partial t} + \frac{\partial (u_i \omega_k^\gamma - u_k \omega_i^\gamma + e_{ki\mu} \pi_\mu^\gamma)}{\partial x_i} = 0.$$

To close the system (24), it is necessary to define sources of interphase friction π_k^γ , for example, in the form

$$\pi_k^\gamma = \kappa^{\gamma\alpha} \rho E_{w_k^\alpha}, \quad \gamma, \alpha = 2, \dots, K.$$

References

- Godunov S K (1961). An Interesting Class of Quasilinear Systems. *Soviet Math. Dokl.* **2**, pp 947–949.
 Godunov S K (1986). Lois de Conservation et Integrales D'énergie des Equations Hyperboliques. in Carasso, Raviart, Serre (Editors). *Nonlinear Hyperbolic Problems*, Springer, pp 135–149 .

- Godunov S K and Romensky E I (1998). *Elements of Continuum Mechanics and Conservation Laws*. Nauchnaya Kniga, Novosibirsk.
- Ruggeri T and Müller I (1999). *Rational Extended Thermodynamics*. Springer.
- Godunov S K (1972). Symmetric Form of Magnetohydrodynamics Equations. *Numerical Methods of Continuum Mechanics (Novosibirsk, Russian)* **3**, no 1, pp 26–34.
- Friedrichs K O (1978). Conservation Laws and the Laws of Motion in Classical Physics. *Comm. Pure Appl. Math.* **31**, pp 23–131.
- Godunov S K and Romensky E I (1995). Thermodynamics, Conservation Laws, and Symmetric Forms of Differential Equations in Mechanics of Continuous Media. Hafez and Oshima (Editors). *Comput. Fluid Dynamics Review 95*, John Wiley & Sons, pp 13–19.
- Romensky E I (1998). Hyperbolic Systems of Thermodynamically Compatible Conservation Laws in Continuum Mechanics. *Math. Comput. Modelling*, **28**, pp 115–130.
- Godunov S K, Mikhailova T Yu and Romensky E I (1996). Systems of Thermodynamically Coordinated Laws of Conservation Invariant under Rotation. *Siberian Math. J.*, **37**, pp 690–705.
- Mikhailova T Yu (1997). Thermodynamically Consistent Conservation Laws with Unknowns of Arbitrary Weight, *Siberian Math. J.*, **38**, pp 528–538.
- Romensky E I (1995). Symmetric Form of the Equations of a Nonlinear Maxwell Medium. *Siberian Adv. Math.* **5**, pp 133–150.
- Landau L D and Lifshitz Ye M (1982). *Electrodynamics of Continuous Media*. Nauka, Moscow.
- Landau L D and Lifshitz Ye M (1988). *Fluid Mechanics*. Pergamon Press.
- Resnyansky A D, Milton B E and Romensky E I (1997). A Two-Phase Shock-Wave Model of Hypervelocity Liquid Jet Injection into Air. Proc. JSME Centennial Grand Congress, Int. Conf. on Fluid Engineering, Tokyo, Japan, pp 943–947

DEVELOPMENT AND APPLICATION OF HIGH-RESOLUTION ADAPTIVE NUMERICAL TECHNIQUES IN SHOCK WAVE RESEARCH CENTER

T. SAITO

Email: saito@bellanca.ifs.tohoku.ac.jp

P. VOINOVICH

Emails: vpeter@vpeter.ioffe.rssi.ru

vpeter@ceres.ifs.tohoku.ac.jp

vpeter@cri.univ-lille1.fr

voino@hermes.fnb.maschinenbau.tu-darmstadt.de

E. TIMOFEEV

Emails: timo@ceres.ifs.tohoku.ac.jp

eugene@sunphys.ioffe.rssi.ru

AND

K. TAKAYAMA

Email: takayama@ifs.tohoku.ac.jp

Shock Wave Research Center

Institute of Fluid Science, Tohoku University

2-1-1 Katahira, Aoba-ku, Sendai, 980-8577 Japan

Abstract.

A second-order accurate Godunov-type scheme has been implemented using 2-D and 3-D locally-adaptive unstructured grids. The fundamentals of the method, specifics of vector- and massively parallel processing, and some sample applications are presented in this paper.

1. Introduction

During the recent several years, a family of efficient numerical methods which combine locally-adaptive unstructured grid technology with second-order quasi-monotone finite-volume Godunov-type schemes in two and three

dimensions have been implemented and applied to various problems of shock dynamics at the Shock Wave Research Center (SWRC) of the Institute of Fluid Science (IFS), Tohoku University, Sendai, Japan. The variety of experimental facilities with modern registration and flow visualization equipment available at SWRC provide reliable quantitative data for verification of new computer codes, while powerful and permanently upgrading computer resources and post-processing tools of the Supercomputer Center at IFS form a good basis for large-scale CFD projects.

This paper intends outlining basic motives for the selected code development strategy, fundamentals of the scheme and grid adaptation techniques, essential aspects of data management including dynamic data structures and specifics of their operation in vector and massively-parallel supercomputer environment, as well as presenting a broad spectrum of applications ranging from basic research to safety analysis and engineering.

From a numerical point of view, different problems of shock wave dynamics involving shock reflection, diffraction, refraction, vorticity generation etc. have much in common, namely heterogeneous and rapidly changing wave patterns which typically contain singularities or sharp flow features like pressure and density jumps, slip surfaces and material interfaces whose interaction with one another and with the boundaries may drastically change the whole flow structure. This makes a shock-fitting or front-tracking method ineffective as a universal numerical tool in shock wave research. Our choice hence was made in favor of a shock-capturing scheme.

Because the total number of nodes in a grid is usually prescribed by the required resolution at discontinuities, which becomes more critical when the latter are numerous and/or a shock propagation has to be traced for a long distance, dynamic adaptation of the grid to the solution and high-accurate schemes might contribute considerably to the memory- and CPU-efficiency of the numerical method.

A range of shock-capturing schemes have been introduced recently utilizing unstructured grids. The latter become popular for their extreme flexibility to complex geometries and for non-fixed underlying data structures which naturally support local variations in the grid resolution by insertion or deletion of some grid nodes. The grid generation and adaptation methods are best developed for unstructured grids composed of triangles in 2-D and tetrahedra in 3-D.

Below we firstly present the 2-D version of our adaptive numerical technique with the respective sample applications. Essentials of the 3-D implementation are discussed thereafter together with the specifics of vector and massively-parallel processing, also followed by representative examples.

2. Governing equations in two dimensions

Though some efforts were made and results achieved in the development and application of adaptive Navier-Stokes solvers, we restrict this paper to the inviscid, non-heat-conducting Eulerian gas model. The mass, momentum and energy conservation laws written in an integral form for a two-species gas mixture underlie the mathematical model. For two dimensions the governing equations are:

$$\int_{\text{Vol}} U dV \Big|_{t_0}^t + \int_{t_0}^t dt \int_{\sigma} (F_x n_x ds + F_y n_y ds) = 0,$$

$$U = \begin{pmatrix} \rho \\ \rho_1 \\ \rho u_x \\ \rho u_y \\ \rho e \end{pmatrix}, \quad F_x = \begin{pmatrix} \rho u_x \\ \rho_1 u_x \\ \rho u_x^2 + p \\ \rho u_x u_y \\ (\rho e + p) u_x \end{pmatrix}, \quad F_y = \begin{pmatrix} \rho u_y \\ \rho_1 u_y \\ \rho u_x u_y \\ \rho u_y^2 + p \\ (\rho e + p) u_y \end{pmatrix},$$

where $\vec{u} = (u_x, u_y)$ – velocity vector, p – pressure, $\rho = \rho_1 + \rho_2$ – density of the gas mixture composed of species “1” and “2”, $e = \epsilon + 0.5u^2$ – specific total energy. The mixture is considered to be a perfect gas with the specific internal energy given by $\epsilon = (1/(\gamma - 1))p/\rho$, where γ is the polytropic index. Vol represents a gas volume bounded by the closed surface σ with the outward normal $\vec{n} = (n_x, n_y)$.

3. The unstructured finite-volume method

An incremental triangulation of the given computational domain results in a boundary-fitted unstructured grid composed of triangular area elements. The grid generation code Ref. (Galyukov and Voinovich, 1993) has been used in this work. The node-centered nonoverlapping control volumes are established about the grid nodes, i.e. vertices of grid triangles, where the dependent gas dynamic variables are given. The control volumes around each node are bounded by medians of triangular grid elements. So, each edge ij between nodes i and j is associated with vector area \vec{S}_{ij} .

Numerous FCT, TVD, TVB, ENO, flux-splitting and high-order Godunov schemes to compute fluxes through the areas \vec{S}_{ij} where comprehensively tested using 1-D and 2-D structured grids. It has been shown that a second-order Godunov-type scheme proposed by Rodionov (Rodionov, 1987) for steady-state supersonic flows and modified later for transient flows, see (Fursenko et. al., 1992), (Fursenko et. al., 1993a), (Fursenko et. al., 1993b), possesses better properties in view of the compromise between computational efficiency and accuracy for practical applications. It can also be easily written for arbitrarily shaped control volumes typical for unstructured grids.

The scheme represents a conservative predictor-corrector method providing second order temporal and spatial accuracy for smooth one-dimensional solutions.

To enhance spatial accuracy, a linear reconstruction of primitive gas dynamic variables $V = (\rho, \rho_1, u_x, u_y, p)^T$ is used within a control volume. The effective gradient is chosen as a minimum of the mean gradient $(\overline{\nabla V})_i$ and doubled gradients $(\nabla V)_e$ in the triangles e sharing node i using a minmod limiter:

$$\left(\frac{\partial V}{\partial \alpha}\right)_i = \text{minmod}_e \left[\left(\frac{\partial V}{\partial \alpha}\right)_i, 2\left(\frac{\partial V}{\partial \alpha}\right)_e \right], \quad \alpha = x, y.$$

The mean gradient at node i is determined as volume-weighted average over neighboring triangles e :

$$(\overline{\nabla V})_i = \frac{1}{\text{Vol}_i} \sum_e (\nabla V)_e \text{Vol}_e,$$

where Vol_e is the contribution to Vol_i by triangle e .

The predictor step employs no Riemann solver thus contributing to method's efficiency. For the control volume surrounding node i predictor solution \tilde{U}_i is obtained as follows:

$$\tilde{U}_i \text{Vol}_i = U_i^n \text{Vol}_i - \Delta t \sum_j \{ F_x (V_{ij}^{\text{in}}) n_x S_{ij} + F_y (V_{ij}^{\text{in}}) n_y S_{ij} \},$$

where V_{ij}^{in} – primitive gas dynamic variables at the inner side of the control volume surface:

$$V_{ij}^{\text{in}} = V_i + 0.5(\nabla V)_i \vec{ij}.$$

Subscript j denotes grid points surrounding node i ; superscript n – time level; Vol_i – value of control volume i ; S_{ij} – area of the interface separating volumes i and j ; $\vec{n} = (n_x, n_y)$ – unit vector of the outward normal to the surface of control volume i .

Then for each volume i and time level $(n+1)$ the corrector step is written as follows:

$$U_i^{n+1} \text{Vol}_i = U_i^n \text{Vol}_i - \Delta t \sum_j \{ F_x (W_{ij}) n_x S_{ij} + F_y (W_{ij}) n_y S_{ij} \},$$

where W_{ij} – primitive gas dynamic variables obtained from the Riemann problem solution for $\tilde{V}_{ij}^{\text{in}}$, $\tilde{V}_{ij}^{\text{out}}$:

$$\tilde{V}_{ij}^{\text{in}} = 0.5 \left(\tilde{V}_i + V_i + (\nabla V)_i \vec{ij} \right); \quad \tilde{V}_{ij}^{\text{out}} = 0.5 \left(\tilde{V}_j + V_j + (\nabla V)_j \vec{ji} \right);$$

\tilde{V}_i, \tilde{V}_j are the results of the predictor step. An exact solution of the 1-D Riemann problem at the interface is used.

It can be seen that although the predictor step is non-conservative, its results are used in a way ensuring conservative property of the scheme as a whole.

4. The local 2-D grid adaptation

A local transient adaptation of the grid to the solution peculiarities (shocks, contact and slip surfaces etc.) can provide a powerful tool for improvement of both the accuracy and efficiency.

An expression of the following form combining first- and second-order differences has been employed in this work for the refinement/coarsening sensor:

$$E_i = \max_e \left(\frac{|(\nabla U)_e - \overline{(\nabla U)}_i|}{|(\nabla U)_e| + |\overline{(\nabla U)}_i| + \epsilon |U_i| d^{-1}} \right),$$

where e denotes triangles sharing node i , d – characteristic scale of cell i (it's "diameter"), ϵ – filter coefficient. This criterion is consistent with the hybrid type non-oscillatory schemes, as extra nodes are added and resolution increases right there where scheme's accuracy is reduced by the limiter.

All the grid nodes are labeled according to E_i values. If $E_i > T_r$, all the triangles with vertex i become fractionized unless the highest allowed refinement level is achieved. If $E_i < T_c$, the node i is removed provided that it is not a node of the initial grid and that no topologically non-triangular elements appear in the resulting grid. The topological rules governing the refinement/coarsening procedure adopted in our developments are essentially the same as those introduced by Löhner (Löhner, 1987). The standard refinement pattern corresponds to cutting a "parent" triangle into 4 regular "child" triangles, which could later be subdivided further. An auxiliary subdivision mode producing two "irregular" child triangles from a parent one is used in transitional zones between higher and lower grained cells. These irregular triangles are never cut further; in case of a subsequent refinement, the parent triangle is first restored and then cut into four.

The refinement always overrides derefinement if a race condition occurs. The full (ρ) or partial (ρ_1) density values were normally used in practical computations in place of U in the sensor depending on the problem under study.

An original node-based data structure has been implemented in the computer code to support dynamic variations in the grid. Twelve words per node are used to store nodes' cross-references. No memory to store

“parent” or “child” cells is required. Triangles and their edges are described implicitly through the references of neighboring nodes. A list of actual grid nodes and vacant memory units is used to access grid-related data without any sorting or renumbering.

On a scalar hardware, the refinement/coarsening procedure being called at each time step typically takes less than 5% of the total CPU time used by the solver and contributes considerably (by more than an order of magnitude) to the CPU- and memory efficiency of the compound method as compared to a scheme using uniformly spaced fine grids.

5. Two-dimensional applications

Here we present a brief review of some applications of the above numerical technique to 2-D and axisymmetric problems. Three of them related to the shock wave interaction with material interfaces in gases are considered in more detail in the following subsections.

The shock wave reflection over wedges is well studied theoretically and experimentally and therefore represents a popular benchmark problem for shock-capturing CFD codes. Numerical results on this test using our 2-D adaptive Euler code can be found in Ref. (Galyukov et. al., 1997).

Shock wave focusing due to reflection at curvilinear concave surfaces was considered in (Babinsky et. al., 1998) and (Babinsky et. al., 1996) for cylindrical cavities and in (Sun and Takayama, 1996) for a circular reflector.

Diffraction of shock waves over a convex corner was extensively simulated and analyzed in (Sun and Takayama, 1997).

The transient gas motion and supersonic flow formation as a result of nozzle starting process was simulated and discussed in (Saito et. al., 1999).

The extreme resolution potentialities of the proposed method were applied to the analysis of an axisymmetric shock wave implosion in a specifically designed compartment in (Timofeev et. al., 1998a) and to the detailed study of regular-to-Mach reflection transition in (Timofeev et. al., 1999).

Attenuation of shock and acoustic waves represents a topical technical problem. Simulation of weak shock wave propagation along a channel with multiple damping cavities at the walls was performed in (Sasoh et. al., 1997) and (Sasoh et. al., 1998).

A modification to the numerical technique allowing internal moving boundaries within the main computational domain has been presented in (Sun et. al., 1997) with an application to shock-induced separation of particles from a solid wall.

Most of the cited studies involve both the numerical simulation and laboratory experiments complementing essentially each other in the way that a good agreement in their results confirms validity of the numerical data,

which in turn contain information hardly being obtained experimentally. To considerably facilitate quantitative comparisons with the experiment, a special post processing technology has been developed presenting CFD numerics in an experiment-like form, as for example interferograms or shadowgraphs, see for instance (Babinsky et. al., 1998) and (Babinsky et. al., 1996).

5.1. SHOCK WAVE REFRACTION OVER A PLANE INTERFACE

The oblique shock wave refraction over a planar interface separating two gases of different wave impedance has been extensively studied theoretically (Henderson, 1989), (Zeng and Takayama, 1996), experimentally (Abd-El-Fattah et. al., 1976), (Abd-El-Fattah and Henderson, 1978), (Zeng and Takayama, 1996), and numerically (Henderson et. al., 1991), (Zeng and Takayama, 1996). Here we demonstrate application of the locally adaptive numerical technique to this problem.

Despite the substantial simplicity of initial conditions and low number of parameters in the problem setting, the resulting flow structure and wave pattern may become rather complex for some cases of refraction, as it takes place by the slow-fast refraction at large angles of incidence, when a so-called irregular “twin Mach reflection type” refraction or “twin von Neumann” refraction occurs. A computed result in the form of density isolines for this type of refraction is given in Figure 1. We reproduced initial parameters of the problem from (Zeng and Takayama, 1996): the ambient or incident gas is air, the receiving gas is He, the incident shock wave Mach number 1.4, and the angle of incidence 77° . To accomplish the analogy with the experiment, a layer of dense gas was introduced between air and He simulating the thin plastic film which initially separates two gases preventing them from mixing and forming a planar interface before the incident shock arrives and ruptures it. An excellent correlation with the experiment was found when comparing the computed flow pattern with the experimental one obtained using a double exposure holographic interferometry method (Zeng and Takayama, 1996). A high resolution of material interfaces, even essentially distorted by the shock wave refraction, is clearly seen in the figure, as a result of the local grid refinement technique used in the computation.

5.2. INTERACTION OF A SHOCK WAVE WITH A HELIUM BUBBLE

The problem of interaction of a shock wave with isolated gaseous inhomogeneities became of considerable interest as a possible building block of more complex phenomena accompanying shock-induced mixing which occurs in various scientific and technological applications, see e.g. (Haas and Sturte-

vant, 1987). Here we present an instantaneous computed flow field for a $M_s = 1.2$ shock wave interaction with a cylindrical He bubble in air. As has been stated in (Picone and Boris, 1988), the problem is a very difficult one from a numerical point of view, as it involves non-steady compressible flows with shocks which generate non-steady, complex rotational flows embedded in a compressible medium. The presence of two fluids of different density separated by sharp interface (bubble boundary) represents a further complication.

As can be seen from Figure 2, the described adaptive numerical technique treats the problem adequately ensuring a very high resolution of both the shock waves and material interfaces within the shock-capturing environment. An early phase of the Richtmyer-Meshkov instability can be observed at the downstream (right-hand) side of the deformed He bubble, which is in a good qualitative correlation with the experiment (Haas and Sturtevant, 1987). A quantitative analysis of this instability and comparison against the experiment should involve more careful alignment of computational and experimental parameters.

5.3. SIMULATION OF A RICHTMYER-MESHKOV INSTABILITY

The final example of application of the developed adaptive numerical technique to transient problems with strongly perturbed material interfaces in gases to be presented in this paper is a developed stage of the Richtmyer-Meshkov (RM) instability caused by the interaction of a primarily converging shock in air with an axisymmetric helium-filled cavity. The general flow development can be described as follows. The incident cylindrical converging shock wave approaches the initial perfectly axisymmetric helium bubble from exterior. A moderate angular variation in the shock intensity of mode four (4 wave lengths per 360°) was imposed to simulate conditions of the experimental facility used by Takayama et al. employing a horizontal co-axial annular shock tube with the strut-supported inner part (Takayama et. al., 1987).

The first interaction of the converging shock with the bubble boundary occurs in the slow-fast direction and forms a transmitted converging shock wave in He and a diverging expansion wave in ambient air. The material interface is accelerated toward the symmetry line and a minor RM instability starts developing on it, though at a rather slow rate. The transmitted cylindrical shock wave collapses at the center and gives rise to a reflected diverging shock which propagates in the helium volume outwards. This first diverging shock interacts with the interface in the fast-slow direction, which is much more sensitive to the RM instability, as has been already demonstrated in the previous section. It is this second shock-interface in-

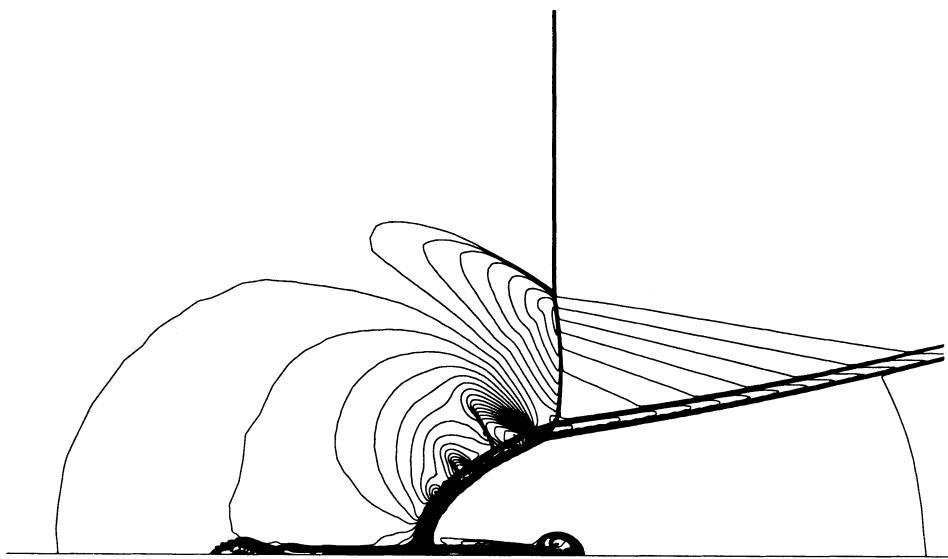


Figure 1. Shock wave $M_s = 1.4$ refraction over a plane Air/He interface. A thin layer of dense gas simulates the plastic film used in the experiments to maintain initial interface.

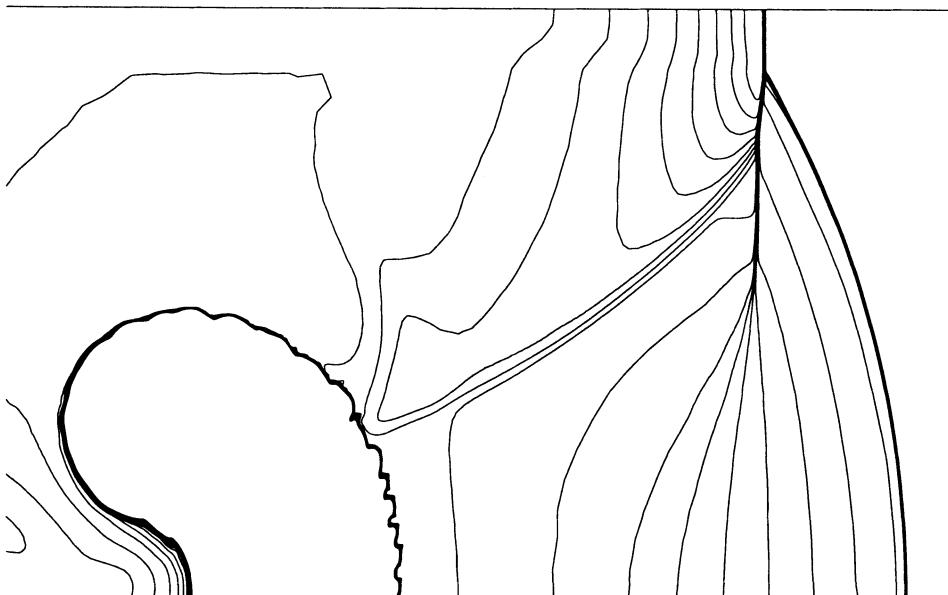


Figure 2. $M_s = 1.2$ shock wave interaction with a cylindrical He bubble in air.

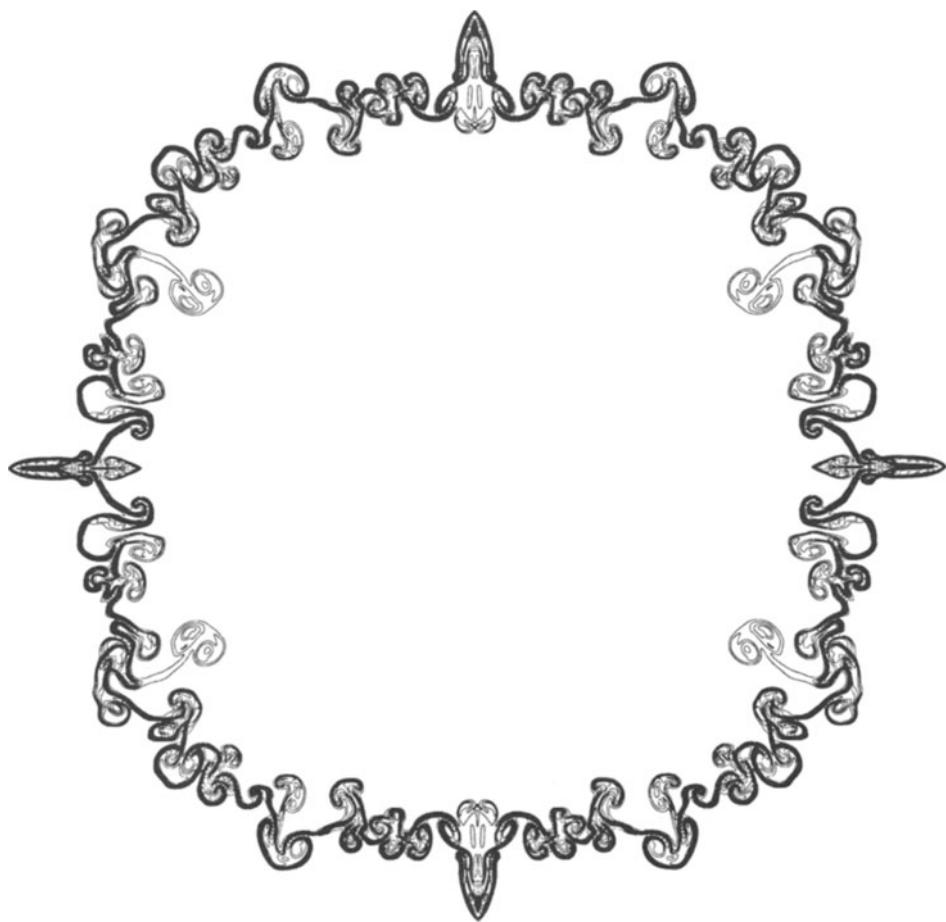


Figure 3. A developed stage of Richtmyer-Meshkov instability induced by the interaction of a converging cylindrical shock wave with an axisymmetric helium-filled cavity.

teraction which dramatically amplifies the development rate of instability generated by the first slow-fast interaction. A transmitted diverging shock wave leaves the interaction zone outwards while a secondary converging shock is reflected inwards to collapse and reflect again at the center.

By the moment when the second diverging shock approaches the interface, the latter becomes highly disturbed so that its shape is far from axisymmetric. For that reason the second diverging shock passes through the interface outwards without creating a new inward-going symmetric shock, so no further shocks collapse and reflect at the center. Hence, two reflected shock waves can be expected for registration in the ambient air after the primary shock wave implodes at the center of the helium-filled cavity.

Figure 3 presents a developed stage of the RM instability observed in this flow. The averaged square-shaped deformation of the interface is due to the above-mentioned mode four variation in the primary shock intensity. The main instability mode obtained in our computations corresponds to the grid spacing in the initial non-adapted grid, which means that either a slight non-uniformity in the initial conditions for the material interface was preserved despite the fine local enrichment in the grid nodes, or a minor anisotropy is exhibited by the scheme so that the regular grid patterns changing from one largest parent triangle to another could trigger the observed mode of instability. Further analysis into the problem is under way now to clarify the sources and reduce effects of method-dependent instability.

6. The 3-D numerical method

6.1. THE EULER SOLVER

The numerical scheme used in the 3-D solver, see Ref. (Timofeev et. al., 1997a), is a straightforward extension of one described in section 2 above. For the three dimensional grid composed of tetrahedra, the control volumes about the grid nodes (tetrahedra's vertices) are confined and separated from one another by triangular area elements built up within every tetrahedron, with the exception of the external sides of control volumes at the domain boundary which coincide with the respective boundary sides of tetrahedra. The separating triangular elements within a tetrahedron are formed by planes passing through the tetrahedron's centroid, the centroid of a tetrahedron's side, and the center of an edge, thus forming 12 elements in each tetrahedron so that two such elements contribute to one interface of six in 6 pairs of neighboring nodes involved in the tetrahedron. The rest of the algorithm fully corresponds to its two-dimensional counterpart.

6.2. GRID ADAPTATION IN 3-D

The grid refinement strategy employed here is similar to that introduced by Löhner (Löhner, 1989). The regular refinement mode corresponds to the subdivision of a tetrahedron into 8 child tetrahedra (all the parent tetrahedron's edges and faces split). Two extra auxiliary subdivision types are allowed splitting one tetrahedron into four (only one face split) and into two (only one edge split) smaller tetrahedra to smoothly connect grid regions with different number of successive subdivisions. All the refinement cases are fully reversible; the regular child tetrahedra only may undergo further refinement. The grid adaptation procedure is invoked at every time step rearranging the grid under control of the adaptation criteria and a sensor based on a combination of first and second derivatives along the grid edge. The tetrahedra are typically subdivided and new nodes inserted in the vicinity of solution singularities like shock fronts and contact surfaces improving cardinally the local resolution and overall accuracy. Some previously inserted child elements are removed restoring a more coarse and finally the original grid where the solution becomes sufficiently smooth, thus dramatically reducing the number of nodes in the adaptive grid.

6.3. THE DATA STRUCTURES

A linear ordering of neighbor nodes about a given node is extensively used in data handling in the 2-D code. As no linear ordering of randomly placed nodes is possible in 3-D, the node-based data structure becomes inefficient in this case. A tetrahedra-based homogeneous data structure maintains the grid connectivity and its dynamic variations. Tetrahedra refer to their vertices (grid nodes) and edges; the latter refer to the nodes they connect; the child grid elements (tetrahedra and edges) refer to their respective parents. This reference system is slightly redundant compromising between memory and CPU efficiency. No hierarchical tree structure is used in the code to handle the grid objects. All the grid elements, including tetrahedra, edges and nodes, are addressed (but not referenced) indirectly through dynamic address tables which support variations in the grid. The data associated with the grid (coordinates of grid nodes, gasdynamic variables, etc.) can be stored sparsely, while the address tables are always compact, i.e. contain, without gaps, randomly ordered addresses of actual grid objects, which are followed by addresses of vacant memory units to be used for the new elements emerging due to refinement.

7. Supercomputer implementation

The large-scale 3-D simulations require extensive computer resources commonly associated with supercomputers. Two alternative supercomputer platforms are currently available at IFS representing major branches in high-performance computer architectures: a vector-parallel system and a massively parallel one. A high level of automation in code optimization for execution on these systems is provided by the respective pre-processors and compilers. Some features of the algorithm however may dramatically influence the resulting efficiency of the code.

7.1. VECTOR PROCESSING

Two essentially independent modules execute every time step of the numerical method: the unstructured Euler solver and the grid adaptation procedure, which require comparable CPU time mildly dominated by the solver on a single-processor scalar hardware. Recurrent updates of indirectly addressed data within a loop are regularly used in the solver inhibiting vectorization of the respective loops. To allow vectorization, the recurrently treated objects are combined into groups with only single update within a group, so that a sub-loop over a group becomes vectorizable. Vectorization of the adaptation procedure is limited to vector processing of the innermost loops searching the data arrays. Profiling the vectorized code on CRAY C90 in real computations indicates that both the computation modules are again comparable in their CPU usage with the same moderate dominance by the solver.

7.2. MASSIVELY PARALLEL PROCESSING

Since the massively parallel processing (MPP) is widely recognized as the main direction in future supercomputing, a parallel version of the code has been developed targeting the CRAY T3D MPP system, which is a MIMD architecture with distributed memory. The key point of the adopted parallelization strategy is achieving of the best possible load balance by uniform distribution of the address tables among the processors. The approach turned out to be successful due to the very fast interprocessor exchange in CRAY T3D, which fundamentally reduces importance of data locality. The objects' grouping is also employed to prevent erroneous recurrent updates. A master-slave technique has been implemented to perform the essentially serial grouping procedure concurrently with parallel data processing. More details concerning vector and parallel implementations of the code are given in (Voinovich et. al., 1998).

8. Three-dimensional applications

8.1. A 3-D SHOCK-CYLINDER INTERACTION

The results of numerical and experimental analysis of strongly unsteady 3-D wave patterns by the interaction of an initially planar incident shock wave with an oblique cylinder are discussed in (Timofeev et. al., 1997a) and (Timofeev et. al., 1997b). The subject is mostly of theoretical interest as a comparatively simple example of truly three-dimensional shock wave diffraction and transition from regular to Mach reflection. The problem setting can be seen from Figure 4 presenting a developed wave structure for the incident shock wave intensity $M_s = 2.8$ and cylinder inclination angle $\theta = 60^\circ$.

Specifics of the problem from a numerical point of view is that sufficiently high resolution is needed to resolve details of the wave structure and determine the transition location with reasonable accuracy. The trouble stems from the fact that the triple line and its 3-D trajectory are tangent to the cylinder surface. As the Mach stem has to be at least one grid cell high to manifest itself, simple considerations result in the following relations between the desired linear l and angular α accuracy and grid spacing δ relative to the cylinder radius R : $\delta/R \sim l^2/R^2 \sim \sin^2\alpha$. Thus, for instance, $\delta \approx 0.03R$ for $\alpha = 10^\circ$ and $\delta \approx 0.001R$ for $\alpha = 2^\circ$. In practice, more than one grid cell is needed to detect the Mach stem undoubtedly, and the above estimates become even more severe. As the specified accuracy is only needed at the shock waves, the adaptive grid refinement technique orders of magnitude reduces the required amount of grid nodes. In our computations we achieved an accuracy of $\alpha \approx 5^\circ$ with the largest grid size of less than 500,000 nodes which is approximately equivalent to $\sim 10,000,000$ nodes by uniform gridding. An instant adaptive grid pattern is shown in Fig. 4, bottom.

8.2. BLAST WAVES OVER TERRAINS

A series of problems related to blast wave propagation over realistic terrains has been simulated recently, see (Voinovich et. al., 1998), (Saito et. al., 1997), (Timofeev et. al., 1998b). The complex geometry of real topographies and man-made structures stimulates application of unstructured grids for the numerical analysis. The high resolution critically needed only at the blast wave front and reflected shock waves close to the ground surface and structures' walls is maintained by the local grid refinement at a restricted amount of total nodes in the grid.

An example of the numerical simulation of blast waves induced by an explosive volcanic eruption in the geometry of Mt. Aso Nakadake volcano,

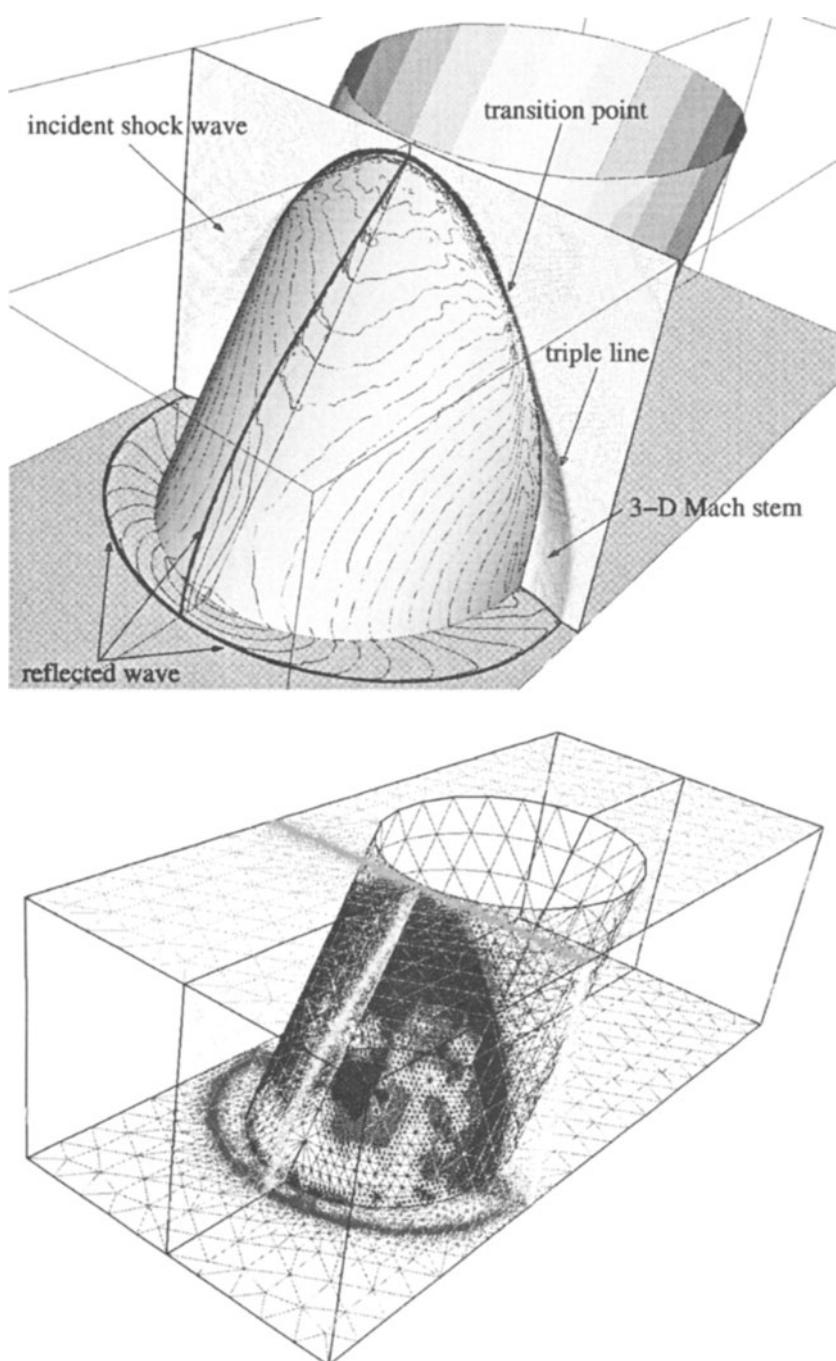


Figure 4. An instant of 3-D shock-cylinder interaction: a pressure iso-surface and density contour lines (top), the respective grid pattern (bottom).

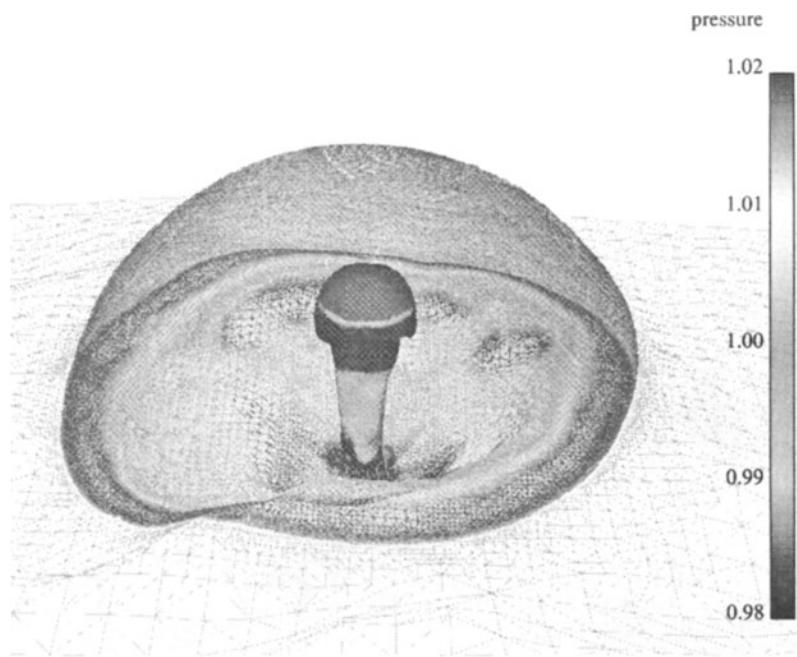


Figure 5. Eruption-induced blast wave over a volcanic terrain.

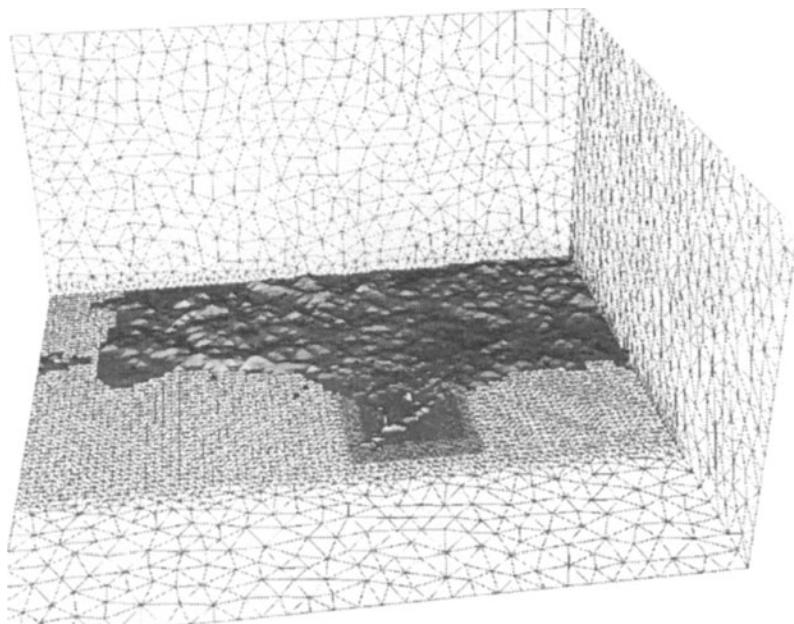


Figure 6. Domain for a blast simulation: general view.

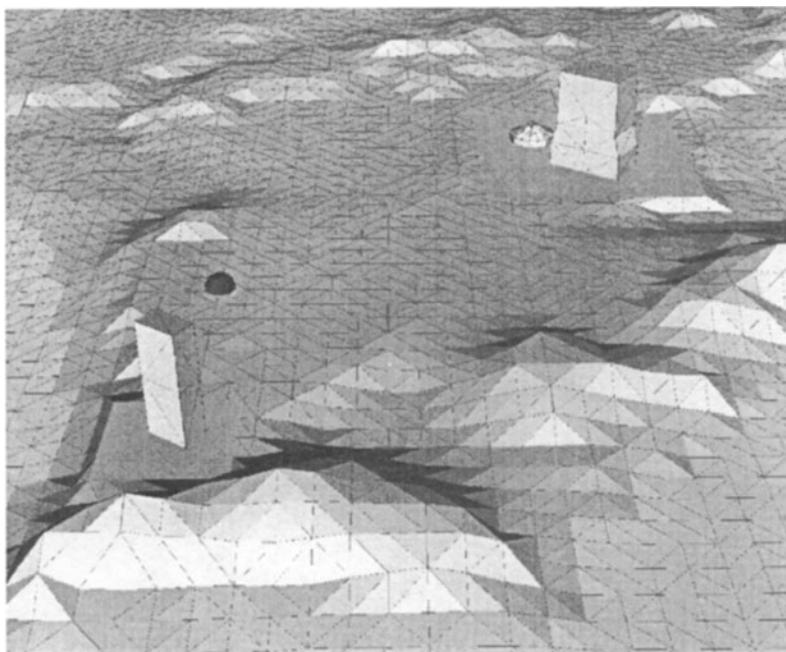


Figure 7. Domain for a blast simulation: a close-up of the central part.

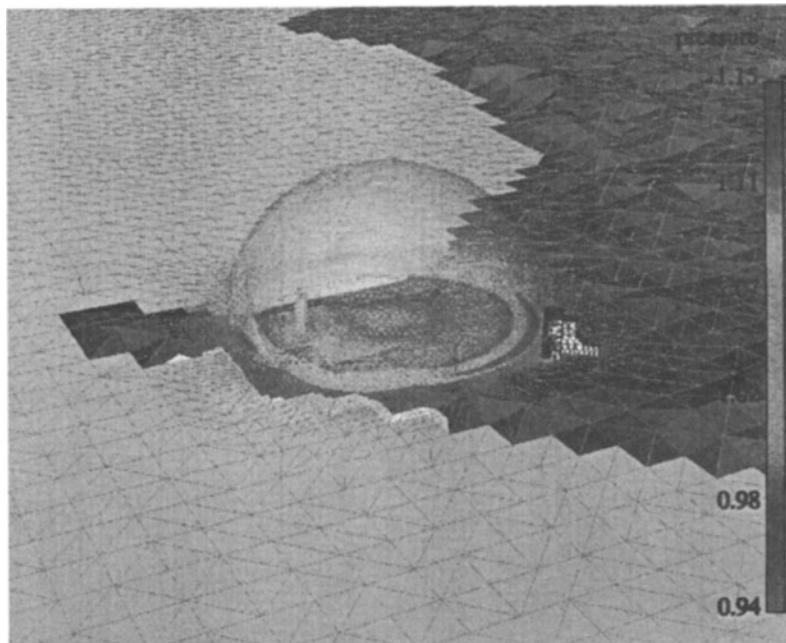


Figure 8. Simulated blast wave.

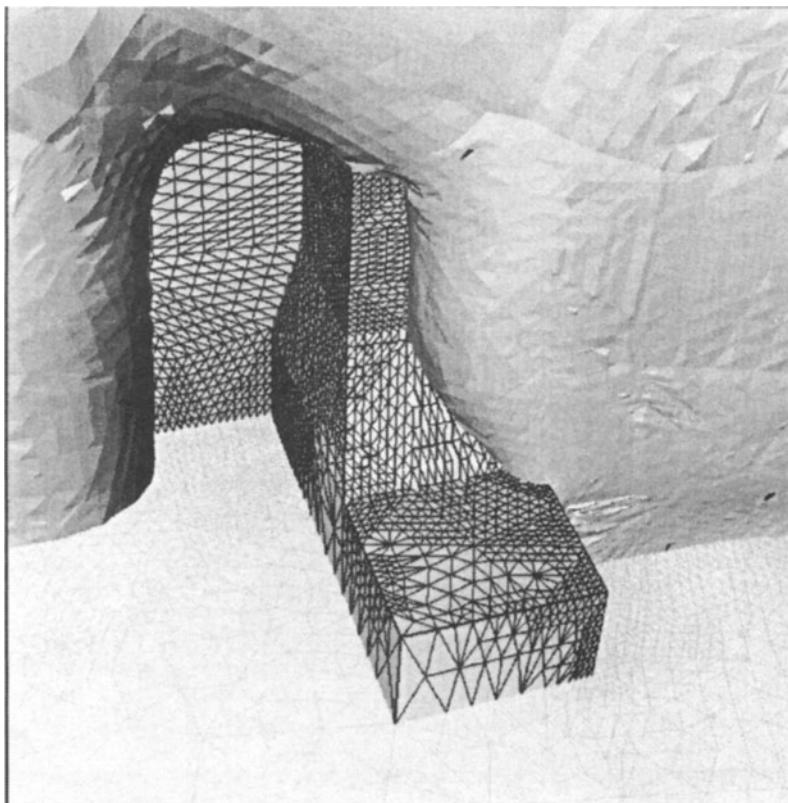


Figure 9. Diffraction of the blast wave over a building.

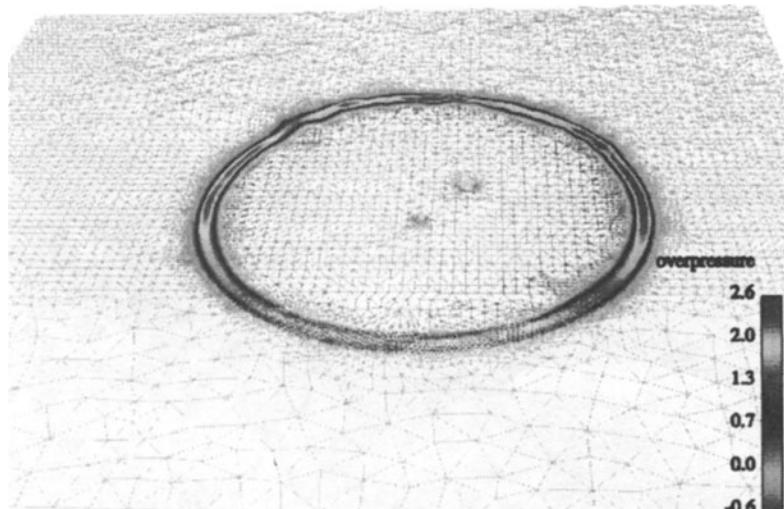


Figure 10. Long-distance blast wave propagation.

Kyushu, Japan are given in Figure 5. The blast wave, high-temperature jet rising up from the volcano mouth, and the grid pattern at the surface can be seen in the Figure. The simulations revealed noticeable qualitative distinctions in pressure histories at specified locations depending on different models of energy release by the eruption, which were suggested in preparation for the future field measurements.

Numerous simulations have been performed in the interests of safety analysis at the launch site of NASDA Space Center, Tanegashima, Japan. Some of the results are discussed in (Timofeev et. al., 1998b). The below figures depict most recent simulations (Voinovich et. al., 1999). Figure 6 presents a general view of the computational domain while an enlargement of the vicinity of the launch site is given in Figure 7 (the dark hemisphere indicates the location of an imaginary explosion). Some details of the geometry can also be seen from Figure 8 with an instant image of the simulated blast wave. In Figure 9 a small fragment of the domain is given presenting the interaction of the blast wave with a construction unit. Even though the characteristic length scale of the unit is approximately 1/50 of the distance to the explosion center and approximately 1/600 of the domain size, the blast wave appears sharply resolved and specific details of the wave interactions can be observed clearly.

In some computations, a long range propagation of the blast wave was analyzed. The overpressure becomes very low at a large distance (a few percent of the atmospheric pressure) and even minor disturbances caused by the blast wave interaction with the outer domain boundaries may seriously influence the results. The advantage of adaptive grids in this specific relation is that one can use computational domains large enough to prevent the blast wave from reaching any outward boundary during the observation time without considerable increase in the nodes number, as the grid refinement can be simply deactivated in the regions out of interest and/or coarser background meshes can be used there (see Fig. 10).

The grids for the blast wave simulations, as well as for the shock-cylinder interactions presented in the previous subsection were created by the grid generation software (Galyukov and Voinovich, 1994).

8.3. AN ENGINEERING APPLICATION IN 3-D

An application of the developed techniques and codes to a three-dimensional engineering problem is presented in (Voinovich et. al., 1999). The gas flow under consideration is a strongly transient one which occurs in the intake and exhaust manifolds of automotive engines. A data conversion utility has been coded to transform the geometry and unstructured grid data generated by the software embedded into the CAD system in use to data formats

accepted by our unstructured solvers. To further increase portability of the 3-D unstructured CFD software making it essentially independent of commercial post-processors, a set of post-processing routines has been developed which can run on any computer system presenting computed results in a form of color PostScript files. The respective images are not displayed here because most of the visual information is lost being rendered in a gray-scale mode.

9. Conclusion

Our experience accumulated during massive application of the 2-D and 3-D locally-adaptive shock-capturing methods based on unstructured grids to various problems of shock wave dynamics results in a very positive estimation of major principles underlying the developed techniques and codes, though a labor-intensive period of coding and debugging preceded.

The productive idea of using an exact solution of the Riemann problem to compute fluxes between control volumes within the shock-capturing approach introduced by S.K.Godunov 30 years ago works well in the unstructured grid environment contributing to robustness of modern CFD codes.

Acknowledgements

The authors would like to thank Alexander Galyukov of SoftImpact Ltd., St.Petersburg, Russia for his crucial assistance in generation of 2-D and 3-D unstructured grids. The safety analysis project presented in subsection 8.2 was supported by NASDA, project coordinator Mr.Y.Hyodo.

References

- Abd-El-Fattah AM, Henderson LF, Lozzi A (1976). Precursor shock waves at a slow/fast gas interface. *J. Fluid Mech.* **76**, pp 157-176.
- Abd-El-Fattah AM, Henderson LF (1978). Shock waves at a fast-slow gas interface. *J. Fluid Mech.* **86**, pp 15-32.
- Babinsky H, Onodera O, Takayama K, Timofeev E, Voinovich P (1996). The influence of geometric variations on the shock focusing in cylindrical cavities. In: Shock Waves, Proc. of the 20th Int. Symp. on Shock Waves (Pasadena, USA, 23-28 July 1995), Sturtevant B, Shepherd JE, Hornung HG (Editors), World Scientific Publishing, **1**, pp 495-500.
- Babinsky H, Onodera O, Takayama K, Saito T, Voinovich P, Timofeev E (1998). The influence of entrance geometry of circular reflectors on the shock wave focusing. *Computers and Fluids* **27**, No. 5-6, pp 611-618.
- Fursenko AA, Sharov DM, Timofeev EV, Voinovich PA (1992). Numerical simulation of shock wave interactions with channel bends and gas nonuniformities. *Computers and Fluids* **21**, pp 377-396.

- Fursenko AA, Mende NP, Oshima K, Sharov DM, Timofeev EV, Voinovich PA (1993a). Numerical simulation of propagation of shock waves through channel bends. *Computational Fluid Dynamics Journal* **2**, pp 1-36.
- Fursenko AA, Sharov DM, Timofeev EV, Voinovich PA (1993b). High-resolution schemes and unstructured grids in transient shocked flow simulation. *Lecture Notes in Physics* **414**, pp 250-254.
- Galyukov AO, Voinovich PA (1993). Two-dimensional triangular unstructured grid generator. Advanced Technology Center, St. Petersburg, Russia (unpublished).
- Galyukov A, Voinovich P (1994). Three-dimensional unstructured tetrahedral grid generator. Advanced Technology Center, St. Petersburg, Russia (unpublished).
- Galyukov A, Voinovich P, Timofeev EV (1997). In: Shock wave reflection over wedges: a benchmark test for CFD and experiments. *Shock Waves* **7**, No. 4, pp 196-199.
- Haas J-F, Sturtevant B (1987). Interaction of weak shock waves with cylindrical and spherical gas inhomogeneities. *J. Fluid Mech.* **181**, pp 41-76.
- Henderson LF (1989). On the refraction of shock waves. *J. Fluid Mech.* **198**, pp 365-386.
- Henderson LF, Colella P, Puckett EG (1991). On the refraction of shock waves at a slow-fast gas interface. *J. Fluid Mech.* **224**, pp 1-27.
- Löhner R (1987). The efficient simulation of strongly unsteady flows by the finite element method. *AIAA Paper* 87-0555.
- Löhner R (1989). Adaptive h-refinement on 3-D unstructured grids for transient problems. *AIAA Paper* 89-0365.
- Picone JM, Boris JP (1988). Vorticity generation by shock propagation through bubbles in a gas. *J. Fluid Mech.* **189**, pp 23-51.
- Rodionov AV (1987). Improvement of the approximation order in the Godunov scheme. *Zh. Vychisl. Mat. Mat. Fiz.* **27**, pp 1853-1860 (in Russian).
- Saito T, Voinovich P, Timofeev E, Hayakawa S, Onodera O, Takayama K (1997). Propagation of blast waves induced by volcano eruptions: new experimental and numerical tools. In: Shock Waves, Proc. of the 21st Int. Symp. on Shock Waves (Great Keppel Island, Australia, 20-25 July 1997), Houwing AFP (Editor-in-Chief), Panther Publishing and Printing, **2**, pp 1431-1435.
- Saito T, Timofeev E, Sun M, Takayama K (1999). Numerical and experimental study of 2-D nozzle starting process. Proceedings of the 22th Int. Symp. on Shock Waves, Imperial College, London, 1999 (to be published).
- Sasoh A, Matsuoka K, Nakashio K, Timofeev E, Takayama K, Saito T (1997). Attenuation of weak shock waves along pseudo-perforated walls. In: Shock Waves, Proc. of the 21st Int. Symp. on Shock Waves (Great Keppel Island, Australia, 20-25 July 1997), Houwing AFP (Editor-in-Chief), Panther Publishing and Printing, **2**, pp 749-754.
- Sasoh A, Matsuoka K, Nakashio K, Timofeev E, Takayama K, Voinovich P, Saito T, Hirano S, Ono S, Makino Y (1998). Attenuation of weak shock waves along pseudo-perforated walls. *Shock Waves* **8**, No 3, pp 149-159.
- Sun M, Takayama K (1996). A holographic interferometric study of shock wave focusing in a circular reflector. *Shock Waves* **6**, pp 323-336.
- Sun M, Takayama K (1997). The formation of a secondary shock wave behind a shock wave diffracting at a convex corner. *Shock Waves* **7**, pp 287-295.
- Sun M, Takayama K, Timofeev E, Voinovich PA (1997). Numerical simulation of the aerodynamic shock-cylinder interaction. In: Shock Waves, Proc. of the 21st Int. Symp. on Shock Waves (Great Keppel Island, Australia, 20-25 July 1997), Houwing AFP (Editor-in-Chief), Panther Publishing and Printing, **2**, pp 1481-1486.
- Takayama K, Kleine H, Groenig H (1987). An experimental investigation of the stability of converging cylindrical shock waves in air. *Exp. Fluids* **5**, pp 315-322.
- Timofeev E, Takayama K, Voinovich P (1997a). Numerical and experimental observation of three-dimensional unsteady shock wave structures. *AIAA Paper* 97-0070.

- Timofeev E, Takayama K, Voinovich P, Sisljan J, Saito T (1997b). Numerical and experimental study of three-dimensional unsteady shock wave interaction with an oblique cylinder. In: Shock Waves, Proc. of the 21st Int. Symp. on Shock Waves (Great Keppel Island, Australia, 20–25 July 1997), Houwing AFP (Editor-in-Chief), Panther Publishing and Printing, **2**, pp 1487-1492.
- Timofeev E, Sokolov I, Voinovich P, Saito T, Takayama K (1998a). Numerical simulation of an intense shock wave implosion in an axisymmetric compartment. *Review of High Pressure Science and Technology* **7**, pp 909-911.
- Timofeev E, Voinovich P, Takayama K, Saito T, Hyodo Y, Nakamura H (1998b). Adaptive unstructured simulation of three-dimensional blast waves with ground surface effect. *AIAA Paper* 98-0544.
- Timofeev E, Skews BW, Voinovich PA, Takayama K (1999). The influence of unsteadiness and three-dimensionality on regular-to Mach reflection transitions: a high-resolution study. Proceedings of the 22th Int. Symp. on Shock Waves, Imperial College, London, 1999 (to be published).
- Voinovich P, Timofeev E, Takayama K, Saito T, Galyukov A (1998). 3-D unstructured adaptive supercomputing for transient problems of volcanic blast waves. *AIAA Paper* 98-0540.
- Voinovich PA, Timofeev EV, Saito T, Takayama K, Hyodo Y, Galyukov AO (1999). An adaptive shock-capturing method in real 3-D applications, Proceedings of the 22th Int. Symp. on Shock Waves, Imperial College, London, 1999 (to be published).
- Zeng S, Takayama K (1996). On the refraction of shock wave over a slow-fast gas interface. *Acta Astronautica* **38**, No. 11, pp 829-838.

INTERFACES, DETONATION WAVES, CAVITATION AND THE MULTIPHASE GODUNOV METHOD

RICHARD SAUREL

Institut Universitaire des Systèmes Thermiques Industriels

5 rue Enrico Fermi

13453 Marseille Cedex 13

FRANCE

richard@iusti.univ-mrs.fr

Abstract. We have recently proposed a compressible two-phase unconditionally hyperbolic model able to deal with a wide range of applications: interfaces between compressible materials, shock waves in condensed multiphase mixtures, homogeneous two-phase flows (bubbly and droplet flows) and cavitation in liquids. Here we generalise the formulation to an arbitrary number of fluids, with mass and energy transfers, and the associated Godunov method is extended to multidimensions. This is necessary for modelling interaction of detonation waves in multiphase mixtures with interfaces separating the energetic material from inert ones. Thus, the model is able to solve detonation problems without mixture equation of state and dynamic interface creation for cavitating flows in multidimensions.

1. Introduction

Difficult problems arise in the modelling of flows involving mixtures. These mixtures may have a physical origin as in classical multiphase flows or may be due to numerical inaccuracies and artificial mixing. This occurs in the computation of interfaces separating two compressible fluids of different physical properties.

In this introduction these difficulties are first reviewed for flows involving fluid interfaces, secondly for flows involving homogeneous mixtures, and finally for flows where interfaces appear naturally: cavitating flows. The solution strategy is then developed and is valid for any of these applications.

Short review of methods for compressible flows with interfaces

Compressible multifluid flows occur in many situations when fluids have different physical or thermodynamical properties and are separated by interfaces. Well-known examples are the Richtmyer Meshkov instabilities in gas dynamics, the behaviour of a gas bubble in a liquid under shock wave etc.

Usually, the mathematical model in these situations is rather simple: it is based on compressible Euler or Navier Stokes equations. All the difficulty does not rely on the model but on the solution procedure and numerical method. Indeed, the various materials are governed by different equations of state (EOS) and artificial mixing occurs at the interfaces when an Eulerian scheme is used. This mixing is due to numerical diffusion of contact discontinuities. In this artificial mixture, computation of the pressure, the temperature, the sound speed is so wrong that methods fail at the second time step (negative pressure). Also, mixture rules and mixture equations of state based on physical arguments also fail because the mixing is purely numerical. Another reason is that most of the equations of state for liquids and solids (even for gases (Abgrall, 1996)) have limited domain of validity. The interface is the physical location where flow parameters are close to the limits of validity. Hence a careful and clean treatment of interfaces is mandatory.

Basically, two classes of methods are able to solve more or less accurately interface problems. The **first class** corresponds to methods that eliminate the numerical diffusion at the interfaces. Consequently, the artificial mixing problem is eliminated too. A short review about these methods is given in (Saurel and Abgrall, 2000). These methods can be listed as :

- Lagrangian methods (Benson, 1992),
- Arbitrary Lagrangian - Eulerian (ALE) methods (Farhat and Roux, 1991),
- Front Tracking methods (Harten and Hyman, 1983),(Glimm *et al.*, 1998),
- Interface reconstruction methods (Hirt and Nichols, 1981), (Youngs, 1982),
- Level Set methods (Dervieux and Thomasset, 1981),(Karni, 1996), (Fedkiw *et al.*, 1999).

Lagrangian and ALE methods are not well adapted for flows with very large distortions, and for inflow and outflow boundary conditions. Front tracking methods combine several flow solvers and are very difficult to code. Interface reconstruction methods are not conservative at the interface and unclear regarding compressible flows (the literature is too poor to evaluate these methods). Level set methods seem very interesting with the approach of Fedkiw (Fedkiw *et al.*, 1999). Indeed the same equations are solved everywhere with the same flow solver. Simplicity and generality are key points for the design of efficient numerical schemes.

The **second class** of methods allows numerical diffusion at the interfaces.

The interface is solved as a numerical diffusion zone, as classically done with shock capturing methods. In this context, simplicity and generality of the algorithm is favoured. The aim that is followed with these methods is a building of a scheme that works on a fixed grid, that allows interface deformations as large as possible, that deals with inflow and outflow boundary conditions in a simple way, and that uses the same numerical scheme for all computational cells (shocks, interfaces, rarefaction waves). Here, two types of methods exist:

- Methods based on the Euler equations (Karni, 1996), (Shyue, 1998), (Saurel and Abgrall, 2000),
- Methods based on multiphase flows equations (Saurel and Abgrall, 1999). The first class, based on the Euler equations consist in rather simple methods, non conservative, very efficient regarding computer resources but limited to rather simple physical situations.

The last method is based on another model involving a large number of equations (multiphase flow model). This model is certainly more difficult to solve and more expensive in computer resources, but the resolution can be achieved for all mesh points with the same method. The model complexity is balanced by its generality and its capabilities to solve difficult problems involving mixtures and interfaces. Regarding interface problems with Euler or Navier-Stokes model, an artificial mixing zone appears. With this model mixing is naturally involved: pure phases, interfaces or true mixtures are considered as multiphase mixtures.

Compared to the others methods for interface computation this one insures conservativity of the mixture: total energy, mass and momentum of the mixture are preserved. Consequently, the interface temperature is accurately determined.

But the most important feature is that this method does not work only for interface problems. It is able to model non-equilibrium two-phase flows as well as flows with interfaces. This has serious implications for the modelling of shock waves in compressible mixtures, detonation waves in heterogeneous materials, and any flows involving compressible mixtures. It also posses a major feature: the method is able to dynamically create interfaces. This is very important for the simulation of cavitating flows (or flows with mass transfer). So this method is able to deal with a wide range of applications as we are detailing now.

Some remarks on the modelling of detonation waves in solid energetic materials

Most detonation codes for solid energetic materials are based on the same kind of model as those of gaseous detonations: reactive Euler equations with mixture equation of state and more or less complicated chemical kinetics. But in the context of solid explosives, the assumptions required for

the building of a mixture equation of state are very questionable. Building of mixture equation of state is based on a combination of the pure material equations of state and some thermodynamic assumptions to couple them. The first thermodynamic relation is the pressure equilibrium between each phase. This assumption can easily be justified for these problems. The main difficulty arise from the second equilibrium relation: temperature or density equilibrium, or others equalities between thermodynamic functions. The density equilibrium assumption is obviously wrong: there is no physical reason for the gas and the solid phase to have the same density. The temperature equilibrium assumption is wrong too. This assumption is the one used in gaseous and most solid detonation models. In gaseous detonations, there is no phase change but only a change in the chemical composition of the mixture occurs. Indeed, the molecular collisions are so intense that the assumption of temperature equilibrium between the various components is valid. For solid explosive detonations, this no longer holds. The detonation is composed of a shock wave propagating in the solid material and followed by a reaction zone where energy is released with a finite rate. During the energy release the reactive solid material transforms to gaseous products and the reaction zone always corresponds to a multiphase mixture. In this multiphase mixture, the assumption of thermal equilibrium is totally wrong. The reason is that the size of solid elemental particles is much bigger than the molecular size in gas mixtures, and molecular collisions cannot make uniform the temperature in the gas and within the particles.

A way to circumvent this problem is to determine more thermodynamic variables for the multiphase mixture. The consequence is to replace the Euler equations by a multiphase model. In this context, each phase will be governed by his own set of partial differential equation, closed by the equation of state of the corresponding pure material. The multiphase model for interface modelling will be extended for this purpose.

Some remarks on the modelling of cavitating flows

Cavitation occurs in liquids or solids when a pressure drop is large enough that the resulting thermodynamical state corresponds to a point in the two-phase region of the phase diagram. Such a situation may occur when a strong rarefaction wave propagates into a liquid or when inertial effects induce pressure drop. Cavitation is difficult to model because interfaces have to be created from a pure phase (liquid) to another phase (gas). According to our knowledge, the modelling of cavitating flows is achieved in specific situations only:

- homogeneous bubbly flows (Tan and Bankoff, 1984), (Massoni *et al.*, 1999),
- study of a single or limited number of interfaces (cavitation pockets) (Molin *et al.*, 1997),

- homogeneous cavitation with mixture models (Saurel *et al.*, 1999).

Models for bubbly flows are restricted in applications because the gas phase must be initially present and also because the flow topology is fixed. For instance, cavitation pockets cannot be predicted by this type of model.

The cavitation pockets model does not allow external bubbly flows. Also, the liquid phase is considered incompressible, thus restricting the domain of application. But the main limitation is that the interfaces need to be initially settled: interface creation is not allowed.

Cavitation with mixture models allows dynamic interface creation (in a certain sense). These models are based on the compressible Euler or Navier-Stokes equations closed by an equation of state valid for all the states of the fluid: pure liquid, pure vapour and two-phase mixture. But again, there are some limitations with this approach. Non-equilibrium states are forbidden, the flow topology is ignored and the mass transfer is assumed to occur instantaneously.

It is now clear that previous cavitation models are restricted to specific applications. The multiphase model we propose for interfaces and detonation waves can also be used for cavitation problems. Of course, it cannot cover all range of interest but possess several features that render it more general. It is able to consider compressibility of all phases, it takes into account uncondensable gases, it is able to model metastable mixtures, to create dynamically interfaces, and also, at least in theory, the model is able to make coexisting bubbly flows and cavitation pockets.

2. The multiphase model

The averaging method of Drew and Passman (Drew and Passman, 1998) applied to the compressible Navier-Stokes equations of the various constituents is used to obtain the multiphase flow model. All dissipative terms are neglected everywhere except at the interfaces. This model is developed in Saurel and Abgrall (Saurel and Abgrall, 1999) for a two-phase system. It is inspired from the formidable work of Baer and Nunziato (Baer and Nunziato, 1986) where a two-phase model is proposed to study the deflagration-to-detonation transition in solid energetic materials. The main difference between Baer and Nunziato work and the present one is related to the use of special relaxation terms and the new numerical method. We introduce here the notion of infinitely fast relaxation regarding pressure and velocity. That makes possible the numerical treatment of interface problems, and open the model to a wider range of applications (interfaces, detonations, cavitation, and others multiphase systems). The model is composed of a

set of five partial differential equations for each phase k .

$$\begin{aligned} \frac{\partial \alpha_k}{\partial t} + u_i \nabla \alpha_k &= \mu(P_k - P'_k) + m_k / \rho_X \\ \frac{\partial \alpha_k \rho_k}{\partial t} + \nabla(\alpha_k \rho_k u_k) &= m_k \\ \frac{\partial \alpha_k \rho_k u_k}{\partial t} + \nabla(\alpha_k \rho_k u_k \otimes u_k + \alpha_k P_k) &= P_i \nabla \alpha_k + m_k u_i + F_{dk} \\ \frac{\partial \alpha_k \rho_k E_k}{\partial t} + \nabla(u_k (\alpha_k \rho_k E_k + \alpha_k P_k)) &= \\ P_i u_i \nabla \alpha_k + m_k E_{ki} + F_{dk} u_i + Q_{ki} + \mu P_i (P_k - P'_k) \\ \frac{\partial N_k}{\partial t} + \nabla(N_k u_k) &= \dot{N}_k \end{aligned} \quad (1)$$

with averaged interface conditions:

$$\begin{aligned} \sum_k m_k &= 0 \\ \sum_k P_i \nabla \alpha_k + m_k u_i + F_{dk} &= 0 \\ \sum_k P_i u_i \nabla \alpha_k + m_k E_{ki} + F_{dk} u_i + Q_{ki} + \mu P_i (P_k - P'_k) &= 0 \end{aligned} \quad (2)$$

The volume fraction α_k is defined by the volume occupied by phase k over the total volume. The saturation constraint imposes $\sum \alpha_k = 1$. Density, velocity, pressure and total energy are represented respectively by ρ , u , P and $E = e + 1/2uu$. The subscripts k and i are related to phase k and interface averaged variables respectively.

The left-hand sides of System (1) are classical. On the right-hand side of the same equations appear the mass transfer m_k , the drag force F_{dk} , the heat transfer Q_i and the non conservative terms $P_i \nabla \alpha_k$ and $P_i u_i \nabla \alpha_k$. The $\mu(P_k - P'_k)$ and $\mu P_i (P_k - P'_k)$ terms are related to the pressure relaxation process. They are of capital importance.

The first equation of System (1) expresses the evolution of the phase volume fractions. It is obtained from averaging of an indicator function equal to 1 in phase k and 0 elsewhere. This equation is a simplification of a more general volume fraction evolution equation accounting for inertial effects (rebounding bubbles for example) and others interface kinematics considerations. Here, for mathematical, numerical and physical reasons, the model is closed with this simplified equation.

The last equation of System (1) represents the evolution of the number density of individual entity composing phase k . For instance, if phase k is the gas phase filling bubbles, N_k then represents the density number of bubbles. Knowledge of the density number of elemental particles is important for determining the surface of mass, momentum and energy exchanges between phases. The term \dot{N}_k models break-up or coalescence of elemental particles. When the assumption of spherical elemental particles is not valid, the exchange surface determination is a more acute problem. Such a

difficulty occurs when the flow changes its topology. Determination of the interfacial area in the general case is still an open problem. The interested reader will find information in (Drew and Passman, 1998).

Before giving details about the various terms, let us give a simple picture of the physical meaning of the non conservative terms $P_i \nabla \alpha_k$ and $P_i u_i \nabla \alpha_k$. The basic one-dimensional Euler equations averaged over a duct of variable cross section A :

$$\begin{aligned} \frac{\partial A\rho}{\partial t} + \frac{\partial (A\rho u)}{\partial x} &= 0 \\ \frac{\partial A\rho u}{\partial t} + \frac{\partial A(\rho u^2 + P)}{\partial x} &= P \frac{\partial A}{\partial x} \\ \frac{\partial A\rho E}{\partial t} + \frac{\partial A u (\rho E + P)}{\partial x} &= P \frac{\partial A}{\partial t} \end{aligned} \quad (3)$$

In two-phase systems, the volume fraction α is sometimes used as a surface fraction. If this analogy is retained and the temporal derivative $\frac{\partial A}{\partial t}$ is replaced by a space derivative $u \frac{\partial A}{\partial x}$ by the means of the volume fraction evolution equation, the one dimensional averaged Euler equations and the multiphase model match. This means that the non-conservative terms in the multiphase model have the same effects as the duct variation cross section terms. Their effects are well known in steady flows: acceleration of subsonic flows in area restriction for example. This simple picture can be important for the derivation of numerical schemes, or results analysis.

This also means that the multiphase model, in a certain sense, couples several Euler systems in fictitious ducts of variable cross section. These "ducts" have permeable walls for the various transfers, they move with the flow at velocity u_i and they expand with the pressure differential $\mu(P_k - P'_k)$ at a rate controlled by μ .

Closure relations not depending on the physical processes. They are related to the determination of the two new interfacial averaged variables P_i and u_i . An accurate estimate of these variables is nearly impossible in the general case and certainly unnecessary for our applications. For all the present applications, the pressures and the velocities will be relaxed instantaneously after each hydrodynamic time step. So, our strategy is to choose interfacial average variables close to the relaxed state. We also need estimates that use only variables determined by the multiphase system and a choice that preserves symmetry is preferred. Indeed, each phase being compressible, there is no reason to prefer a specific phase. Thus, our esti-

mates are:

$$\begin{aligned} P_i &= \sum \alpha_k P_k \\ u_i &= \sum \alpha_k \rho_k u_k / \sum \alpha_k \rho_k \end{aligned} \quad (4)$$

Note that there is a large degree of freedom for these estimates without changing the hyperbolicity of the model. Hyperbolicity in the model is a result of phase compressibility and not of interfacial variables as done in (Bestion, 1990) or (Sainsaulieu, 1995) with conditionally hyperbolic models. Our model is unconditionally hyperbolic (Saurel and Abgrall, 1999).

Closure relations depending on the physical process. In most physical situations, the drag force, the mass and heat transfers, particles break-up and coalescence are finite rate processes and are modelled by empirical closure laws, based on separated effects experiments. We do not enter in the details here. The novelty here is related mainly on pressure and velocity relaxation terms which require some details.

Pressure terms. The model involves non-classical interaction terms regarding the pressure relaxation process: $\mu(P_k - P'_k)$ in the volume fraction evolution equation and $\mu P_i(P_k - P'_k)$ in the energy equation. The first term represents the rate of expansion of the volume fraction α_k in order to the pressures tend towards equilibrium. The physical meaning of this term is very simple. If the various phases are not in pressure equilibrium after the passage of a rarefaction or shock wave, the volume of each phase must vary in order to reach the pressure equilibrium. The variable μ controls the rate at which this equilibrium is reached. The existence of this variable has been shown theoretically according the second law of thermodynamics and mechanics of irreversible processes (Baer and Nunziato, 1986).

When the pressures are in a non-equilibrium state, the elementary particles (bubbles, drops etc.) undergo a 3D microscopic motion making their volume vary in order to the pressures tend towards equilibrium. This 3D motion has not been taken into account in the averaged phase velocities and in the guess of the interfacial velocity (4). The interfacial velocity represents the average motion of the mixture, so the microscopic motion is not considered. Introducing a volume variation function of the pressure differential is a way to correct the estimate for the averaged interfacial velocity, and also a way to take the information from the microscopic media.

We have shown in (Saurel and Abgrall, 1999) that these terms were crucial for the computation of pressure waves in two-phase mixtures (as in shock or detonation waves), but also of paramount importance for restoring the pressure interface condition when solving interfaces between compressible pure materials.

Velocity terms The velocity relaxation term is the most classical one in multiphase systems and is represented by the drag force F_{dk} . What is

non-classical in our approach is to consider an infinite relaxation drag coefficient for specific applications. General drag force may be written under the form: $F_{dk} = \lambda_k(u_k - u'_k)$ where λ_k is a positive finite function (or vector if there are more than 2 fluids). It controls the rate at which velocities tend towards equilibrium. In special physical situations, this function tends to infinity. For example, it is the case in materials with very high deviatoric stress tensor. Imagine gas pores inside a solid set into motion by a strong shock wave. The gas inside the pores will have its motion imposed nearly instantaneously by the surrounding solid. The same situation occurs in metal alloys and metal powders. Such type of situation has been studied recently by Kapila *et al.* (Kapile *et al.*, 1997) in the limit of very high drag coefficients.

When solving interface problems between pure fluids with the multiphase model we have shown that the missing characteristic directions at the interface may be replaced by source terms with infinite pressure and velocity relaxation coefficients (Saurel and Abgrall, 1999). The first interface condition (pressure equality) is restored when the coefficient μ tends to infinity. To restore the second interface condition (velocity equality), an infinite drag coefficient λ must be used. The numerical procedures with infinite relaxation coefficients are detailed in the next section.

3. Numerical method

The previous model can be used for interface computations, detonation waves, cavitation and other physical problems (Saurel and Abgrall, 1999). It can also be easily shown that the summation over all the phases of the mass, momentum and energy equations reduces to the mixture Euler equation. So the mixture is perfectly conservative. Frame invariance of the equations can also be easily demonstrated as well as unconditional hyperbolicity. The solution strategy we adopt is based on an operator splitting:

$$U_i^{n+1} = L_S^{\Delta t} L_R^{\Delta t} L_H^{\Delta t} U_i^n \quad (5)$$

L_S represents the integration operator for source terms: the mass and energy finite rate transfers. When velocity and pressure relaxation are also finite rate, the same operator is used instead of L_R , the infinite relaxation operator. L_S is a standard ODE solver, dependant on the problem stiffness. The infinite relaxation operator L_R is not easy to develop. We have detailed it in (Saurel and Abgrall, 1999) in the context of two fluids only. We generalise it in the following for an arbitrary number of materials, governed by arbitrary equations of state. This is of major importance for the applications with detonations in hydrocodes, where a large number of materials coexist. But the major difficulties rely in the hyperbolic solver L_H which is

now detailed. The basic elements are given in (Saurel and Abgrall, 1999). We generalise it here to multidimensions.

HYPERBOLIC OPERATOR

The numerical method applies at all mesh point: single phase, two-phase and at the interfaces. For the sake of simplicity and generality regarding complex equations of state we have retained the simplest ingredients for the construction of a high-resolution scheme for multiphase flows with arbitrary equations of state. The Riemann solver is chosen for an easy implementation with the various models and equations of state even though the accuracy can be improved.

The hyperbolic system involves several difficulties. Non-conservative terms and a non-conservative equation (the volume fraction evolution equation) are present. We have proposed in (Saurel and Abgrall, 1999) an efficient way to discretise these terms. The major guideline for the building of the numerical scheme can be stated as follows: *If a multiphase flows evolves under uniform pressure and velocity conditions, these flow variables must remain uniform during time evolution.*

This guide has been systematically exploited in the context of the Euler equation and it has shown that it was providing an efficient discretisation scheme for non conservative equations even if velocity and pressure were not initially uniform (Saurel and Abgrall, 2000). The method has been developed in (Saurel and Abgrall, 1999) in 1D for two phases only. We extend it here to multidimensions and arbitrary number of fluids. The two-dimensional hyperbolic system to solve for phase k reads:

$$\begin{cases} \frac{\partial \alpha_k}{\partial t} + u_i \frac{\partial \alpha_k}{\partial x} + v_i \frac{\partial \alpha_k}{\partial y} = 0 \\ \frac{\partial U}{\partial t} + \frac{\partial F(U)}{\partial x} + \frac{\partial G(U)}{\partial y} = H(U) \frac{\partial \alpha_k}{\partial x} + I(U) \frac{\partial \alpha_k}{\partial y} \end{cases} \quad (6)$$

with $U = (\alpha_k \rho_k, \alpha_k \rho_k u_k, \alpha_k \rho_k v_k, \alpha_k \rho_k E_k, N_k)^T$,

$$F(U) = \begin{pmatrix} \alpha_k \rho_k u_k \\ \alpha_k (\rho_k u_k^2 + P_k) \\ \alpha_k \rho_k u_k v_k \\ \alpha_k u_k (\rho_k E_k + P_k) \\ N_k u_k \end{pmatrix}, \quad G(U) = \begin{pmatrix} \alpha_k \rho_k v_k \\ \alpha_k \rho_k u_k v_k \\ \alpha_k (\rho_k v_k^2 + P_k) \\ \alpha_k v_k (\rho_k E_k + P_k) \\ N_k v_k \end{pmatrix},$$

$$H(U) = (0, P_i, 0, P_i u_i, 0) \text{ and } I(U) = (0, 0, P_i, P_i v_i, 0).$$

The basic ingredients of the finite volume method used here are described in (Toro, 1997) in the context of the Euler equations. We consider

a computational cell i in the two-dimensional (x, y) space. We note n_{ij} the external unit normal vector of the j side of cell i : $n_{ij} = (n_{ijx}, n_{ijy})^T$. We also note $K = (F, G)$ the tensor of fluxes and $S = (H, I)$ the non conservative vectors.

We first have to consider the second equation of system (6): $\frac{\partial U}{\partial t} + \nabla \cdot K = S \cdot \nabla \alpha_k$. Integrating this equation over a fixed control volume V delimited by its edges, leads to:

$$V \frac{\partial U}{\partial t} + \sum_{sides} \int T^{-1} F(TU) dL = \int (S \cdot \nabla \alpha_k) dV$$

where T is the rotation matrix and T^{-1} is its inverse.

First order time and space approximation yields the following result:

$$U_i^{n+1} = U_i^n - \Delta t / V_i \sum_{j=1}^4 T_{ij}^{-1} L_{ij} \widehat{F}_{ij}^* + \Delta t (H(U_{ij}^n) \Delta_x + I(U_{ij}^n) \Delta_y)$$

where Δ_x and Δ_y are numerical approximations of $\frac{\partial \alpha_k}{\partial x}$ and $\frac{\partial \alpha_k}{\partial y}$, L_{ij} is the length of the j side of cell i and \widehat{F}_{ij}^* is the HLL (Harten *et al.*, 1983) numerical flux on the corresponding cell boundary in the rotated frame of reference (along the normal n_{ij}):

$$\widehat{F}_{ij}^* = (S_{Rj} \widehat{F}_{Lj} - S_{Lj} \widehat{F}_{Rj} + S_{Rj} S_{Lj} (\widehat{U}_{Rj} - \widehat{U}_{Lj})) / (S_{Rj} - S_{Lj})$$

where $\widehat{U}_{ij} = T_{ij} U_{ij}$, $\widehat{F}_{ij} = F(\widehat{U}_{ij})$, S_L and S_R are the right and left wave speeds in the HLL solver. Since $T_{ij}^{-1} \widehat{F}_{ij} = n_{ij} K_{ij}$ the Godunov method reads:

$$U_i^{n+1} = U_i^n - \lambda \sum_{j=1}^4 L_{ij} \Phi + \Delta t (H(U_{ij}^n) \Delta_x + I(U_{ij}^n) \Delta_y) \quad (7)$$

with $\Phi = (S_{Rj} n_{ij} K_{Lj} - S_{Lj} n_{ij} K_{Rj} + S_{Rj} S_{Lj} (U_{Rj} - U_{Lj})) / (S_{Rj} - S_{Lj})$ and $\lambda = \Delta t / V_i$. To determine to discretisation formulas for the non-conservative terms and equations, we consider a multiphase mixture under uniform pressure and velocity conditions and develop the various steps of the Godunov method (Godunov *et al.*, 1979). There are some situations where it is necessary to consider sliding effects along a contact discontinuity or an interface. Such situations are explained in (Saurel and Abgrall, 2000). In those cases, the discretisation method accounting for sliding effects as detailed in this reference is preferred. Here, we assume that these effects are not of major importance for the actual applications.

So, under the assumptions $P_k = P_i = P$ and $\vec{V}_k = \vec{V}_i = \vec{V}$, the mass conservation equation first reads: $(\alpha_k \rho_k)_i^{n+1} = (\alpha_k \rho_k)_i^n - \lambda \sum_{j=1}^4 L_{ij} \Omega$ with

$$\Omega = [n_{ij} \cdot \vec{V} (S_{Rj}(\alpha_k \rho_k)_{Lj} - S_{Lj}(\alpha_k \rho_k)_{Rj}) + S_{Rj} S_{Lj} ((\alpha_k \rho_k)_{Rj} - ((\alpha_k \rho_k)_{Lj}))]/(S_{Rj} - S_{Lj})$$

Under the same assumptions, the x-momentum equation reads:

$$(\alpha_k \rho_k u)_i^{n+1} = (\alpha_k \rho_k u)_i^n - \lambda \sum_{j=1}^4 L_{ij} \Omega + \Delta t P_i \Delta_x \text{ with :}$$

$$\Omega = \left[(n_{ijx} (u^2 (S_{Rj}(\alpha_k \rho_k)_{Lj} - S_{Lj}(\alpha_k \rho_k)_{Rj}) + P(S_{Rj} \alpha_{k_{Lj}} - S_{Lj} \alpha_{k_{Rj}})) + (n_{ijy} u v (S_{Rj}(\alpha_k \rho_k)_{Lj} - S_{Lj}(\alpha_k \rho_k)_{Rj}) + u S_{Rj} S_{Lj} ((\alpha_k \rho_k)_{Rj} - (\alpha_k \rho_k)_{Lj}))) \right] / (S_{Rj} - S_{Lj})$$

By multiplying the mass equation by u and subtracting it to the momentum one, in order to maintain the velocity uniform at the next time step, the non conservative term must be discretised by:

$$\Delta_x = 1/V_i \sum_{j=1}^4 L_{ij} n_{ijx} (S_{Rj} \alpha_{k_{Lj}} - S_{Lj} \alpha_{k_{Rj}}) / (S_{Rj} - S_{Lj}) \quad (8)$$

The same developments for the y-momentum equation yields the discretisation formula of $\frac{\partial \alpha_k}{\partial y}$:

$$\Delta_y = 1/V_i \sum_{j=1}^4 L_{ij} n_{ijy} (S_{Rj} \alpha_{k_{Lj}} - S_{Lj} \alpha_{k_{Rj}}) / (S_{Rj} - S_{Lj}) \quad (9)$$

Finally, by doing the same type of calculation for the energy equation we obtain the discretisation formula for the volume fraction evolution equation, under the constraint that the pressure must not have any variation during the time step. The results is:

$$(\alpha_k)_i^{n+1} = (\alpha_k)_i^n - \lambda \sum_{j=1}^4 L_{ij} \Omega \quad (10)$$

with

$$\Omega = \left(n_{ij} \cdot \vec{V} (S_{Rj} \alpha_{k_{Lj}} - S_{Lj} \alpha_{k_{Rj}}) + S_{Rk} S_{Lk} (\alpha_{k_{Rj}} - \alpha_{k_{Lj}}) \right) / (S_{Rj} - S_{Lj})$$

This equation is nothing else than a numerical approximation of $\frac{\partial \alpha_k}{\partial t} + \vec{V} \cdot \nabla \alpha_k = 0$.

The term factor of \vec{V} in Ω is the discretisation of $\nabla \alpha_k$, while the other term is a viscous one. It is important to note that discretisation of non conservative terms is strongly dependent of the Riemann solver, and discretisation of the non conservative equation makes a viscous term also dependent of the solver appear.

The Godunov method (7), with non-conservative approximations (8) and (9) and non-conservative scheme (10) constitutes the first order two-dimensional method for System (6). Second-order extension follows MUSCL strategy (vanLeer, 1979). The predictor step is done under primitive variable

formulation. This choice of variables insures that pressure and velocity will remain uniform after the predictor step, when starting from uniform conditions. Predicted variables are computed at the center of each cell boundary with:

$$W_{ij}^{n+1/2} = W_i^n + (x_{cij} - x_{ij})\delta_x W_{ij} + (y_{cij} - y_{ij})\delta_y W_{ij} \\ - \Delta t/2(A(W_{ij}^n)\delta_x W_{ij} + B(W_{ij}^n)\delta_y W_{ij})$$

where (x_{cij}, y_{cij}) and (x_i, y_i) are the cell boundary and control volume center coordinates respectively, W is the primitive variables vector, $A(W)$ and $B(W)$ the Jacobian matrix, and $\delta_x W_{ij}$, $\delta_y W_{ij}$ the limited slopes along each direction. The primitive variables vector reads: $W = (\alpha_k, \rho_k, u_k, v_k, P_k, N_k)^T$ and the Jacobian matrixes:

$$A(W) = \begin{pmatrix} u_i & 0 & 0 & 0 & 0 \\ \rho_k/\alpha_k(u_k - u_i) & u_k & \rho_k & 0 & 0 \\ (P_k - P_i)/\alpha_k\rho_k & 0 & u_k & 0 & 1/\rho_k \\ 0 & 0 & 0 & u_k & 0 \\ \rho_k c_{ki}^2/\alpha_k(u_k - u_i) & 0 & \rho_k c_k^2 & 0 & u_k \\ 0 & 0 & N_k & 0 & 0 \end{pmatrix}$$

$$B(W) = \begin{pmatrix} v_i & 0 & 0 & 0 & 0 \\ \rho_k/\alpha_k(v_k - v_i) & v_k & 0 & \rho_k & 0 \\ 0 & 0 & v_k & 0 & 0 \\ (P_k - P_i)/\alpha_k\rho_k & 0 & 0 & v_k & 1/\rho_k \\ \rho_k c_{ki}^2/\alpha_k(v_k - v_i) & 0 & 0 & \rho_k c_k^2 & v_k \\ 0 & 0 & N_k & 0 & v_k \end{pmatrix}$$

Then, by developing the Godunov scheme over a time step under velocity and pressure uniformity constraints, the resulting corrector step is:

$$(\alpha_k)_i^{n+1} = (\alpha_k)_i^n - \lambda \sum_{j=1}^4 L_{ij} \Omega \text{ and}$$

$$U_i^{n+1} = U_i^n - \Delta t/V_i \sum_{j=1}^4 T_{ij}^{-1} L_{ij} \widehat{F_{ij}^{*n+1/2}} \\ + \Delta t(H(U_{ij}^{n+1/2})\Delta_x + I(U_{ij}^{n+1/2})\Delta_y), \text{ with}$$

$$\Omega = (n_{ij} \cdot \vec{V}_i^{n+1/2} (S_{Rj}^{n+1/2} \alpha_{k_{Lj}}^{n+1/2} - S_{Lj}^{n+1/2} \alpha_{k_{Rj}}^{n+1/2}) + \\ S_{Rk}^{n+1/2} S_{Lk}^{n+1/2} (\alpha_{k_{Rj}}^{n+1/2} - \alpha_{k_{Lj}}^{n+1/2})) / (S_{Rj}^{n+1/2} - S_{Lj}^{n+1/2}),$$

$$\Delta_x = 1/V_i \sum_{j=1}^4 L_{ij} n_{ijx} (S_{Rj}^{n+1/2} \alpha_{k_{Lj}}^{n+1/2} - S_{Lj}^{n+1/2} \alpha_{k_{Rj}}^{n+1/2}) / (S_{Rj}^{n+1/2} - S_{Lj}^{n+1/2})$$

$$\Delta_y = 1/V_i \sum_{j=1}^4 L_{ij} n_{ijy} (S_{Rj}^{n+1/2} \alpha_{k_{Lj}}^{n+1/2} - S_{Lj}^{n+1/2} \alpha_{k_{Rj}}^{n+1/2}) / (S_{Rj}^{n+1/2} - S_{Lj}^{n+1/2})$$

This scheme is stable under the standard CFL condition, based on the largest wave speed. We now examine the various source and relaxation operators in the specific context of infinitely fast relaxation processes.

SOURCE AND RELAXATION OPERATORS

The solution of the overall problem has a physical meaning only after the relaxation procedure application. So this step is of paramount importance.

As given by Equation (5), we have to solve source terms (finite rate relaxation) as must done for example with mass transfer. This step is classical and not detailed here. We also have to deal with relaxation terms (infinite rate relaxation) for pressure and velocity. We first explain this last one.

Velocity relaxation operator

For each phase k we have to solve the ODE system:

$$\begin{aligned} \frac{\partial \alpha_k}{\partial t} &= 0 \\ \frac{\partial \alpha_k \rho_k}{\partial t} &= 0 \\ \frac{\partial \alpha_k \rho_k u_k}{\partial t} &= \lambda(u_{k'} - u_k) \\ \frac{\partial \alpha_k \rho_k E_k}{\partial t} &= \lambda u_i(u_{k'} - u_k) \end{aligned} \quad (11)$$

where the relaxation coefficient λ tends to infinity. That means that for any arbitrary small time increment, the velocities must be equal.

The combination of the mass and the momentum equations yields: $\frac{\partial u_k}{\partial t} = \lambda(u_{k'} - u_k)/(\alpha_k \rho_k)$ for phase k and $\frac{\partial u_{k'}}{\partial t} = -\lambda(u_{k'} - u_k)/(\alpha_{k'} \rho_{k'})$ for phase k' . Subtracting the first equation to the second and integrating leads to the expected result: $u_k^* - u_{k'}^* = 0$. Then, summing the same equations and integrating yields the relaxed velocity:

$$u_k^* = \sum (\alpha_k \rho_k u_k)_0 / \sum (\alpha_k \rho_k)_0 \quad (12)$$

the subscript 0 indicates the solution obtained from the hyperbolic solver. Note that this relaxed velocity corresponds to the estimate we have proposed for the averaged interfacial velocity in (4). So, in every situation where velocities are relaxed instantaneously, the estimated (4) is an accurate one. Note also that relaxation procedure is an exact one. It is also a straightforward extension of the two-fluid case as derived in (Saurel and Abgrall, 1999).

It now remains to update the internal energies, since System (12) involves relaxation terms in the energy equation. Again, combination of the mass, momentum and energy equations and exact integration leads to the result:

$$e_k^* = e_{k0} + 1/2(u_k^* - u_{k0})^2 \quad (13)$$

We now examine the pressure relaxation step for an arbitrary number of fluids.

Pressure relaxation operator

We proposed in (Saurel and Abgrall, 1999) a procedure valid only for two

fluids. Here, we improve the accuracy of the relaxation pressure step and we generalise it to an arbitrary number of fluids. For any phase k we have to solve the ODE system:

$$\begin{aligned}\frac{\partial \alpha_k}{\partial t} &= \mu(P_k - P_{k'}) \\ \frac{\partial \alpha_k \rho_k}{\partial t} &= 0 \\ \frac{\partial \alpha_k \rho_k u_k}{\partial t} &= 0 \\ \frac{\partial \alpha_k \rho_k E_k}{\partial t} &= -\mu P_i (P_k - P_{k'})\end{aligned}\tag{14}$$

where the relaxation coefficient μ tends to infinity.

Combination of the volume fraction, mass, momentum and energy equations yields: $\alpha_k \rho_k \frac{\partial e_k}{\partial t} = -P_i \frac{\partial \alpha_k}{\partial t}$ with P_i given by Equation (4). Since $m_k = \alpha_k \rho_k = \text{const}$ then $d\alpha_k = -m_k d\rho_k / \rho_k^2$. The energy equation becomes: $\frac{\partial e_k}{\partial t} = P_i \frac{\partial 1/\rho_k}{\partial t}$. By using a trapezoidal approximation: $e_k^* - e_k^0 = (P_i^* + P_i^0)/2(1/\rho_k^* - 1/\rho_k^0)$, where variables marked with an asterisk are the relaxed ones, and variables marked with 0 are obtained from the velocity relaxation step. This corresponds to N equations (N is the number of fluids) with $2N+1$ unknowns: N energies, N densities and P_i^* . N equations of state are available for closure: $e_k^* = e_k(P_i^*, \rho_k^*)$. The saturation constraint $\sum \alpha_k = 1$ provides the last equation.

The system to solve now reads (we suppress the symbol *):

$$\begin{aligned}2\rho_1 \rho_1^0 (e_1 - e_1^0) + (P_i + P_i^0)(\rho_1 - \rho_2^0) &= 0 \\ \dots \\ 2\rho_N \rho_N^0 (e_N - e_N^0) + (P_i + P_i^0)(\rho_N - \rho_N^0) &= 0 \\ \sum m_k / \rho_k - 1 &= 0\end{aligned}\tag{15}$$

The solution of this non-linear system is obtained with the Newton Raphson

method. We set $X = \begin{pmatrix} \rho_1 \\ \dots \\ \rho_N \\ P_i \end{pmatrix}$ and

$$F(X) = \begin{pmatrix} 2\rho_1 \rho_1^0 (e_1 - e_1^0) + (P_i + P_i^0)(\rho_1 - \rho_2^0) \\ \dots \\ 2\rho_N \rho_N^0 (e_N - e_N^0) + (P_i + P_i^0)(\rho_N - \rho_N^0) \\ \sum m_k / \rho_k - 1 \end{pmatrix}.$$

Newton method reads: $D(X^{l-1}) \Delta X^l = -F(X^{l-1})$ where l designates the current iteration and $\Delta X^l = (X^l - X^{l-1})$.

The solution is obtained when $\Delta X^l < \epsilon$. $D(X)$ represents jacobian matrix

of the non linear system ($D(X) = \frac{\partial F(X)}{\partial X}$) and is given by:

$$D(X) = \begin{pmatrix} A_1 & 0 & \dots & 0 & B_1 \\ & \ddots & & & \ddots \\ & & \ddots & & \ddots \\ & & & A_N & B_N \\ m_1/\rho_1 & \dots & \dots & m_N/\rho_N & 0 \end{pmatrix}$$

with $A_k = 2\rho_k^0(e_k - e_k^0) + 2\rho_k\rho_k^0e_{k\rho} + (P_i + P_i^0)\rho_k$ and $B_k = 2\rho_k\rho_k^0e_{k\rho} + (\rho_k - \rho_k^0)$. This procedure is robust and accurate. It has been used in all test problems.

4. Test problems

We consider here test problems involving interfaces, cavitation, shock and detonations in one and two space dimensions. All test problems involve instantaneous pressure and velocity relaxation. Other test problems with finite rate relaxation (two velocities) and other applications are available in (Saurel and Abgrall, 1999).

Water - air shock tube

A shock tube filled on the left side with high-pressure liquid water and on the right side with air is considered. This test problem consists in a classical shock tube with two fluids and has an exact solution. On this test problem, standard methods based on the Euler equations fail at the second time step.

Each fluid is governed by the Stiffened Gas equation of state (Godunov *et al.*, 1979):

$P = (\gamma - 1)\rho e - \gamma P_{inf}$ where γ and P_{inf} are constant parameters.

The initial data are: $\rho_l = 1000\text{kg/m}^3$, $P_l = 10^9\text{Pa}$, $u_l = 0$, $\gamma_l = 4.4$,

$P_{inf,l} = 6.10^8\text{Pa}$, $\alpha_l = 1 - \epsilon$ ($\epsilon = 10^{-6}$) if $x < 0.7$; $\rho_g = 50\text{kg/m}^3$,

$P_g = 10^5\text{Pa}$, $u_g = 0$, $\gamma_g = 1.4$, $P_{inf,g} = 0$, $\alpha_g = 1 - \epsilon$ otherwise.

A mesh with 1000 cells is used to show convergence. The corresponding results are shown in Figure 1 at time 229 ms. On this test case, the right and left chamber contain nearly pure fluids: the gas volume fraction in the water chamber is only 10^{-6} and inversely in the gas chamber. It clearly appears that the correct waves speed are reproduced by the method, and that the method converges to the correct solution. This test problem shows that the method works correctly on interface problems.

One-dimensional cavitation tube

Consider a tube filled with water and imagine that the left part of this tube is set to motion to the right, and the left part is set to motion in the

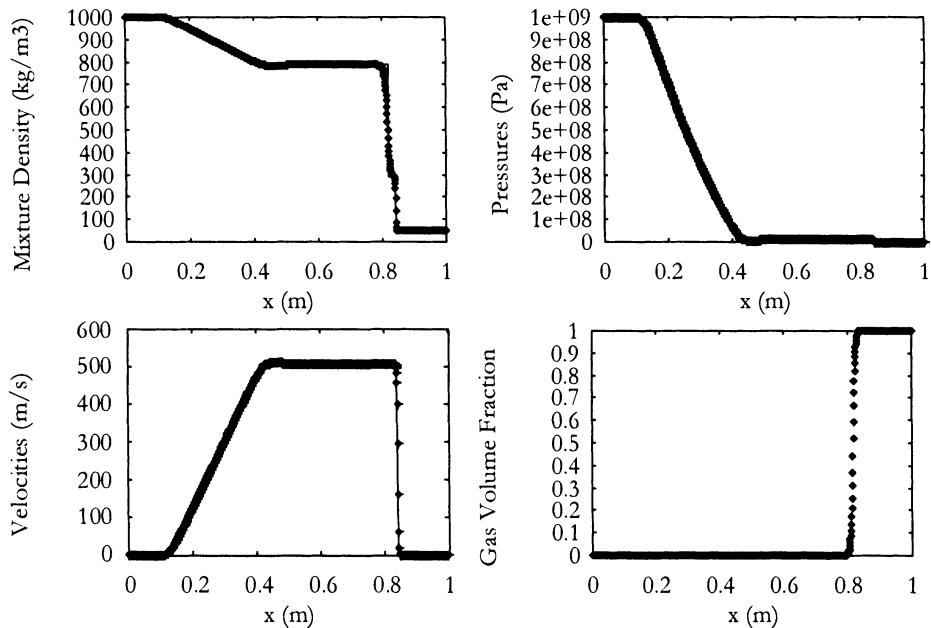


Figure 1. Water-air shock tube. Computed solution with 1000 cells (symbols) and exact solution (lines)

opposite direction. In such situation, the pressure, density and internal energy decrease across the rarefaction waves in order that the velocity reaches zero at the center of the domain. The pressure decreases until the saturation pressure at the local temperature is reached. Then the mass transfer appears, a part of the liquid becomes gas, and the flow becomes a two-phase mixture.

When this type of problem is solved with the Euler equations, so with a single fluid model and an appropriate equation of state for the liquid, the pressure becomes completely wrong (i.e. negative). The reason is that the liquid EOS is no longer valid when the pressure becomes sub-atmospheric. Indeed the liquid transform to gas, and the gas has neither the same EOS, nor the same behaviour. With the multiphase model, each phase is described by its own EOS.

Because mass and energy transfer will strongly affect the results and also because there are strong uncertainty about these correlation, we consider a simplified problem with a small fraction of gas initially present in the liquid (1% gas of its volume) and we remove the mass transfer. From this initial situation where the gas and liquid are at atmospheric pressure, we set into motion the right part of the tube at 100m/s, and the left part at -100m/s. The results are shown on Figure 2.

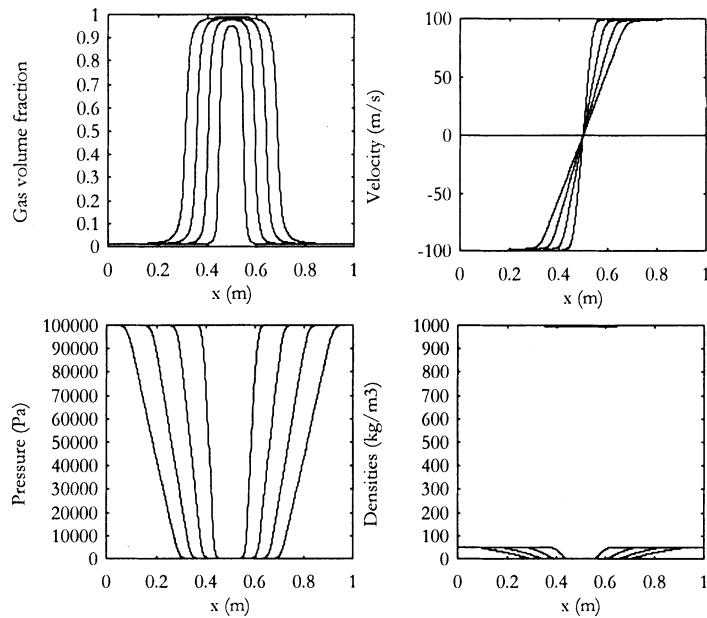


Figure 2. 1D cavitation tube. Interfaces appear dynamically, pressure remains positive and liquid density remains inside the domain of validity of the EOS.

The liquid density decreases slightly but remains closed to the initial one: it remains liquid. Gas density decreases across the rarefaction waves and decreases again due to the pressure relaxation process. The gas density at the centre of the tube is very low. The pressure relaxation process makes the gas volume fraction to increase thus creating two interfaces propagating to the right and to the left. So this method has the capability to create dynamically interfaces starting from a pure (or nearly pure) liquid. This feature has important applications for specific problems, as shown with the next example.

Two-dimensional cavitation around an obstacle in a supersonic liquid flow

The two-dimensional unsteady calculations initiate with the same mixture as previously : 99% water and 1% gas at atmospheric pressure and temperature flowing over an obstacle under supersonic conditions. The obstacle surface is treated as a rigid wall with centerline symmetry. The outflow and upstream boundaries are treated as nonreflecting boundaries. At the inflow boundary on the left-hand side the fluid is pure liquid and moving at 2000 m/s. The obstacle has conical-shaped leading and trailing edges and a cylindrical centre body. Since the inflow is supersonic, a detached shock

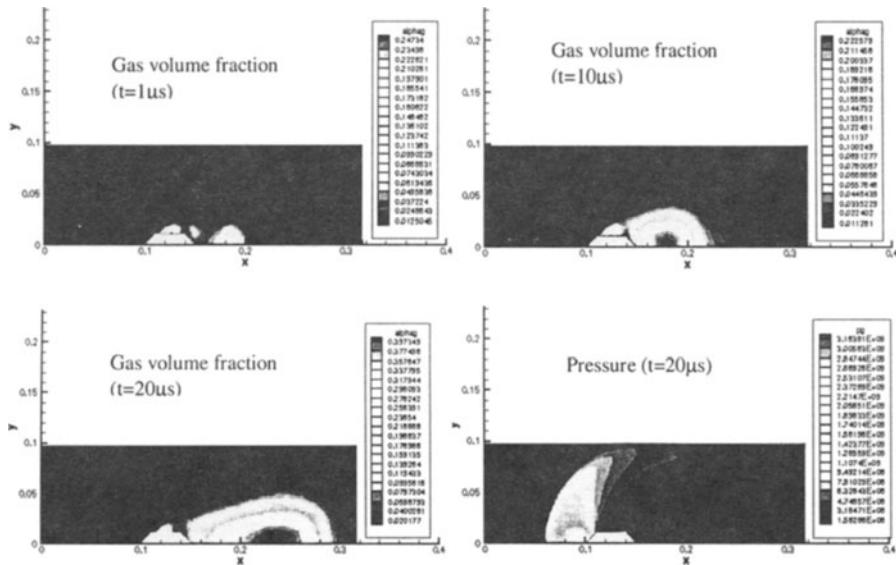


Figure 3. 2D supersonic liquid flow around a cylindrical projectile. Cavitation pockets appear.

wave is expected in front of the obstacle as represented on the pressure contours in Figure 3. On the two angular points connecting the cones to the cylindrical portion, strong rarefaction waves are expected, possibly capable of inducing cavitation. Volume fraction contours are presented in Figure 3 at 1, 10 and 20 μs to illustrate the unsteady formation of cavitation pockets and bubble tearing.

Detonation test problems

We have shown in (Saurel and Abgrall, 1999) that the model and method were able to compute very strong shock waves in two-phase mixtures (Mixture Hugoniot test problem). We evaluate here the capabilities of the method to compute detonation waves in solid energetic materials. Compared to the Mixture Hugoniot test problem, it now involves mass and energy transfers. Reference test problems are rather rare on detonations and nearly absent regarding multiphase detonations.

A detonation wave in solid explosive consists in a shock wave followed by a reaction zone where the solid transforms to gas with an energy release. So, all reaction zone of solid explosives consists in a two-phase mixture. The interest of the multiphase model, as explained previously in section 2, is that no mixture equation of state are needed and that each phase will have its own density, energy and temperature.

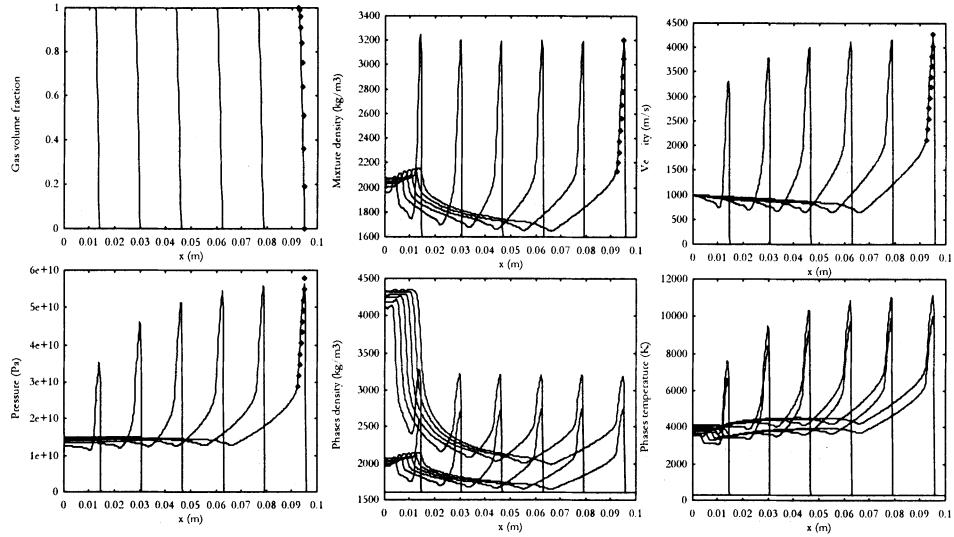


Figure 4. Detonation test problems. Mixture variables are compared to ZND calculation inside the reaction zone. Results also show that fluids temperatures and densities are never in equilibrium.

An unusual test problem is chosen in order to compare the solution of the multiphase model to an exact solution. We consider that solid and gas phases are governed by the same equation of state, and to simplify the analysis we chose the ideal gas equation of state $P = (\gamma - 1)\rho e$ with $\gamma = 3$ for both the phases. Doing this, the multiphase model solution must degenerate to the single phase Euler solution for all mixture variables.

Within the detonation wave reaction zone, the reactive Euler equations are exactly solved between the shock point (Neumann spike) and the point moving with the sonic velocity relative to the shock front (CJ point). This classical calculation corresponds to the ZND model solution and can be performed exactly in simplified situation as given in Fickett and Davis (Fickett and Davis, 1979), or by the resolution of an ODE problem with a nearly exact accuracy. Here, we use the conditions of the ZND problem given in the Fickett and Davis's book.

We use exactly the same explosive data and the whole flow is solved in an unsteady regime. After the shock-to-detonation transition, a stable detonation wave must be obtained. When this detonation is stable (the last curves of the previous figure) we compare the reaction zone obtained from the computation and the exact solution. The comparison is only possible in the reaction zone because the ZND model is valid only in this part. Also, the comparison is possible only on mixture flow variables: the ZND problem has never been solved for multiphase mixtures. Results are shown in Figure

4.

It appears that the numerical solution converges to the exact one. The other important results is that the densities and temperatures of the two fluids never reach an equilibrium inside the reaction zone, even on this basic test problem where the equation of state and material properties are exactly the same. This result is in fact obvious: the gas phase receives energy from the solid phase, while the solid phase does not receive anything.

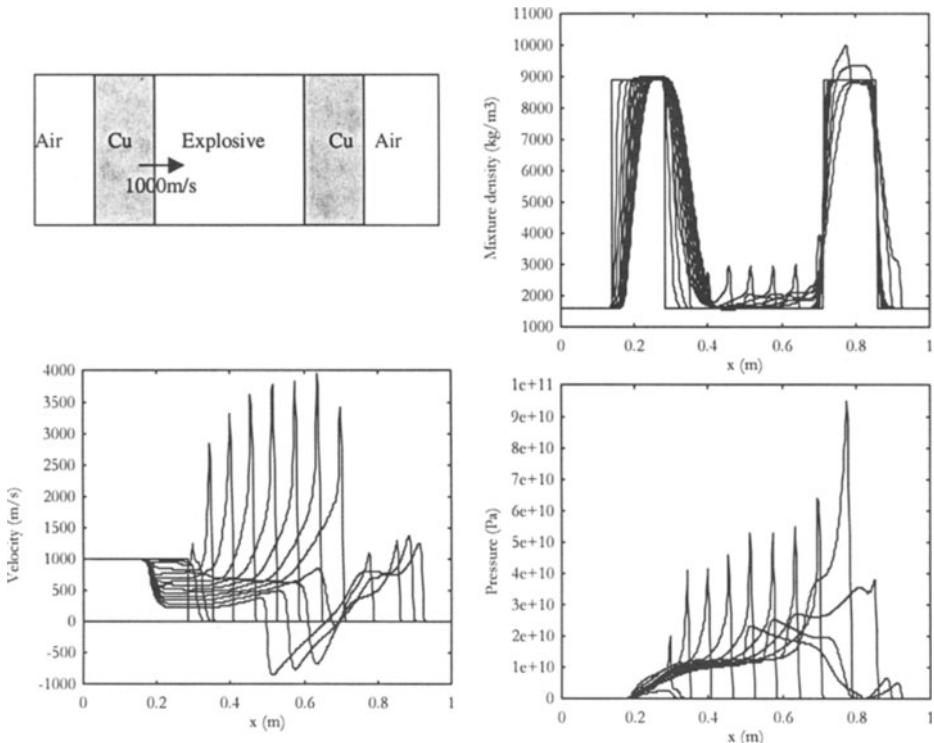


Figure 5. Impact of a copper plate over an explosive, transition to detonation and shock interaction with a copper target.

We end this presentation by a last test case involving the interaction of a detonation wave with a target material: an inert copper plate. Introduction of this new material involves a third fluid and two supplementary interfaces. Here we use all the capabilities of the model: three fluids are present, with four interfaces, a multiphase zone inside the explosive, with mass and energy transfers. The pressures and the velocities are relaxed with the procedure for an arbitrary number of fluids. Simple equations of state are used, but the method works for complicated ones. Note that this

situation with interaction of detonation waves with inert materials is very representative of detonation applications.

Results are represented in Figure 5. We just show some mixture variables that are easy to analyse.

5. Conclusion

We have shown some important applications of the model and numerical method. It is able to:

- create dynamically interfaces in cavitating flows,
- make interact multiphase mixture with interfaces as in the detonation interaction problem.

Besides, it does not need any mixture equation of state and provides thermodynamic variables of each phase. It is also conservative regarding the mixture, even at the interfaces. This provides accurate energy and temperature computation at the interfaces.

Others applications are possible.

Acknowledgments

This work has been partially supported by DGA/ETCA/CEG Gramat, CEA/DAM Bruyeres le Chatel, RENAULT Direction de la Recherche and PEUGEOT. The author is particularly grateful to Gerard Baudin (CEG), Serge Gauthier (CEA) and Lionel Sainsaulieu (RENAULT). He also addres special thanks to Olivier Lemetayer, Jacques Massoni and Eric Daniel for their help with the codes and daily support.

References

- Abgrall, R. (1996) How to prevent pressure oscillations in multicomponent flow calculations: A quasi conservative approach. *J. Comp. Phys.*, **125**, pp 150-160
- Baer, M.R. and Nunziato, J.W., (1986) A two-phase mixture theory for the deflagration-to-detonation transition (DDT) in reactive granular materials. *Int. J. of Multiphase Flows*, **12**, pp 861 - 889
- Benson, D.J. (1992) Computational methods in Lagrangian and Eulerian hydrocodes. *Comp. Meth. in Appl. Mech. and Eng.*, **99**, pp 235-394
- Bestion, D. (1990) The physical closure laws in the CATHARE code. *Nuc. Eng. and Design*, **124**, pp 229-245
- Dervieux, A., and Thomasset, F. (1981) Multifluid incompressible flows by a finite element method. *Lect. Notes in Physics*, **141**, pp 158-163 (1981)
- Drew, D.A. and Passman, S.L.(1998) Theory of multicomponent fluids. Springer New York. Applied Mathematical Sciences **135**
- Farhat, C. and Roux, F.X. (1991) A method for finite element tearing and interconnecting and its parallel solution algorithm *Int. J. Num. Meth. Eng.* , **32**, pp 1205-1227

- Fedkiw, R.P., Aslam, T., Merriman, B. and Osher, S. (1999) A non-oscillatory Eulerian approach to interfaces in multimaterial flows (the Ghost Fluid Method). *J. Comp. Phys.*, **152:2**, pp 457-492
- Fickett, W. and Davis, W.C (1979) Detonation. *University of California Press*, Berkeley
- Glimm, J., Grove, J.W., Li, X.L., Shyue, K.M., Zhang, Q., and Zeng, Y. (1998) Three dimensional front tracking. *SIAM J. Sc. Comp.*, **19**, pp 703-727
- Godunov S.K., Zabrodine, A., Ivanov, M., Kraiko, A. and Prokopov, G. (1979) Resolution numerique des problemes multidimensionnels de la dynamique des gaz. *Editions Mir*, Moscou
- Harten, A. and Hyman, M.J. (1983) Self adjusting grid methods for one-dimensional hyperbolic conservation laws. *J. Comp. Phys.*, **50**, pp 235-269
- Harten, A., Lax, P.D. and van Leer, B. (1983) On upstream differencing and Godunov type schemes for hyperbolic conservation laws. *SIAM Rev.*, **25(1)**, pp 33-61
- Hirt, C.W. and Nichols, B.D. (1981) Volume Of Fluid (VOF) method for the dynamics of free boundaries. *J. Comp. Phys.*, **39**, pp 201-255
- Kapila, A., Son, S., Bdzil, J., Menikoff, R. and Stewart, D. (1997) Two-phase modeling of DDT: structure of the velocity-relaxation zone. *Phys. Fluids* **9** : 12, pp 3885-3897
- Karni, S. (1996) Hybrid multifluid algorithms. *SIAM Journal of Scientific Computing* , **17:5**, pp 1019-1039
- Massoni, J., Saurel, R., Baudin, G. and Demol, G. (1999) A mechanistic model for shock initiation of solid explosives. *Phys. Fluids* , **11**: 3, pp 710-736
- Molin, B., Dieval, L., Marcer, R. and Arnaud, M. (1997) Modelisation instationnaire de poches de cavitation par la methode VOF. *6eme Journees de l'Hydrodynamique*. ISSN 1161-1847
- Sainsaulieu, L. (1995) Finite volume approximations of two phase fluid flows based on an approximate Roe-type Riemann solver. *J. Comp. Phys.*, **121**, pp 1-28
- Saurel, R., and Abgrall, R., (1999) A multiphase Godunov method for compressible multifluid and multiphase flows, *J. Comp. Phys.*, **150**, pp 425-467
- Saurel, R., and Abgrall, R., (2000) A simple method for compressible multifluid flows. *SIAM J. Sc. Comp.*, Vol. 21, n3, pp 1115-1145
- Saurel, R., Cocchi, J.P., and Butler, P.B., (1999) A numerical study of the cavitation effects in the wake of a hypervelocity underwater projectile. *J. Prop. and Power* , **15:4**, pp 513-522
- Tan, M.J., and Bankoff, S.G., (1984) Propagation of pressure waves in bubbly mixtures. *Phys. Fluids*, **27:5** , pp 1362-1369
- Toro, E.F. (1997) Riemann solvers and numerical methods for fluid dynamics. *Springer Verlag* , Berlin.
- Shyue, K.M. (1998) An efficient shock-capturing algorithm for compressible multicomponent problems. *J. Comp. Phys.*, **142**, 208
- van Leer, B. (1979) Toward the ultimate conservative difference scheme V. A second order sequel to Godunov's Method. *J. Comp. Phys.*, **32**, p101
- Youngs, D.L. (1982) Time dependent multi-material flow with large fluid distortion. Eds. K.W. Morton and M.J. Baines, Academic Press

ONE-DIMENSIONAL CALCULATION OF UNSTEADY OPEN CHANNEL FLOW USING ADAPTIVE MESH REFINEMENT

JENS SCHRAMM, SUSANNE ENK, JÜRGEN KÖNGETER

*Institute of Hydraulic Engineering and Water Resources Management,
University of Technology Aachen (RWTH Aachen),
Kreuzherrenstrasse, 52056 Aachen, Germany
Email: schramm@iww.rwth-aachen.de*

Abstract. In the present study a one-dimensional numerical model for the simulation of unsteady flow in compound open channels is combined with an adaptive mesh refinement. To solve the applied Saint-Venant Equations the Godunov first-order upwind scheme with the source terms discretised according to Garcia-Navarro and Vázquez-Cendón (Garcia-Navarro and Vázquez-Cendón, 1998) was used. To minimize the errors due to an inflexible grid, an effective adaptive mesh refinement algorithm with refinement both in time and space has been implemented into the model.

1. Introduction

Almost all major hydraulic engineering activities involve computation of water level and discharge along a river reach. Computation of the water level is essential for the determination of the effect of a hydraulic structure on the channel, inundation of land due to dam, weir or barrage construction and estimation of the flood zone. Because of practical importance, the calculation of water level and discharge along a river reach has been a topic of continued interest to hydraulic engineers. Today, complex hydraulic phenomena occurring in rivers can be simulated with two- or three-dimensional numerical models. So far, these models can not be easily used in engineering practice due to high computational costs. On the other hand, one-dimensional models offer the possibility of cheaply evaluating the response of the hydraulic system to a variety of practical situations. The one-dimensional models developed so far are mostly applied to rivers which

are characterized by large discharges and homogeneous and compact geometry with subcritical flow. Contrary, small natural rivers are characterized by a highly irregular cross section configuration with abrupt contractions and expansions in channel width, varying boundary roughness along the wetted perimeter and the flow changing from sub- to supercritical and vice versa. When applying one of the existing one-dimensional models to one of these small rivers, spatial oscillations are often observed in the calculated water level and discharge.

To overcome the described problems that are due to the dominant source term a new one-dimensional hydraulic model was developed. It combines the Godunov first order upwind scheme (Godunov, 1959; LeVeque, 1992) with the source term discretised according to Garcia-Navarro and Vázquez-Cendón and an adaptive mesh refinement.

2. Fundamental Equations

The one-dimensional Saint-Venant Equations, which are based on the conservation of mass and momentum, can be used to describe the unsteady flow in natural rivers with composite cross sections and considerably varying boundary roughness (Cunge, Holly and Verwey, 1980). They can be written as the following system of equations:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}(x, U)}{\partial x} = \mathbf{R}(x, U) \quad (1)$$

so that

$$\mathbf{U} = (A, Q)^T \quad (2)$$

$$\mathbf{F} = \left(Q, \frac{Q^2}{A} + gI_1(x, A) \right)^T \quad (3)$$

$$\mathbf{R} = (q, gI_2(x, A) + gA(S_0 - S_f))^T \quad (4)$$

\mathbf{U} represents the vector of the dependent variables. A is the wetted area from which the water level can be derived and Q is the discharge. \mathbf{F} represents the vector of the numerical flux. The gravitational acceleration is represented by g . I_1 multiplied by g represents the pressure force at the cross section. Point and nonpoint sources are represented by q . I_2 multiplied by g is the increase of pressure force due to the width variation for an infinitesimal channel length. S_0 represents the slope of the channel. The energy dissipation due to roughness S_f will be considered through empirical friction laws, e.g. Manning. The Jacobian matrix \mathbf{J} of the Saint-Venant Equation is

$$\mathbf{J} = \frac{\partial \mathbf{F}}{\partial \mathbf{U}} = \begin{pmatrix} 0 & 1 \\ c^2 - u^2 & 2u \end{pmatrix} \quad (5)$$

with the eigenvalues and eigenvectors of \mathbf{J}

$$\lambda^{1,2} = u \pm c; \quad \mathbf{e}^{1,2} = (1, u \pm c)^T \quad (6)$$

To solve the Saint-Venant Equation, we use the Godunov first order scheme for non-linear systems (Toro, 1999)

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i^n + \frac{\Delta t}{\Delta x} [\mathbf{F}_{i-\frac{1}{2}} - \mathbf{F}_{i+\frac{1}{2}}]. \quad (7)$$

To compute water level and discharge at time $n+1$ the numerical flux difference $\Delta \mathbf{F} = \mathbf{F}_{i+\frac{1}{2}} - \mathbf{F}_{i-\frac{1}{2}}$ has to be computed for every interface. According to Roe (Roe, 1981), an approximated Jacobian matrix $\tilde{\mathbf{J}}$ can be used in order to locally linearize the Saint-Venant Equations. The eigenvalues $\tilde{\lambda}^{1,2}$ and the eigenvectors $\tilde{\mathbf{e}}^{1,2}$ of $\tilde{\mathbf{J}}$ satisfy for every pair of nodal points ($i, i+1$) in the one-dimensional grid. The numerical flux becomes:

$$\Delta \mathbf{F} = \tilde{\mathbf{J}}_{i+\frac{1}{2}} \Delta \mathbf{U}_{i+\frac{1}{2}} = \sum_k \left(\tilde{\lambda}_k \tilde{\alpha}_k \tilde{\mathbf{e}}_k \right)_{i+\frac{1}{2}}. \quad (8)$$

Following the approach of Garcia-Navarro and Vázquez-Cendón with b being the width of the cross section, $\Delta b = \frac{1}{2}(b_i - b_{i-1})$, z being the altitude of the cross section and $\Delta z = \frac{1}{2}(z_i - z_{i-1})$, the source term \mathbf{R} can be discretised as

$$\mathbf{R} \Delta x = \sum_k \tilde{\beta}_k \tilde{\mathbf{e}}_k \text{ with } \tilde{\beta}_1 = -\tilde{\beta}_2 = \frac{g}{2\tilde{c}} \left(-\tilde{A} \Delta z - \tilde{A} \Delta x \tilde{S}_f + \frac{\tilde{A}^2}{\tilde{b}^2} \Delta b \right). \quad (9)$$

With the averaged values for velocity \tilde{u} and celerity \tilde{c} defined as

$$\tilde{u}_{i+\frac{1}{2}} = \frac{\frac{Q_{i+1}}{\sqrt{A_{i+1}}} + \frac{Q_i}{\sqrt{A_i}}}{\sqrt{A_{i+1}} + \sqrt{A_i}} \quad \text{and} \quad \tilde{c}_{i+\frac{1}{2}} = \sqrt{\frac{g}{2} \left(\left(\frac{A}{b} \right)_i + \left(\frac{A}{b} \right)_{i+1} \right)} \quad (10)$$

and the wave strength $\Delta \mathbf{U} = \mathbf{U}_R - \mathbf{U}_L = \sum_k \alpha_k \mathbf{e}_k$ with

$$\alpha_1 = \frac{(\tilde{u} + \tilde{c})(A_{i+1} - A_i) - (Q_{i+1} - Q_i)}{2\tilde{c}} \quad (11)$$

$$\alpha_2 = \frac{-(\tilde{u} - \tilde{c})(A_{i+1} - A_i) + (Q_{i+1} - Q_i)}{2\tilde{c}} \quad (12)$$

the discretization of the system becomes with $C = \mathbf{U}_i^n + \frac{\Delta t}{\Delta x}$:

$$U_i^{n+1} = C \cdot \left[\left(\sum_{k+} \left(\tilde{\lambda}_k \tilde{\alpha}_k - \tilde{\beta}_k \right) \tilde{\mathbf{e}}_k \right)_{i-\frac{1}{2}} + \left(\sum_{k-} \left(\tilde{\lambda}_k \tilde{\alpha}_k - \tilde{\beta}_k \right) \tilde{\mathbf{e}}_k \right)_{i+\frac{1}{2}} \right] \quad (13)$$

where $k\pm$ indicates the positive or negative eigenvalues at every interface.

3. Adaptive Mesh Refinement

Flow phenomena in natural rivers occur in different length scales. To minimize the errors due to an inflexible numerical grid, it is desirable to use mesh refinement to cluster grid points in regions where they are most needed. For example around shocks or in other regions where the solution shows steep gradients. For time-depending problems the region of refinement must move adaptively with the related flow phenomena. An effective adaptive mesh refinement strategy with refinement in both space and time has been developed by Berger and Oliger (Berger and Oliger, 1984). The refinement of the numerical grid is done by an arbitrary even integer ratio. Further recursive refinement can be done within these patches to an arbitrary depth.

The adaptive mesh refinement algorithm described by Berger and Oliger (Berger and Oliger, 1984) uses an error estimation procedure based on Richardson extrapolation to determine the regions where the resolution of the solutions is insufficient. In the present work a different criteria that identifies steep gradients in both water height h and discharge Q has been used. The relation

$$|\Delta h \Delta x| > \epsilon ; \quad |\Delta Q \Delta x| > \epsilon \quad (14)$$

is applied with

$$\Delta h = h_i - h_{i-1} ; \quad \Delta Q = Q_i - Q_{i-1} ; \quad \Delta x = \frac{1}{2} (x_i + x_{i+1}) . \quad (15)$$

The left side of the relation corresponds to the error of the solution in the vicinity of a discontinuity if Δh or ΔQ is a function of the grid. A crucial point of the adaptive mesh refinement when applied to natural rivers is the manner in which division and unification of subgrids are coordinated. A failing dam, for example, causes two waves, a bore and a sunk, to move in opposite directions from the location of the dam. Unification takes place when two flood waves merge at a river junction. The refinement algorithm was modified in the following way that if a reorganization of the subgrids has to be carried out, it is guaranteed that a unification takes place first on the coarse upper level grid before the subgrids will unite. A division of the numerical grid will always take place first on the subgrids before the coarse upper level grids will separate.

4. Numerical Results

Flow in small rivers often changes from sub- to supercritical and vice versa (hydraulic jump). Therefore, the numerical model has to prove that it prop-

erly captures the discontinuities. First, the model is applied to a standard test case developed by MacDonald et al. (MacDonald, Baines, Nichols and Samuels, 1997) for a prismatic cross-section to prove the reliability of the model. The discharge is set to $Q = 20 \text{ m}^3/\text{s}$ and the Manning coefficient

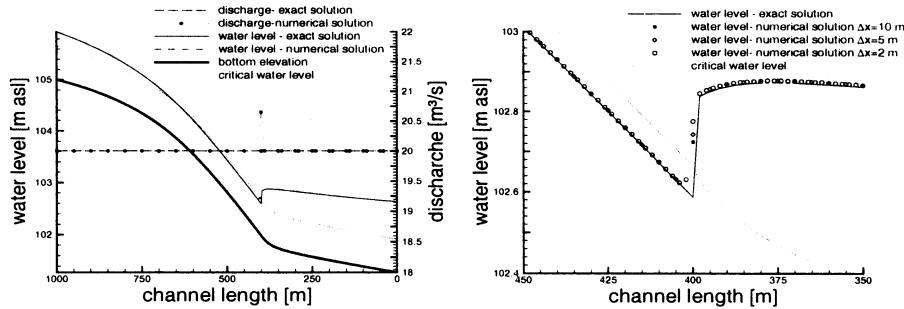


Figure 1. Exact (full line) and numerical (dotted line) solution for water depth $h(x)$. Left plots correspond to grid spacing of $\Delta x = 10 \text{ m}$. Right plots correspond to grid spacing of $\Delta x = 10 \text{ m}$, $\Delta x = 5 \text{ m}$ and $\Delta x = 2 \text{ m}$.

is set to $n = 0.03 \text{ s}/\text{m}^{\frac{1}{3}}$. The mathematical formulation of bed slope and water level are derived in detail by MacDonald et al. In the left picture the results of the numerical solution for the water level (dotted line) correspond very well with the analytical solution (solid line). The right picture illustrates in detail the calculated water depth compared to the analytical solution in the region of the hydraulic jump for different grid spacings. One sees that the numerical solution becomes very accurate with increasing grid spacing.

In order to show that the presented combination of the Godunov first order upwind scheme and an adaptive mesh refinement is a good strategy for unsteady problems a wave propagation is proposed. The hydraulic model is applied to the Jüchener Bach (river length 10.950 km), a small river in North-Rhine-Westphalia/Germany with highly irregular cross-section configurations. The river has a mean discharge of about $Q = 0.081 \text{ m}^3/\text{s}$. At the beginning of the simulation a waste water treatment plant that is located close to the river has a constant discharge of $Q = 0.064 \text{ m}^3/\text{s}$. The discharge of the waste water treatment plant is then rapidly increased to a discharge of $Q = 0.318 \text{ m}^3/\text{s}$ which lasted six minutes. Figure 2 illustrates the change in discharge and water level over time in the river reach. One can see that the increasing discharge causes a large wave traveling downstream with the amplitude decreasing due to retention. Additionally, a small wave is traveling upstream for a short distance. The increase of water level in the vicinity of the plant is about 0.10 m and therefore not illustrated. The

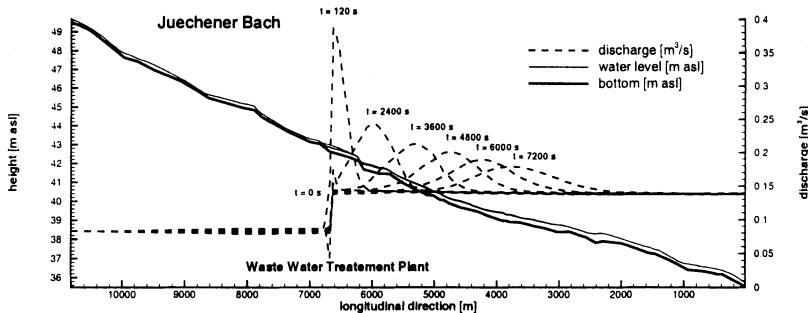


Figure 2. Wave, caused by a rapidly increasing discharge at a waste water treatment plant, moving downstream

model produces very accurate predictions for unsteady flows compared to measured values.

5. Conclusions

In the present study a one-dimensional numerical model using the Godunov first-order upwind scheme for the simulation of unsteady flow in compound open channels was combined with an adaptive mesh refinement. The model produces very accurate flow prediction and is able to deal with rapidly changing discharges, heterogeneous geometry and point sources.

References

- Berger M J, Oliger J (1984). Adaptive Mesh Refinement for Hyperbolic Partial Differential Equations. *J. Comp. Physics*, **53**, pp. 484-512.
- Cunge JA, Holly FM and Verwey A (1980). Practical Aspects of Computational River Hydraulics. Monographs and Surveys in Water Resources Engineering 3. Pitman Publishing Limited.
- Garcia-Navarro P and Vázquez-Cendón M E (1998). Roe's Schemes for 1D Irregular Geometries. *Hydroinformatics 1998, Babovic & Larsen, Balkema, Rotterdam*.
- Godunov S K (1959). A Finite Difference Method for the Computation of Discontinuous Solutions of the Equations of Fluid Dynamics. *Mat. Sb.* **47**, pp 357-393.
- LeVeque R J (1992). Numerical Methods for Conservation Laws. Birkhäuser Verlag.
- MacDonald I, Baines M J, Nichols N K, Samuels P G (1997). Analytic Benchmark Solutions for Open-Channel Flows *Int. J. Hydraulic Engineering*, **123(11)**, pp. 1041-1044.
- Roe P L (1981). Approximate Riemann Solvers, Parameter Vectors and Difference Schemes. *J. Comp. Physics*, **43**, pp. 357-372.
- Toro E F (1999). Riemann Solvers and Numerical Methods for Fluid Dynamics. Second Edition, Springer-Verlag.

ERROR ESTIMATES FOR GODUNOV-TYPE SCHEMES IN THE PRESENCE OF SOURCE TERMS

HANS JOACHIM SCHROLL

Department of Mathematical Sciences

Norwegian University of Science and Technology

N-7491 Trondheim, Norway

Email: schroll@math.ntnu.no

Abstract. The present paper reviews a particular argument how to obtain error estimates for Godunov-type schemes applied to conservation laws with source terms. In the case of stiff sources —so-called relaxation problems— a special monotonicity property of the source term is required. Several examples satisfying this requirement are presented. However, this paper is not a complete review on the topic. The reader is referred to the contribution by Tadmor at the same conference, the lecture notes (Tadmor, 1998) and the proceedings (Aregba-Driollet and Natalini, 1999).

1. The original Godunov scheme

Godunov's numerical method to approximate the homogeneous conservation law

$$u_t + f(u)_x = 0 \quad (1)$$

is an iterative application of the following two steps.

1. Compute cell averages $u_i^n = \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} u(t_n, x) dx$ on a given grid.
2. Solve the conservation law $u_t + f(u)_x = 0$ for $t_n < t < t_{n+1}$ with piecewise constant data $u(t_n, x) = u_i^n$ for $x_{j-1/2} \leq x < x_{j+1/2}$.

For small time steps, the exact solution in step 2 is given by solving local Riemann problems at each cell interface. This can be done exactly and the resulting scheme reads

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + \frac{f(u^*(u_j^n, u_{j+1}^n)) - f(u^*(u_{j-1}^n, u_j^n))}{\Delta x} = 0 .$$

Here, $u^*(u^-, u^+)$ is the solution of the homogeneous Riemann problem

$$u_t + f(u)_x = 0 , \quad u(x, 0) = \begin{cases} u^-, & x < 0 \\ u^+, & 0 < x \end{cases}$$

evaluated along the ray $x/t = 0$. Thus, the scheme consists of a projection step, projecting into the space of piecewise constants, and a solution step. Because of L_1 -contractivity of solutions of the conservation law (1) the L_1 -error does not increase during the solution step. In fact, one can show that the global error in L_1 is of order $\sqrt{\Delta x}$. This was an important contribution due to Kuznetsov (Kuznetsov, 1976), see also (Lucier, 1985). Moreover, the analysis can be extended including a source term in step 2. This leads to an “error estimate” of order $\sqrt{\Delta x}$. However, the result is of little interest, because there is no numerical scheme to solve the inhomogeneous conservation law exactly, even with piecewise constant data.

2. The splitting error

A natural idea to include source terms into numerical schemes is to apply operator splitting. For example the solution of

$$u_t + f(u)_x = g(u) \tag{2}$$

may be approximated by iterating the following two steps.

1. Solve $v_t = g(v)$ for $t_n < t < t_{n+1}$ with data $u(t_n, x)$.
2. Solve $u_t + f(u)_x = 0$ for $t_n < t < t_{n+1}$ with data $v(t_{n+1}, x)$.

It was shown in (Langseth, Tveito and Winther, 1996) that for scalar conservation laws such a splitting approach leads to a global error of order Δt in L_1 . Introducing the notation $E(\Delta t)u^n := u^n + \Delta t g(u^n) = v^{n+1}$ for the forward Euler method approximating step 1 as well as $H(\Delta t)v^{n+1} := u^{n+1}$ for the entropy solution of the homogeneous conservation law in step 2, they show that

$$\|u(T, \cdot) - [H(\Delta t)E(\Delta t)]^N u(0, \cdot)\|_1 = \mathcal{O}(\Delta t) ,$$

where u is the entropy solution of (2) and $T = N\Delta t$. In particular, when using a Godunov scheme in step 2, the dominating error is due to the Godunov scheme and not caused by operator splitting. Omitting all the technicalities the argument can be summarized as follows. Consider one time step from $t = 0$ to $t = \Delta t$. The ordinary differential equation is approximated by a forward Euler step. Then the homogeneous conservation law is solved exactly with the data from the Euler step

$$w_t + f(w)_x = 0 , \quad w(0, \cdot) = E(\Delta t)u(0, \cdot) .$$

Observe that $w(\Delta t, \cdot) = H(\Delta t)E(\Delta t)u(0, \cdot)$, thus the goal is to show that $\|w(\Delta t, \cdot) - u(\Delta t, \cdot)\|_1 = \mathcal{O}(\Delta t^2)$. To establish the estimate, Langseth, Tveito and Winther introduce $\alpha = w - (\Delta t - t)g(u(0, \cdot))$, where obviously $\alpha(\Delta t, \cdot) = w(\Delta t, \cdot)$ and hence it is sufficient to show $\|\alpha(\Delta t, \cdot) - u(\Delta t, \cdot)\|_1 = \mathcal{O}(\Delta t^2)$. The function α is governed by the conservation law

$$\alpha_t + f(\alpha + (\Delta t - t)g(u(0, \cdot)))_x = g(u(0, \cdot)) , \quad \alpha(0, \cdot) = u(0, \cdot) .$$

Using the Kružkov formalism, it is possible to estimate the difference $\alpha - u$ in L_1 , leading to the desired bound. But, what is the reason behind this result? Here is a formal argument. In a first step, consider

$$\beta_t + f(\beta)_x = g(u(0, \cdot)) , \quad \beta(0, \cdot) = u(0, \cdot) .$$

For the difference $e = u - \beta$ it follows that

$$e_t + (ae)_x = g(u) - g(u(0, \cdot)) , \quad e(0, \cdot) = 0 ,$$

where $a = (f(u) - f(\alpha))/(u - \beta)$. Therefore $\frac{d}{dt}\|e(t, \cdot)\|_1 \leq c\|u(t, \cdot) - u(0, \cdot)\|_1 = \mathcal{O}(t)$ and hence $\|u(\Delta t, \cdot) - \beta(\Delta t, \cdot)\|_1 = \mathcal{O}(\Delta t^2)$. It remains to show that $\epsilon = \beta - \alpha$ is of order Δt^2 . The equation for ϵ becomes

$$\epsilon_t + (f(\beta) - f(\alpha + (\Delta t - t)g(u(0, \cdot))))_x = 0 , \quad \epsilon(0, \cdot) = 0 .$$

Introducing $b = [f(\beta) - f(\alpha + (\Delta t - t)g(u(0, \cdot)))]/[\beta - \alpha - (\Delta t - t)g(u(0, \cdot))]$, this can be written as

$$\epsilon_t + (b\epsilon)_x = (\Delta t - t)(bg(u(0, \cdot)))_x ,$$

where now the source term is of order $\Delta t - t$, thus $\frac{d}{dt}\|\epsilon(t, \cdot)\|_1 = \mathcal{O}(\Delta t - t)$ and finally, $\|\epsilon(\Delta t, \cdot)\|_1 = \mathcal{O}(\Delta t^2)$.

3. Stiff source terms

In this section we review an argument which was developed to estimate the error of a semi-implicit Godunov-type scheme applied to a conservation law with a stiff source term

$$u_t + f(u)_x = \frac{1}{\delta}g(u) , \quad 0 < \delta \ll 1 .$$

For the source term we assume monotonicity in the sense

$$\text{sign}(u - v)(g(u) - g(v)) \leq 0 . \quad (3)$$

This inequality will be used below and it allows to generalize to an interesting class of hyperbolic systems with relaxation. The scheme is given by

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + \frac{f(u^*(u_j^n, u_{j+1}^n)) - f(u^*(u_{j-1}^n, u_j^n))}{\Delta x} = \frac{1}{\delta} g(u_j^{n+1}).$$

Under the CFL-condition $\frac{\Delta t}{\Delta x} \|f'\|_\infty \leq 1$ and assuming well prepared initial data such that $\|g(u(0, \cdot))\|_1 = \mathcal{O}(\delta)$ it was shown in (Schroll and Winther, 1996) that

$$\|u(t_n, \cdot) - u^n\|_1 \leq M \Delta t , \quad (4)$$

where M is a uniform constant, independent of δ . For the analysis it is again useful to interpret the scheme as a sequence of solution and projection steps.

1. Solve the conservation law $\bar{u}_\tau = f(\bar{u})_\xi = 0$ for $t_n < \tau < t_{n+1}$ with piecewise constant data $\bar{u}(t_n^+, \xi)$.
2. Compute $\bar{u}(t_{n+1}, \xi) = \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} \bar{u}(t_{n+1}^-, \eta) d\eta$, $\xi \in [x_{j-1/2}, x_{j+1/2}]$.
3. Perform a backward Euler step

$$\bar{u}(t_{n+1}^+, \xi) = \bar{u}(t_{n+1}, \xi) + \frac{\Delta t}{\delta} g(\bar{u}(t_{n+1}^+, \xi)) . \quad (5)$$

Due to the CFL-condition it is true that $\bar{u}(t_n^+, \xi)$ interpolates the discrete data generated by the scheme. Therefore, the goal is to bound $\|\bar{u}(t_n^+, \cdot) - u(t_n, \cdot)\|_1$. The tool for such a bound is the Kružkov inequality. For u it reads

$$\frac{\partial}{\partial t} |u - k| + \frac{\partial}{\partial x} \text{sign}(u - k) (f(u) - f(k)) \leq \frac{1}{\delta} \text{sign}(u - k) g(u) . \quad (6)$$

Here, it is important to choose the constant right, namely $k = \bar{u}(t_n^+, \xi)$. By construction it holds for $\bar{u} = \bar{u}(\tau, \xi)$

$$\frac{\partial}{\partial \tau} |\bar{u} - c| + \frac{\partial}{\partial \xi} \text{sign}(\bar{u} - c) (f(\bar{u}) - f(c)) \leq 0 . \quad (7)$$

However, this inequality is only valid in (t_n^+, t_{n+1}^-) . The gap from t_{n+1}^- to t_{n+1}^+ is governed by projection and the backward Euler formula. For this step we derive

$$|\bar{u}(t_{n+1}^+, \xi) - c| \leq |\bar{u}(t_{n+1}, \xi) - c| + \frac{\Delta t}{\delta} \text{sign}(\bar{u}(t_{n+1}^+, \xi) - c) g(\bar{u}(t_{n+1}^+, \xi)) \quad (8)$$

by subtracting a constant from (5) and multiplying by $\text{sign}(\bar{u}(t_{n+1}^+, \xi) - c)$. In order to compare to u , we choose $c = u(t, x)$ in both (7) and (8). At this stage the analysis follows the classical argument. Both Kružkov inequalities

(6) and (7) are tested by the same test function $\varphi_\epsilon = \omega_\epsilon(x - \xi)\omega_\epsilon(t - \tau)$, where ω_ϵ is a mollifier. The integral forms are integrated once more in (τ, ξ) and (t, x) respectively and added together, using (8). For one single time step one arrives at

$$\|u(t_{n+1}, \cdot) - \bar{u}(t_{n+1}^+, \cdot)\|_1 \leq \|u(t_n, \cdot) - \bar{u}(t_n^+, \cdot)\|_1 + \frac{1}{\delta} \int \int \int \int \text{sign}(u - \bar{u}(t_{n+1}^+, \xi)) (g(u) - g(\bar{u}(t_{n+1}^+, \xi))) \varphi_\epsilon d^4 + \dots$$

Here, the monotonicity of the source term (3) is used to eliminate the parameter δ . The rest is bounded appropriately. For details the reader is referred to the original article (Schroll and Winther, 1996). In fact, the global error is bounded by $M\sqrt{\Delta x}$ uniformly in δ (c.f. 4)! The result was generalized to the multidimensional case by Shen (Shen, 1999).

4. Applications

The technique described above to derive error estimates also applies to systems of conservation laws with relaxation, where the source term has special monotonicity properties. An example is the following model for adsorption

$$\begin{aligned} (u + a)_t + f(u)_x &= 0 \\ \delta a_t &= A(u) - a , \quad 0 < \delta \ll 1 \end{aligned}$$

where $A' \geq 0$. The crucial property of the source term is

$$[\text{sign}(u^1 - u^2) - \text{sign}(a^1 - a^2)] [\left(a^1 - A(u^1) \right) - \left(a^2 - A(u^2) \right)] \leq 0 .$$

The error estimate for the corresponding scheme was established in (Schroll, Tveito and Winther, 1997) and generalized to the two-dimensional case in (Shen, Tveito and Winther, 1996).

Another example where the analysis applies is a model of viscoelasticity, usually written as

$$\begin{aligned} u_t + \sigma_x &= 0 \\ (\sigma - f(u))_t + \frac{1}{\delta} (\sigma - \mu f(u)) &= 0 , \quad 0 < \delta \ll 1 \end{aligned} \tag{9}$$

with $0 < \mu < 1$, $f' > 0$ and $f'' \geq 0$. Introducing $v = f(u) - \sigma$, the system is equivalent to

$$\begin{aligned} \sigma_t + \frac{1}{(f^{-1})'(\sigma+v)} \sigma_x &= \frac{1}{\delta} (\mu v - (1-\mu)\sigma) \\ v_t &= \frac{1}{\delta} ((1-\mu)\sigma - \mu v) . \end{aligned}$$

Here, the (bilinear) source term is monotone in the sense

$$(\text{sign}(\sigma) - \text{sign}(v)) (\mu v - (1-\mu)\sigma) \leq 0 .$$

Again the error estimate for the scheme was derived in the preprint (Shen, Tveito and Winther, 1997). The zero relaxation limit for the viscoelasticity model (9) was also studied in (Luo and Natalini, 1998) and (Yong, 1998).

Finally, we observe that also the so-called relaxation approximation

$$\begin{aligned} u_t + v_x &= 0 \\ v_t + au_x &= \frac{1}{\delta}(f(u) - v) , \quad 0 < \delta \ll 1 \end{aligned}$$

for a conservation law $w_t + f(w)_x = 0$ is a system suitable for the above described analysis. In characteristic variables $\bar{u} = (u - v/\sqrt{a})/2$, $\bar{v} = (u + v/\sqrt{a})/2$ the system reads

$$\begin{aligned} \bar{u}_t - \sqrt{a}\bar{u}_x &= \frac{1}{\delta}(M_-(u) - \bar{u}) \\ \bar{v}_t + \sqrt{a}\bar{v}_x &= \frac{1}{\delta}(M_+(u) - \bar{v}) , \end{aligned}$$

where $M_{\pm}(u) = (u \pm f(u)/\sqrt{a})/2$. Note that $u = \bar{u} + \bar{v}$. By the subcharacteristic condition $\|f'\|_{\infty} \leq \sqrt{a}$, $M'_{\pm} \geq 0$ and furthermore $M'_+ + M'_- = 1$. Therefore, the source term is negative in the sense

$$\begin{aligned} \text{sign}(\bar{u}^1 - \bar{u}^2)(M_-(u^1) - M_-(u^2) - (\bar{u}^1 - \bar{u}^2)) + \\ \text{sign}(\bar{v}^1 - \bar{v}^2)(M_+(u^1) - M_+(u^2) - (\bar{v}^1 - \bar{v}^2)) \leq 0 . \end{aligned}$$

This again allows the error analysis as described in the previous section. Indeed the error estimate for relaxation schemes was recently proved by Liu and Warnecke (Liu and Warnecke, 1999), see also (Aregba-Driollet and Natalini, 1999).

References

- Aregba-Driollet D and Natalini R (1999). Discrete kinetic schemes for systems of conservation laws. In: Hyperbolic problems: theory, numerics, applications, Vol. I (Zürich, 1998) Internat. Ser. Numer. Math. **129**, Birkhäuser, Basel (1999), pp 1–10.
- Kružkov S N (1970). First order quasi linear equations with several space variables. *Math. USSR. Sb.* **10**, pp 217–243.
- Kuznetsov N N (1976). Accuracy of some approximate methods for computing the weak solutions of a first order quasi linear equation. *USSR. Comp. Math. and Math. Phys.* **16**, pp 105–119.
- Langseth J O, Tveito A and Winther R (1996). On the convergence of operator splitting applied to conservation laws with source terms. *SIAM J. Numer. Anal.* **33**, pp 843–863.
- Liu H L and Warnecke G (1999). Convergence rates for relaxation schemes approximating conservation laws. Preprint.
- Lucier B J (1985). Error bounds for the methods of Glimm, Godunov and LeVeque. *SIAM J. Numer. Anal.* **22**, pp 1074–1081.
- Luo T, Natalini R (1998) BV solutions and relaxation limit for a model in viscoelasticity. *Proc. Roy. Soc. Edinburgh Sect. A* **128**, pp 775–795.
- Schroll H J, Tveito A and Winther R (1997). An L1-error bound for a semiimplicit difference scheme applied to a stiff system of conservation laws. *SIAM J. Num. Anal.* **34**, pp 1152–1166.

- Schroll H J and Winther R (1996). Finite difference schemes for scalar conservation laws with source terms. *IMA J. Num. Anal.* **16**, pp 201-215.
- Shen W (1999). Error bounds of finite difference schemes for multi-dimensional scalar conservation laws with source terms. *IMA J. Numer. Anal.* **19**, pp 77-89.
- Shen W, Tveito A, Winther R (1996). A system of conservation laws including a stiff relaxation term; the 2D case. *BIT* **36**, pp 786-813.
- Shen W, Tveito A, Winther R (1997). On the zero relaxation limit for a system modeling the motions of a viscoelastic solid. Preprint.
- Tadmor, E (1998). Approximate solutions of nonlinear conservation laws. In: Advanced numerical approximation of nonlinear hyperbolic equations, Lecture Notes in Mathematics, 1697, Springer, Berlin (1998), pp 1-149.
- Yong W A (1998) A difference scheme for a stiff system of conservation laws *Proc. Roy. Soc. Edinburgh Sect. A* **128**, pp 1403-1414.

A MULTI-DIMENSIONAL EULER SOLVER

R. SCHWANE

European Space Agency, ESTEC

Postbus 299, 2200AG Noordwijk, Netherland

Emails: RSCHWANE@ESTEC.ESA.NL

Abstract.

In this paper a numerical algorithm for the solution of the multi-dimensional Euler equations in conservative and non-conservative form is presented. Most existing standard and multi-dimensional schemes use flux balances with assumed constant distribution of variables along each cell edge, which interfaces two grid cells. This assumption is believed to be one of the main reasons for the limited advantage gained from multi-dimensional high order discretizations compared to standard one-dimensional ones. The present algorithm is based on the optimisation of polynomials describing the distribution of flow variables in grid cells, where only polynomials that satisfy the Euler equations in the entire grid cell can be selected. The global solution is achieved if all polynomials and by that the flow variables are continuous along edges interfacing neighbouring grid cells. Results from the present scheme between first and fifth order spatial accuracy are compared to standard first and second order Roe computations for simple test cases demonstrating the gain in accuracy for a number of sub- and supersonic flow problems.

1. Introduction

Today's CFD applications are still predominantly performed with standard solvers, that are applied to multi-dimensional flows by means of a dimensional splitting technique. The approximate Riemann solver is employed in two or three arbitrary, grid-determined directions in the flow field, where for the balance in each grid direction all influences (flow-gradients) from the other directions are neglected. One-dimensional characteristics of the Euler equations, e.g. the constancy of total enthalpy and entropy on streamlines are assumed also for two-dimensional flows in each arbitrary grid direction.

Examples for significant errors from this approximation are the solutions of flux-difference splitting schemes for oblique shear layers (e.g. enthalpy, entropy, density or velocity) where the assumption of constant characteristic variables along grid lines is strongly violated by the unsteady change that occurs when a grid lines passes through the shear layer under an oblique angle. As a result, CFD solutions become significantly dependent on the orientation of the shear layer to the grid lines.

Numerous attempts have been made during the last decade to introduce multi-dimensional features into CFD solvers, e.g., (Deconink et al., 1994; Powell and van Leer, 1989; Roe, 1994) mainly trying to find a multi-dimensional decomposition of the Euler equations based on a cell flux balance.

The results of these schemes which are now maturing to application level have in common, that they show improved resolution for shocks and shear layers mainly for lower approximation orders, however, the comparison of higher order standard schemes with higher order multi-dimensional schemes results in only minor improvements (Deconink et al., 1994; Walpot, 1994).

This is mainly due to the fact that standard schemes and multi-dimensional schemes are based on flux balances for each cell which are a natural consequence of conservative discretization methods, that are mandatory to compute strong shocks. Flux balance formulations have an intrinsic amount of one-dimensionality.

The situation can only be improved, if some information on the distribution of variables in cells is available, which can not be constructed from the flux balance alone since the flux balance, due to the integration over the cell edges, allows only the average cell values as degree of freedom (Schwane, 1999).

In this paper a numerical schemes is presented, which resolves a certain amount of substructure in cells and as such allows for multi-dimensional high order solutions of the Euler equations.

The present algorithm is based on the optimisation of polynomials describing the distribution of flow variables in grid cells, where only polynomials that satisfy the Euler equations in the entire grid cell can be selected. The global solution is achieved if all polynomials and by that the flow variables are continuous along edges interfacing neighbouring grid cells.

Whereas the problem of finding a multi-dimensional decomposition of the Euler equations, that is a set of (cell averaged) variables with the property of being constant (or linear) in arbitrary grid determined directions of the flow field has not yet been solved, the present methodology allows for the variation of flow variables in cells according to the degrees of freedom of the Euler equations, and finds one dimensional polynomial distributions

for the edges of each cell resulting from the distribution of flow variables in cells, that are constant between neighbouring grid cells.

From this point of view the present scheme can be interpreted as a solution to the problem of the multi-dimensional decomposition of the Euler equations, however, at the expense of the introduction of additional degrees of freedom in the solution in form of polynomial coefficients that form the solution in place of the the cell averages used in standard schemes.

Results from the present scheme between first and fifth order spatial accuracy (for arbitrary but constant stream line angles) are compared to standard first and second order Roe computations for simple test cases demonstrating the gain in accuracy and computational efficiency for a number of sub- and supersonic flow problems.

1.1. RELATION TO GODUNOV SCHEMES

Godunov schemes can be characterized as algorithms that solve the time dependent wave propagation problem that is created at the cell edges by the assumption of constant flow variables in grid cells. Different levels of simplifications can be distinguished between different schemes, mainly in terms of the averaging procedure that simplifies the time dependent Riemann solution, emanating from the cell edges, again in a cell wise constant solution at the new time level.

The present technique follows a similar concept, if it is applied to the space-time domain: The time evolution of the flow variables can be considered as a solution of the Riemann or wave propagation problem in the grid cell where here the cell edges are treated approximatively (matching of cell edge distributions).

The present paper deals with a variant of the general time dependent flow problem, to reduce computational costs and allow for a thorough validation of the method against known steady state and analytical results. It assumes steady state distributions of flow variables in grid cells and employs flux balances over the cell edges for the calculation of the time advance of the solution. Since the fluxes at the cell edges are computed from the distributions of the flow variables as determined by the present scheme, the multi-dimensional character of the solution is maintained.

This choice can be considered as an intermediate step towards the application of the present technique to the space-time domain, that would also resolve the physical CFL condition (time accuracy requires $CFL < 1$ for all standard explicit and implicit time discretization methods). This condition, that posts presently a major constraint on time accurate computations for geometries with strongly varying length scales, is based on a characteristic of balance formulations, that has been discussed earlier in the context of the resolution of oblique shear layers (for the steady state space

domain). Due to the similarity of both problems this paper will concentrate on the application of the technique to steady state problems and apply the techniques to the space-time domain at a later stage.

2. Non-conservative Formulation

In this section the basic scheme is presented for the Euler equations in non-conservative formulation. The extension of the scheme to the Navier Stokes equations is straight forward, viscous terms have to be included in the derivation of equations 1 to 4 and the scheme has to allow for variations of h_t and s in streamwise direction in equation 5.

In (Schwane, 1999) a variant of this scheme with conservation properties for compressible flows with strong shocks is presented.

2.1. GOVERNING EQUATIONS

The governing equations are the Euler equations written in non-conservative form in a co-ordinate system that is aligned to the local streamline. ($\tilde{u} = \sqrt{u^2 + v^2}$, $\tilde{v} = 0$)

$$0 = \frac{\partial h_t}{\partial \xi} \quad (1)$$

$$0 = \frac{\partial s}{\partial \xi} \quad (2)$$

$$0 = \left(\frac{\tilde{u}^2}{a^2} - 1 \right) \frac{\partial p}{\partial \xi} + \rho \tilde{u}^2 \frac{\partial \sin(\alpha)}{\partial \eta} \quad (3)$$

$$0 = \frac{1}{\rho} \frac{\partial p}{\partial \eta} + \tilde{u}^2 \frac{\partial \sin(\alpha)}{\partial \xi} \quad (4)$$

with: $h_t = \frac{\kappa}{(\kappa - 1)} \frac{p}{\rho} + \frac{1}{2} (u^2 + v^2)$, $s = \frac{p}{\rho^\kappa}$

h_t is the total enthalpy and s is a measure for the entropy in the flow field. ρ , \tilde{u} and a denote the cell-average of the density, total velocity and speed of sound. Equation 1 and 2 describe transport of enthalpy and entropy on streamlines. Equation 3 and 4 contain still partial derivatives with respect to the stream line coordinates ξ and η and are coupled. This system describes the coupling of the pressure p to the change of flow direction and con- or divergence of the flow under the assumption of iso-enthalpic and isentropic flow conditions.

2.2. SCALAR EQUATIONS

h_t and s are transported along stream lines, consequently all possible solutions in one grid cell should be constant in streamwise direction, whereas

no information is available from the Euler equations for the shape of the polynomial for h_t and s normal to the streamlines. An approximation of order m_{ord} to this constraint (for arbitrary but constant streamline angle) would be

$$P_{h,s}(\eta) = \sum_{i=0}^{m_{ord}} q_i \eta^i \quad (5)$$

with $P_{h,s}$ as the polynomial distribution of h_t or s in each cell. The coefficients q_i of the h_t and s polynomials in equation 5 are determined by the constraint that scalar distributions as h_t and s have to be continuous at cell interfaces.

The overall result for h_t and s follows from the optimal match of polynomials in cells with respect to the minimisation of the integral of the quadratic deviation of the polynomials along edges that interface neighbouring cells.

$$E = \sum_{n,s,e,w} \int_0^1 [P_o(\eta(t)) - P_c(\eta(t))]^2 dt \equiv \text{Min} \text{ with } \eta(t) = \eta_a + t * (\eta_e - \eta_a) \quad (6)$$

η_a, η_e are the curvilinear co-ordinates of each cell vertex, t is the dimensionless co-ordinate along each edge, the subscript c denotes the central cell and o the neighbouring cells. The summation has to be carried out over all cell edges, with n, s, e, w representing the north, south, east and west direction in the grid.

The quadratic minimum constraint, equation 6, can be modified into a system of linear equations in the coefficients q_i by calculating all partial derivatives of E with respect to the coefficients $\frac{\partial E}{\partial q_i} = 0$

After substitution of the polynomial approximations for h_t and s as given in equation 5 into equation 6 and the binomial expansion of the expression we obtain two linear relations that define the coefficients of the central polynomials as function of the coefficients in the neighbouring cells.

$$M_c \vec{q}_{h,c} = \sum_{n,s,e,w} M_o \vec{q}_{h,o}; \quad M_c \vec{q}_{s,c} = \sum_{n,s,e,w} M_o \vec{q}_{s,o} \quad (7)$$

The elements of M_c and M_o can be derived for an approximation of arbitrary order and are given in (Schwane, 1999).

2.3. EXTENSION TO SYSTEMS

Equation 3 and 4 are coupled. Therefore the numerical procedure for scalar distributions in cells has to be extended to systems in which polynomial approximations are found for p and $\sin(\alpha)$ simultaneously, that have to satisfy the Euler equations. Similar to the solution procedure for h_t and s in the previous section, also here, p and $\sin(\alpha)$ is described by polynomials

of arbitrary order. Since equation 3 and 4 are not transport equations, the polynomial now has to allow for variations in ξ and η direction.

$$P_{p,\sin(\alpha)}(\xi, \eta) = \sum_{k=0}^{m_{ord}} \sum_{i=0}^k q_{ki} \xi^{k-i} \eta^i \quad (8)$$

with q_{ki} as the coefficients of the polynomials for p and $\sin(\alpha)$ respectively. Substituting the partial differentials $\frac{\partial p, \sin(\alpha)}{\partial \xi, \eta}$ in equation 3 and 4 with the polynomial form in equation 8 results in a linear system, of equations in the coefficients of p and $\sin(\alpha)$. Each equation has the general form:

$$0 = \sum_i \sum_j a_{ij} \xi^i \eta^j \quad (9)$$

with a_{ij} as the sum of all coefficients of $\xi^i \eta^j$. If we postulate that the polynomial approximations for p and $\sin(\alpha)$ satisfy the Euler equations in the entire cell, we have to demand that all a_{ij} vanish.

This results in a system of linear equations in the coefficients for each Euler equation.

$$\phi_j = [\underline{\phi_p}, \underline{\phi_{\sin(\alpha)}}] * \vec{q} = \vec{0} \quad (10)$$

with the coefficients of the Euler equations 1 to 4 as the elements of the matrices $\underline{\phi_p}$ and $\underline{\phi_{\sin(\alpha)}}$ and \vec{q} as the vector of the polynomial coefficients of p and $\sin(\alpha)$ respectively.

The entire linear system at each cell reads

$$\begin{aligned} \frac{\partial E}{\partial q_i} + \sum_{j=1}^{m_{Eu}} \lambda_j \frac{\partial \phi_j}{\partial q_i} &= 0 \\ \phi_j &= 0 \end{aligned} \quad (11)$$

m_{Eu} gives the number of linear equations in q_i that constrain the coefficients of p and $\sin(\alpha)$ such, that the polynomials satisfy the Euler equations in grid cells. m_{Eu} depends on the approximation order m_{ord} via $m_{Eu} = \sum_{i=1}^{m_{ord}+1} i$

Computing the derivatives of the minimum condition $\frac{\partial E}{\partial q_i}$ and the Euler constraints ϕ_j this system of equations can be rewritten as

$$\begin{bmatrix} M_c & 0 & \phi_p^T \\ 0 & M_c & \phi_{\sin(\alpha)}^T \\ \phi_p & \phi_{\sin(\alpha)} & 0 \end{bmatrix} \begin{bmatrix} \vec{q}_{p,c} \\ \vec{q}_{\sin(\alpha),c} \\ \lambda \end{bmatrix} = \begin{bmatrix} \sum_{n,s,e,w} M_o \vec{q}_{p,o} \\ \sum_{n,s,e,w} M_o \vec{q}_{\sin(\alpha),o} \\ \vec{0} \end{bmatrix} \quad (12)$$

The elements of the M_c and M_o can be derived for polynomials of arbitrary order similar as for the scalar case as a function of the curvilinear co-ordinates of each cell vertex, ξ_a, ξ_e and η_a, η_e . Details on the derivation of the matrices M_c and M_o can be found in (Schwane, 1999)

The described algorithm determines the best match at cell edges in the sense of minimal deviations of scalar distributions in neighbouring cells. Together with the guaranteed satisfaction of the Euler equations in each cell by the choice of the polynomial approximation for h_t and s for equation 1 and 2 or by the explicit introduction of the Euler constraints for the coefficients for the p and $\sin(\alpha)$ distributions $\phi_j = 0$ for equation 3 and 4 the whole algorithm converges towards the optimal approximation to an overall Euler solution based on the chosen order of the polynomials.

3. Numerical Results

3.1. FLOW PAST INFINITE CYLINDER

Figure 1-2 shows the comparison of a solution of the present scheme with a solution of a standard unlimited second order MUSCL type Roe scheme without entropy correction and second order formulation for the boundary conditions for the Ma=.3 flow past an infinite cylinder. The finest grid for the computation with the Roe scheme is four times denser than the grid used for the present scheme.

The present scheme uses a second order multi-dimensional formulation to allow comparison with the Roe scheme. From the iso-pressure plots in Figure 1 can be derived, that the conservation properties of the present scheme on the 68x34 pt. grid are still slightly superior to the standard scheme for the 272x136 pt. grid. Iso-lines are in fact symmetrical, as they should be, since the Euler solution for this irrotational and isentropic test case recovers exactly the compressible potential solution.

For a more detailed quantitative comparison of the solution accuracy Figure 2 shows the pressure distribution at the leeward side stagnation area. The comparison between the different computations shows clearly the improved conservation properties of the present scheme compared to the standard Roe solution for the leeward side stagnation point. This capability of the present scheme is expected to have a strong effect on the prediction accuracy for the location of pressure induced separation for more realistic flow configurations (e.g. airfoils).

3.2. SHEAR LAYER ON DISTORTED GRID

In this test case a straight subsonic shear layer is computed on a smooth grid that models a sinus bump. The grid lines are aligned to the shear flow only upstream of the bump, in the region of the bump the shear is oblique to the grid with grid lines passing through the shear layer under a small angle. As mentioned in the first section, standard non-multi-dimensional solvers are in general very sensitive to such flow phenomena, the multi-dimensional capabilities of the present scheme is expected to result in a clear improvement for this flow case.

Figure 3 shows the grid that was used for the computations with the present scheme. The grids for the Roe scheme are two, four and eight times refined in both directions. In Figure 4 the velocity profiles in the outflow plane are compared. The Roe results show improved resolution of the shear with increasing grid density, but even the Roe solution with 272*136 grid points is not able to reproduce the accuracy of the solution of the present scheme on the 34*17pts grid.

3.3. SUPERSONIC WEDGE-FLOW

The conservative variant of the present scheme, as discussed in detail in (Schwane, 1999), is applied to the flow over a 15 degree compression ramp. First order polynomials are used for the present scheme. The results are compared to two computations with a standard Roe scheme of first and second order. Figure 5 shows the isobars and velocity vectors.

The solutions from the second order Roe scheme and the first order present scheme are very similar with two interior points in the shock, with slightly reduced waviness of the isobars in the shock due to the multi-dimensional properties of the present scheme.

3.4. COMPUTATIONAL EFFICIENCY

In this section we compare the computational efficiency of the present scheme to the 2nd order Roe code that was used for the validation of the results so far.

The Roe solver is explicit in time with a two-step Runge-Kutta scheme, employs second order boundary conditions and for the subsonic test cases is used without any entropy correction. Additionally limiters are set to one for second order spatial accuracy in the entire computational domain. All these settings increase the accuracy of the Roe scheme, but have an adverse effect on the convergence rates.

Figure 6 shows the comparison of the residual versus the CPU time for two calculations for the flow past a cylinder and the oblique shear flow, demonstrating the higher efficiency of the present scheme compared to the Roe solver. For the flow past the infinite cylinder, the present scheme achieves about three times the efficiency of the Roe scheme, however, this gain can be reduced by more efficient relaxation schemes for the Roe solver.

For the shear layer on a distorted grid, however, the gain in accuracy for the multidimensional scheme is significant with still more accurate results on a four times recoarsened grid in each grid direction that the present scheme is more efficient, independently from the relaxation methods used in each scheme. Note that in Figure 6 the Roe solution on the coarsest grid is compared to the present solution, whereas, the Roe scheme on the finest

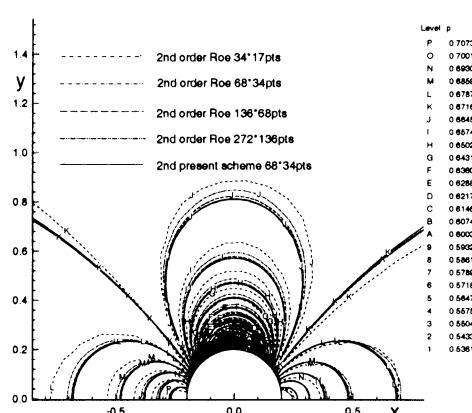


Figure 1. Pressure contours - flow past infinite cylinder

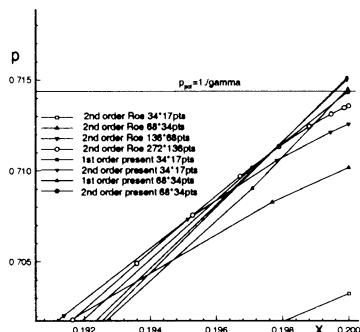


Figure 2. Zoom of Pressure distribution along contour - leeward stagnation region

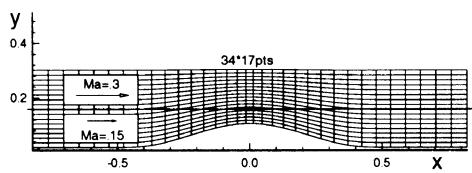


Figure 3. Grid for Shear flow past sinus bump

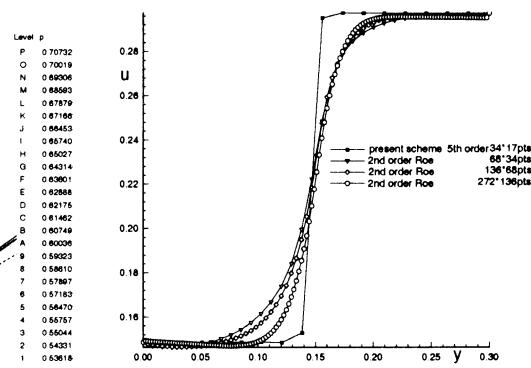


Figure 4. Velocity Profiles in outflow plane

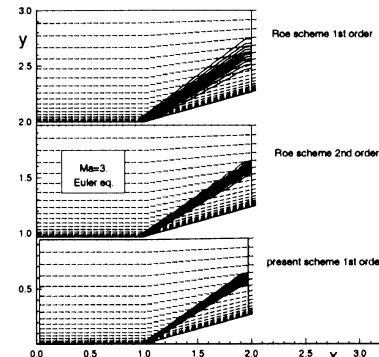


Figure 5. Lines of constant Pressure - Flow over 15 degree wedge

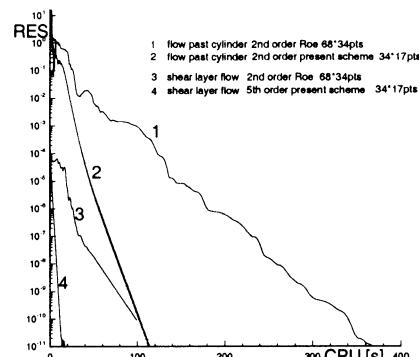


Figure 6. Comparison convergence rate

grid with four times more grid points in each direction, which would lead to an increase of the computational effort by a factor 4^3 , delivers results with still larger errors than the ones from the present scheme on the coarsest grid.

4. Conclusions and Future Work

A numerical scheme for the solution of the Euler equations has been presented that permits to resolve a certain amount of substructure of the solution in the interior of grid cells. This generalisation of the solution allows to avoid the assumption of constant distributions of flow variables and fluxes along cell edges, as commonly used for flux balance formulations.

If the edge distributions, that can be determined by the present scheme with any desired spatial accuracy, are taken into account for the flux integration the accuracy of the resulting global solution is significantly improved.

The scheme has been extended to the Navier Stokes equations, some additional effort had to be spent on the properties of the numerical method for areas of small flow velocities (separated regions), due to the fact that the definition of a stream line angle becomes arbitrary for vanishing velocity components.

For future work two properties of the scheme have to be further explored:

- 1 the inherent freedom in the choice of the local approximation order
- 2 the existence of a unique measure for the local discretization error of the scheme based on the local flow solution that could be used as criterion for the choice of the local approximation order

Both features combined could be used for the design of a scheme with local adaptation of the approximation order, that preserves the simplicity in terms of data and code structure of a non-adaptive scheme.

References

- Deconinck H, Struijs R, Bourgois G and Roe P L High resolution shock capturing cell vertex advection schemes on unstructured grid. VKI LS 1994-05, Computational Fluid Dynamics, March 1994.
- Powell K , van Leer B : A Genuivly Multi-Dimensional Upwind Cell-Vertex Scheme for the Euler-Equations. NASA TM 102029 ICOMP-89-13.
- Roe P L: Multi-dimensional Upwinding: Motivation and Concepts. VKI LS 1994-05, Computational Fluid Dynamics, March 1994.
- Schwane R, Haenel D: An implicit flux-vector splitting scheme for the computation of viscous hypersonic flow. AIAA-paper No. 89-0274, (1989).
- Schwane R: A Multi-dimensional Euler Solver. ESTEC Report: MPA/2020/RS, March 23, 1999.
- Walpot L: Schwane, R.;Bakker, P.G.: Validation of High Resolution Schemes in Flows with Viscous Inviscid Interactions and Non-equilibrium. 12th AIAA Applied Aerodynamics Conference, June 1994, Colorado Springs, CO.

MPDATA—A MULTIPASS DONOR CELL SOLVER FOR GEOPHYSICAL FLOWS

P. K. SMOLARKIEWICZ

*National Center for Atmospheric Research,
Boulder, CO 80307, USA.*

Email: smolar@ncar.ucar.edu

AND

L. G. MARGOLIN

*Los Alamos National Laboratory,
Los Alamos, NM 87545, USA.*

Email: len@lanl.gov

Abstract. This article is a summary of MPDATA, a class of methods for the numerical simulation of fluid flows based on the sign-preserving properties of upstream differencing. MPDATA was designed originally as an inexpensive alternative to flux-limited schemes for evaluating the advection of nonnegative thermodynamic variables (such as liquid water or water vapor) in atmospheric models. During the last decade, MPDATA has evolved from a simple advection scheme to a general approach for integrating the conservation laws of geophysical fluids on micro-to-planetary scales. The purpose of this paper is to outline the basic concepts leading to a family of MPDATA schemes, highlight existing MPDATA options, and to comment on the use of MPDATA to model complex geophysical flows.

1. Introduction

MPDATA—Multidimensional Positive Definite Advection Transport Algorithm; (Smolarkiewicz, 1983), (Smolarkiewicz, 1984)—borrows from the philosophy of Godunov schemes in that it relies on the remarkable properties of upstream differencing. Originally (Smolarkiewicz, 1983), MPDATA was designed as a simple scheme for handling the transport of nonnegative thermodynamic variables (such as water vapor) in atmospheric models.

Even in its most elementary form MPDATA is second-order accurate, sign-preserving, and conservative, while computationally efficient. It is iterative in nature. The first pass is a simple donor cell approximation, which is positive definite but only first-order accurate. The second pass increases the accuracy of the calculation by estimating and compensating the (second-order) truncation error of the first pass. Additional passes can be used to estimate the residual error of the previous pass and approximately compensate for it, leading to successively more accurate solutions of the advection equation.

Over years [see (Smolarkiewicz and Margolin, 1998), for a recent review], MPDATA has been extended to a fully monotone (in the sense of FCT) scheme for advection-diffusion equations in an arbitrary curvilinear framework, and more recently to systems with arbitrary right-hand sides. The efficacy of MPDATA as a general solver for complex fluid problems has been demonstrated in the context of atmospheric dynamics for both compressible and incompressible systems of equations. MPDATA has also been used as an interpolating routine to derive a class of sign-preserving semi-Lagrangian fluid models congruent to the Eulerian MPDATA models.

Within the last decade, MPDATA has evolved from a simple advection scheme to a general approach for integrating the conservation laws of geophysical fluids on micro-to-planetary scales (Smolarkiewicz and Margolin, 1997), (Smolarkiewicz et. al., 1998). In consequence, MPDATA embodies a family of schemes of varying accuracy and levels of complexity. The purpose of this paper is to outline the basic concepts underlying MPDATA schemes, highlight existing options, as well as to comment on the use of MPDATA to model geophysical fluids. Our results of large-eddy simulations of boundary layers (viz., micro scale), as well as modeling Earth climate (viz., global scale) illustrate not only the generality of the approach, but also its advantages compared to more traditional integration methods.

2. Basic MPDATA

The model advection equation for a scalar variable Ψ in one dimension is

$$\frac{\partial \Psi}{\partial t} = -\frac{\partial}{\partial x}(u\Psi) , \quad (1)$$

where the velocity u may vary in space and time. The donor cell (or upstream) approximation to the advection equation (1) is written in flux form

$$\Psi_i^{n+1} = \Psi_i^n - [F(\Psi_i^n, \Psi_{i+1}^n, U_{i+1/2}) - F(\Psi_{i-1}^n, \Psi_i^n, U_{i-1/2})] , \quad (2)$$

where the flux function F is defined in terms of the local Courant number U by

$$F(\Psi_L, \Psi_R, U) \equiv [U]^+ \Psi_L + [U]^- \Psi_R \quad (3a)$$

$$U \equiv \frac{u\delta t}{\delta x} ; [U]^+ \equiv 0.5(U + |U|) ; [U]^- \equiv 0.5(U - |U|) \quad (3b, c, d)$$

The integer and half integer indices correspond to the cell centers and cell walls, respectively. Here δt is the computational time step, and δx is the length of a cell.

Assume for simplicity that the velocity is constant and Ψ nonnegative. A simple truncation analysis, expanding about the time level n and spatial point i , shows that (2) more accurately approximates the advection-diffusion equation

$$\frac{\partial \Psi}{\partial t} = -\frac{\partial}{\partial x}(u\Psi) + \frac{\partial}{\partial x}\left(K\frac{\partial \Psi}{\partial x}\right), \quad (4)$$

where

$$K = \frac{(\delta x)^2}{2\delta t}(|U| - U^2). \quad (5)$$

Thus (2) approximates the solution to the advection equation with a second-order error. To improve the accuracy, it is necessary to construct an estimate of the error and subtract it from (2). The classical one-step Lax-Wendroff scheme is perhaps the most familiar example of such a procedure, using standard centered differences to approximate the second term on rhs of (4). While MPDATA derives from the same general concept, it exploits special properties of the donor cell scheme for approximating and compensating the error.

The donor cell scheme (2) is positive definite for any velocity field and is monotone if the velocity field is constant in space, providing that the Courant number is properly bounded. These properties are lost in any linear combination of donor cell and centered differencing (Godunov, 1959). In these terms, the basic idea underlying all MPDATA schemes can be stated very simply—use a donor cell approximation to the error term. Since the error term is not written in a form to do this directly, it is first rewritten as

$$\text{error}^{(1)} = \frac{\partial}{\partial x}(v^{(1)}\Psi), \quad (6)$$

where

$$v^{(1)} \equiv \frac{(\delta x)^2}{2\delta t}(|U| - U^2)\frac{1}{\Psi}\frac{\partial \Psi}{\partial x}. \quad (7)$$

is a pseudo velocity. The superscript (1) shows that it is the first approximation to the error. Inside the derivative in (6), the diffusive flux in the second term of (4) is multiplied by a factor of Ψ over Ψ —i.e., by unity. However, in the donor cell approximation to (6), the factor in the numerator will be represented using an upstream value whereas the factor in the denominator will be approximated using a centered value. In this way, a nonlinearity is

introduced and a higher-order approximation is found that still preserves positivity.

To compensate for the error between the donor cell solution $\Psi^{(1)}$ and a second-order accurate solution Ψ^{n+1} , we use the error (6) estimated at time level $n + 1$. A first-order accurate estimate of the pseudo velocity (nondimensionalized for convenience) is

$$V_{i+1/2}^{(1)} \equiv (|U| - U^2) \frac{\Psi_{i+1}^{(1)} - \Psi_i^{(1)}}{\Psi_{i+1}^{(1)} + \Psi_i^{(1)}} \equiv (|U| - U^2) A_{i+1/2}^{(1)}, \quad (8)$$

where

$$V^{(1)} = \frac{v^{(1)} \delta t}{\delta x}. \quad (9)$$

In the second pass, we subtract a donor cell estimate of the error to improve the order of the approximation. The equation of the second pass is

$$\Psi_i^{(2)} = \Psi_i^{(1)} - [F(\Psi_i^{(1)}, \Psi_{i+1}^{(1)}, V_{i+1/2}^{(1)}) - F(\Psi_{i-1}^{(1)}, \Psi_i^{(1)}, V_{i-1/2}^{(1)})], \quad (10)$$

which estimates Ψ^{n+1} to the second-order while preserving the sign of Ψ . Note that the stability of the first pass ensures that of the second pass, since $|U| \leq 1 \implies -1 \leq |U| - U^2 \leq 1$ and the assumed nonnegativity of Ψ together with the positivity of the donor-cell scheme assure $|A_{i+1/2}^{(1)}| \leq 1 \forall i$.

The two-pass scheme described above is the most elementary MPDATA; for its illustrations, see (Smolarkiewicz and Margolin, 1998) and references therein. Equation (10) again can be expanded in a Taylor series, the residual error after the second pass estimated as in (6)-(8), and compensated as in (10). The entire process of estimating the residual error and compensating it can be continued, iteration after iteration, reducing the magnitude of the truncation error, which remains however at third order¹. It is worth noting that writing a computer program for such a procedure is extremely simple, as the flux function, the donor cell scheme itself, and the form of the pseudo velocity remain the same in each iteration.

3. General Fluid Algorithm

In modeling atmospheric/oceanic flows—viz. high Reynolds' number low Mach number flows—the governing equations can be always viewed in the form of a generalized transport equation

$$\frac{\partial G\Psi}{\partial t} + \nabla \cdot (\mathbf{v}\Psi) = GR. \quad (11)$$

¹For a discussion of the third-order accurate MPDATA see (Margolin and Smolarkiewicz, 1998)

Here, $G = G(\mathbf{x})$, $\mathbf{v} = \mathbf{v}(\mathbf{x}, t)$, and $R = R(\mathbf{x}, t)$ are assumed to be known functions of the independent variables; G plays the role of the Jacobian of the coordinate transformation from a Cartesian to the curvilinear framework \mathbf{x} (e.g., terrain-following coordinates on a rotating sphere), \mathbf{v} is a generalized “advection” velocity vector $\mathbf{v} = G\dot{\mathbf{x}}$, and R combines all forcings and/or sources. In general, both \mathbf{v} and R are functionals of the dependent variables; see (Smolarkiewicz and Margolin, 1998) for examples.

Generalization of the elementary scheme (section 2) towards an algorithm suitable for integrating (11) invokes extensions to multiple dimensions, variable flow fields, diffusion-advection equations, transporting fields of variable sign, non-Cartesian geometries, inhomogeneous transport problems with arbitrary right hand sides, as well as third-order-accurate and fully monotone transport schemes. These extensions embody nontrivial technical development and contain some subtleties originating, e.g., in the cross derivatives or nonvanishing advective fluxes of the right hand side that appear in the truncation-error analysis. Nonetheless, the overall idea and design of the general algorithm closely follow the approach underlying the elementary scheme; cf. (Smolarkiewicz and Margolin, 1998) for a discussion. Regardless of the complexity of a fluid problem at hand, all MPDATA schemes retain the form of the basic scheme where all subsequent iterations are standard donor cell scheme but with different arguments at each iteration. The first iteration uses physical advective velocity and the transported field, whereas following iterations use pseudo velocities derived from the truncation error analysis, and fields evaluated from the preceding iterations.

A fully second-order accurate MPDATA algorithm, suitable for solving both elastic and anelastic fluid problems with (11), can be compactly written as

$$\Psi_i^{n+1} = MPDATA(\Psi^n + 0.5\delta t R^n, \mathbf{V}^{n+1/2}, G) + 0.5\delta t R_i^{n+1} \quad (12)$$

where MPDATA symbolizes the generalized homogeneous-transport algorithm [section 3.2 in (Smolarkiewicz and Margolin, 1998)], and $\mathbf{V}^{n+1/2}$ is an $\mathcal{O}(\delta t^2)$ estimate of the advective velocity vector at intermediate time level $n + 1/2$ [section 3.4 in (Smolarkiewicz and Margolin, 1998)]. Advecting the auxiliary field $\Psi^n + 0.5\delta t R^n$ not only compensates the truncation error due to the source term but also has the physical interpretation of integrating the forces along a parcel trajectory rather than at the grid point. This makes (12) congruent to semi-Lagrangian approximations and facilitates unified fluid models that integrate the equations of motion, optionally, in the Eulerian (point-wise) or Lagrangian (trajectory-wise) sense (Smolarkiewicz and Margolin, 1997). Note that (12) may be viewed as a paraphrase of Strang splitting (Strang, 1968).

In the literature, there are numerous examples of successful applications of MPDATA to modeling flows on scales from micro to planetary. Over the last decade, MPDATA has been frequently compared with other transport schemes, primarily in the context of passive scalar advection. The assessments of MPDATA's relative strengths and weaknesses reported in the literature depend very much on the schemes included in comparisons, choice of test problems, MPDATA's options, and details of implementation. The most common complaints are that the basic MPDATA is too diffusive, and enhanced MPDATA is too expensive. The most often acknowledged virtues are MPDATA's multidimensionality, robustness, and its underlying simplicity. These advantages carry over to more complex geophysical fluid models; note however that the relative efficiency of advection becomes less important with increasing complexity of the models (Smolarkiewicz and Margolin, 1997), (Smolarkiewicz et. al., 1998), (Smolarkiewicz and Margolin, 1998).

Passive-scalar advection provides only a "static" framework for evaluating behavioral errors of the transport algorithms. In (Margolin et. al., 1999), the authors employed a series of large eddy simulations (LES) of a convective boundary layer to evaluate viscous properties of the fluid solver (12) based solely on the basic second-order-accurate MPDATA. They have shown that in the presence of an explicit subgrid-scale turbulence model, MPDATA reproduces both the benchmark results, generated with a centered-in-time-and-space (CTS) fluid solver, and the available data. Thus, the MPDATA's implicit viscosity is negligible compared to explicit viscosity of the subgrid-scale model. In contrast, in the absence of an explicit subgrid-scale model, their results show that the MPDATA's implicit viscosity serves as an effective "turbulence" model. However, the spectral fall-off in the inertial subrange is less steep than the $-5/3$ consistent with Kolmogorov's law (obtained in the experiment with the explicit subgrid-scale model included) with slope approximately -1 . Also, in consequence of the inherent nonlinearity of MPDATA—common to all higher-order nonoscillatory schemes (Godunov, 1959)—its implicit viscosity is self-adaptive rather than additive (Margolin et. al., 1999).

Present day geophysical fluid models most often use a CTS approach to discretize the dynamics. To mitigate spurious effects due to negative undershoots in the thermodynamic variables, these models usually adopt a "hybrid" approach where different variables are transported with different advection schemes, or even the same variable uses different advection schemes (operators) in the horizontal and the vertical. Here, the genuine multidimensionality and general applicability of MPDATA allow a single scheme for all dependent variables, thus minimizing auxiliary computations. Furthermore, MPDATA's strong (nonlinear) stability—common to all con-

servative sign preserving advection schemes—permits more liberal stopping criteria in iterative elliptic solvers and allows dispensing with various filtering operations often required to stabilize geophysical fluid models. As a result, fluid models based solely on MPDATA appear competitive when compared to established codes of the same category (Smolarkiewicz and Margolin, 1997).

The remarkable efficacy of MPDATA-based fluid solvers becomes especially visible in (Smolarkiewicz et. al., 1999), where (12) has been employed, in the spirit of LES, to model idealized climates of abstract aquaplanets whose radii decrease successively by a factor of 10 (while the Rossby number remains fixed). These calculations spawn a range of fluid motions with the condition number varying by the orders of magnitude and a variety of the flow regimes.

Acknowledgements

The authors gratefully acknowledge partial support of US Department of Energy through the Climate Change Prediction Program while conducting this work.

References

- Godunov S K (1959). Finite Difference Methods for Numerical Computations of Discontinuous Solutions of Equations of Fluid Dynamics. *Mat. Sb.* **47**, pp 271-306.
- Margolin L G and Smolarkiewicz P K (1998). Antidiffusive Velocities for Multipass Donor Cell Advection. *SIAM J. Sci. Comput.* **20**, pp 907-929.
- Margolin L G, Smolarkiewicz P K and Sorbjan Z (1999). Large-eddy Simulations of Convective Boundary Layers Using Nonoscillatory Differencing. *Physica D* **133**, pp 390-397.
- Smolarkiewicz P K (1983). A Simple Positive Definite Advection Scheme with Small Implicit Diffusion. *Mon. Wea. Rev.* **111**, pp 479-486.
- Smolarkiewicz P K (1984). A Fully Multidimensional Positive Definite Advection Transport Algorithm with Small Implicit Diffusion. *J. Comput. Phys.* **54**, pp 325-362.
- Smolarkiewicz P K and Margolin L G (1997). On Forward-in-Time Differencing for Fluids: An Eulerian/Semi-Lagrangian Nonhydrostatic Model for Stratified Flows. *Atmos. Ocean Special* **35**, pp 127-152.
- Smolarkiewicz P K, Grubišić V and Margolin L G (1998). Forward-in-Time Differencing for Fluids: Nonhydrostatic Modeling of Rotating Stratified Flows on a Mountainous Sphere. Numerical Methods for Fluid Dynamics VI, pp 507-513. Baines M J (Editor), Will Print, Oxford.
- Smolarkiewicz P K and Margolin L G (1998). MPDATA: A Finite-Difference Solver For Geophysical Flows. *J. Comput. Phys.* **140**, pp 459-480.
- Smolarkiewicz P K, Grubišić V, Margolin L G and Wyszogrodzki A A (1999) Forward-in-Time Differencing for Fluids: Nonhydrostatic Modeling of Fluid Motions on a Sphere. Recent Developments in Numerical Methods for Atmospheric Modelling, pp 21-43. ECMWF, Reading, UK.
- Strang G (1968). On the Construction and Comparison of Difference Schemes. *SIAM J. Numer. Anal.* **5**, pp 506-517.

ON THE HYPERBOLIC NATURE OF TWO-PHASE FLOW EQUATIONS: CHARACTERISTIC ANALYSIS AND RELATED NUMERICAL METHODS

H. STÄDTKE, B. WORTH, G. FRANCHELLO

Commission of the European Communities

Joint Research Centre - Ispra Establishment

I-21020 Ispra (Varese), Italy

email: herbert.staedtke@jrc.it

Abstract

This paper describes a new approach to the numerical simulation of transient two-phase flow based on a fully hyperbolic two-fluid model of two-phase flow using separated conservation equations for the two phases. Features of the new model include the existence of real eigenvalues, and a complete set of independent eigenvectors which can be expressed algebraically in terms of the major dependent flow parameters. This facilitates the application of numerical techniques specifically developed for high speed single-phase gas flows which combine signal propagation along characteristic lines with the conservation property with respect to mass, momentum and energy. Advantages of the new model for the numerical simulation of one- and two-dimensional two-phase flow are discussed.

1. Introduction

The numerical simulation of gas-liquid two-phase flow processes remains a challenging task, caused mainly by the inhomogeneous nature of the two-phase mixture and the resulting strong thermal and mechanical non-equilibrium conditions. A further difficulty arises from the large variety of different phase distribution patterns (often called flow regimes) which can occur in nature - a fact which has largely hampered the development of a generalized modelling basis and related numerical solution strategies. From a purely theoretical point of view, the classical Navier-Stokes equations are fully applicable for the specific flow domain where either liquid or vapour is present, and together with the jump conditions at the moving interface,

the whole flow processes is completely determined. However at present, and possibly also in the near future, there is little chance for such "direct numerical solution" of flow processes of industrial interest.

The approach described herein starts from an extended "two-fluid" representation of the two-phase flow field where both phases are treated as interpenetrating continua with source terms representing the interfacial transport processes for mass, momentum and energy. Such a model results from a space/time or ensemble-averaging of the basic Navier-Stokes equations. As is the case for all averaging procedures, information on local flow processes, in particular at the interface separating the two phases, is lost and has to be compensated for by additional modelling.

In the present approach, additional models for non-viscous interfacial forces have been included for the description of virtual mass effects and "lift forces". This results in a fully hyperbolic system of governing equations which transforms the set of time-dependent two-phase flow equations into a well-posed initial-boundary value problem. Crucial new aspects of this hyperbolic model of two-phase flow are the existence of real eigenvalues and a corresponding set of linearly-independent eigenvectors, each expressible as algebraic functions of the major dependent flow parameters. This facilitates the application of generalized Godunov-type numerical schemes such as Approximate Riemann Solvers and Flux Vector Splitting techniques which make explicit use of the hyperbolicity of the flow equations.

2. Two Fluid Model

The entire flow domain is assumed to be divided into sub-domains wherein either only liquid or only vapour is present. The two regions are each separated by an assumed infinitesimally thin layer, in the following referred to as the 'interface'. This situation can be described by a distribution function, γ_i , with $i = g$ (gas) or $i = l$ (liquid), defined as

$$\begin{aligned}\gamma_g &= 1, \gamma_l = 0 && \text{where only gas/vapour is present} \\ \gamma_l &= 1, \gamma_g = 0 && \text{where only liquid is present}\end{aligned}$$

with

$$\gamma_g + \gamma_l = 1. \quad (1)$$

Then, for a small finite volume V_ξ (see Fig. 1), the volumetric concentration of the phase i is

$$\alpha_i = \frac{1}{V_\xi} \int_V \gamma_i dV \quad (2)$$

where α_g is known as the void fraction. The gradient of the distribution

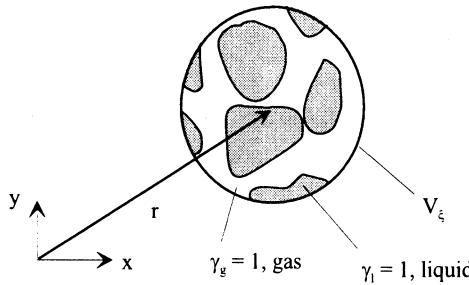


Figure 1: Control volume for space averaging

function γ_i represents a vector field in which $\nabla \gamma_i = 0$ everywhere except at the interface, and where $\nabla \gamma_i$ represents a vector directed into the phase i normal to the interface with an absolute value

$$|\nabla \gamma_i|^{int} \rightarrow \infty. \quad (3)$$

A unit normal vector at the interface directed outwards from the phase i is therefore defined as

$$\vec{n}_i^{int} = -\frac{\nabla \gamma_i^{int}}{|\nabla \gamma_i|^{int}}. \quad (4)$$

Assuming that $|\nabla \gamma_i|^{int}$ has the property of the Dirac delta-function, the interfacial area concentration in a small control volume V_ξ can be expressed as an integral over the space-fixed volume V_ξ

$$a^{int} = \frac{1}{V_\xi} \int_V |\nabla \gamma_i|^{int} dV. \quad (5)$$

Both the void fraction, α_g , as well as the interfacial area concentration, a^{int} , defined by equations (2) and (5), are parameters which essentially characterize the nature of the local two-phase mixture.

2.1 CONSERVATION EQUATIONS

Nearly all current two-phase flow models of practical interest are restricted to averaged flow parameters rather than local flow quantities. The corresponding "macroscopic" separate balance equations for the two phases are obtained by a space and/or time, or ensemble averaging, of the local instantaneous phasic flow equations which leads to what is often referred to as the two-fluid model of two-phase flow. The development of such two fluid models is largely related to the pioneering work of Ishii (Ishii, 1975),

Boure (Boure, 1975), Delhaye and Achard (Delhaye & Achard, 1976), and Drew and Lahey (Drew & Lahey, 1979).

The form of the balance equations presented here has been derived using the concept of distribution functions and the Dirac delta-function as first applied by Gray and Lee (Gray and Lee, 1977) for the volume-averaging of multi-phase systems. With the index $i = g$ (gas or vapour), and $i = l$ (liquid), these balance equations can be written as

mass:

$$\frac{\partial}{\partial t} (\alpha_i \varrho_i) + \nabla \cdot (\alpha_i \varrho_i \vec{v}_i) = \sigma_i^M \quad (6)$$

$$\text{with } \sum_{i=g,f} \sigma_i^M = 0$$

momentum:

$$\begin{aligned} \frac{\partial}{\partial t} (\alpha_i \varrho_i \vec{v}_i) + \nabla \cdot (\alpha_i \varrho_i \vec{v}_i \vec{v}_i) + \alpha_i \nabla p_i = \\ \vec{F}_i^{int} + \vec{F}_i^{ext} + \sigma_i^M \vec{v}_i^{ex} + \nabla \cdot (\alpha_i \bar{\mathbf{T}}_i) \end{aligned} \quad (7)$$

$$\text{with } \sum_{i=g,f} \vec{F}_i^{int} = 0$$

energy:

$$\begin{aligned} \frac{\partial}{\partial t} \left[\alpha_i \varrho_i \left(u_i + \frac{v_i^2}{2} \right) \right] + \nabla \cdot \left[\alpha_i \varrho_i \vec{v}_i \left(h_i + \frac{v_i^2}{2} \right) \right] \\ \nabla (\alpha_i \vec{q}_i) - \nabla (\alpha_i \bar{\mathbf{T}}_i \cdot \vec{v}_i) = \sigma_i^M \left(h + \frac{v_i^2}{2} \right) \\ + \sigma_i^Q + Q_i^{ext} + \vec{F}^{ext} \cdot \vec{v}_i + \vec{F}^{int} \cdot \vec{v}_i^{Fi} \end{aligned} \quad (8)$$

$$\text{with } \sum_{i=g,f} \sigma_i^Q = 0.$$

Using the momentum equation, the kinetic terms can be removed from the energy equations. Assuming further that the following thermodynamic relationship is valid also for the average quantities

$$T_i \delta s_i = \delta u_i - \frac{p_i}{\varrho_i^2} \delta \varrho_i, \quad (9)$$

the energy equation can be simplified by introducing entropy as a new state variable which results in the following balance equation for the phasic entropy:

entropy:

$$\frac{\partial}{\partial t}(\alpha_i \varrho_i s_i) + \nabla \cdot (\alpha_i \varrho_i \vec{v}_i s_i) = \sigma_i^S \quad (10)$$

with

$$\begin{aligned} \sigma_i^S &= \frac{\sigma_i^Q}{T_i} + \frac{Q_i^{ext}}{T_i} + \sigma_i^M s_i + \frac{F_i^{fr}}{T_i}(\vec{v}^{int} - \vec{v}_i) \\ &+ \frac{\sigma_i^M}{T_i} \left[h^{ex} - h_i + \frac{1}{2}(\vec{v}^{ex} - \vec{v}_i)^2 \right]. \end{aligned} \quad (11)$$

In compliance with the second law of thermodynamics, the sum over all the entropy sources must be definite positive

$$\sum_{i=g,f} \sigma_i^S \geq 0.$$

The entropy equation (10) does not give any additional information compared with the energy equations (8), however, due to its more simplified form, it might be worthwhile using the entropy equation instead of the complete energy balance equation for the characteristic analysis of the governing equations.

Using the transport properties of the absolute value of the phasic distribution function $|\gamma_i|$ an additional balance equation for the interfacial area concentration can be derived (Stadtke & Holtbecker, 1991);

interfacial area

$$\frac{\partial}{\partial t}(a^{int}) + \nabla \cdot (a^{int} \vec{v}^{int}) = \sigma^A \quad (12)$$

where the source term σ^A describes the creation (or destruction) of interfacial area due to pressure changes (expansion, compression), phase changes (evaporation, condensation), particle break-up or coalescence, and flow regime transitions.

In most practical applications of the two-fluid model, it is further assumed that both phases have locally the same pressure ($p_g = p_l = p$) and that all source terms on the right-hand sides of the equations (6) to (11) and (12) are algebraic functions of the flow and state parameters of the two phases; an assumption which leads to the "Wallis Model" for inhomogeneous two-phase flow.

The basic equations of the "Wallis Model" are of mixed hyperbolic-elliptic type and are known to have two complex-conjugate eigenvalues with the following consequences: (1) the model does not represent a "well-posed" initial-boundary value problem; (2) the system of equations cannot be transformed into the characteristic form and, therefore, all numerical techniques which make use of the hyperbolic wave-like character of the flow equations cannot be applied; (3) the model does not realistically describe pressure wave phenomena, and so cannot provide realistic critical flow predictions; (4) high wave-number instabilities require specific damping terms in the numerical algorithm and this introduces strong numerical diffusion and artificial viscosity effects, making the solution unphysical.

There are basically two different ways to transform the system of equations for the two-fluid model into a well-posed hyperbolic set:

1. To allow individual pressure values for each of the two phases and to provide additional equations to describe the dynamics of the pressure differential between the phases. This approach results in an increased number of balance equations with the added difficulty of providing an increased number of constitutive relations needed to close the system of equations.
2. To stay with the assumption of equal phasic pressures in the localized domain, but to introduce additional terms in the two momentum equations including time and space derivatives of the average flow and state parameters. These terms can be justified as additional interfacial forces (e.g. "virtual mass" force, Basset force); however, closed analytical expressions could be derived only for specific flow regimes with certain idealized assumptions.

Both methods seek to recover, in a more or less heuristic way, some of the information of the complex local flow processes at the interface which is lost during the averaging procedure. The present approach follows the second assumption which is somewhat easier to implement for practical applications.

2.2 INTERFACIAL MOMENTUM COUPLING

The interfacial forces, introduced in equation (7), can be split into two parts: the interfacial friction force \vec{F}_i^{fr} and the non-viscous forces \vec{F}_i^{nv} including virtual mass effects and lift forces (Drew *et al*, 1979).

$$\vec{F}_i^{in} = \vec{F}_i^{fr} + \vec{F}_i^{nv}. \quad (13)$$

Apart from the non-viscous part of the interfacial forces, \vec{F}_i^{nv} , all the other source terms on the right-hand side of the conservation equations are assumed to be algebraic functions of flow and state parameters of the two phases. It is further assumed that the non-viscous forces do not contribute to the dissipation of mechanical energy.

The criteria used to determine an expression for the non-viscous part of the interfacial forces include:

- the non-viscous interfacial friction terms should not affect the sum of the momentum equations
- the non-viscous interfacial terms do not contribute to the entropy source for the individual phases
- the coefficient matrix should have only real eigenvalues
- the characteristic velocities (eigenvalues) should have physical meaning
- the coefficient matrix should have a complete set of independent eigenvectors
- the system of equations should include, as limiting cases, the single phase flow of gas/vapour ($\alpha_g \rightarrow 1$) or liquid ($\alpha_g \rightarrow 0$), and the homogeneous flow ($v_g = v_l$) condition
- the system of equations should yield implicit "reasonable" values for the two-phase sound velocity in agreement with existing experimental data.

As a result of a large number of analytical trials, the following rather general form has been derived (Stadtke & Holtbecker, 1991):

$$\begin{aligned} \vec{F}_i^{nv} = & \pm \alpha_g \alpha_l \left[k \rho \left(\frac{d^l \vec{v}_g}{dt} - \frac{d^g \vec{v}_l}{dt} \right) + c (\vec{v}_g - \vec{v}_l) \nabla \cdot (\vec{v}_g - \vec{v}_l) \right] \\ & \mp \alpha_g \alpha_l (\vec{v}_g - \vec{v}_l) \left[e (\vec{v}_g - \vec{v}_l) \nabla \cdot \alpha_g + r_g \frac{d^g \rho_g}{dt} + r_l \frac{d^l \rho_l}{dt} \right] \end{aligned} \quad (14)$$

with the total derivative operators

$$\frac{d^l}{dt} = \frac{\partial}{\partial t} + \vec{v}_l \cdot \nabla, \quad \frac{d^g}{dt} = \frac{\partial}{\partial t} + \vec{v}_g \cdot \nabla,$$

and ∇ the usual differential operator $\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right)$ acting on the variables \vec{v}_g , \vec{v}_l , $(\vec{v}_g - \vec{v}_l)$, α_g , ρ_g or ρ_l .

The first term in equation (14) represents the "virtual mass" force in the "objective form" as proposed by Drew et al. (Drew et al, 1979). This force accounts for the reaction resulting from the displacement of adjacent fluid masses in the case of relative acceleration between the phases. Following the above mentioned guidelines, the parameters c , e , r_g and r_l have been determined as follows

$$\left. \begin{aligned} c &= -(\alpha_g \rho_l - \alpha_l \rho_g) & e &= -\frac{\partial c}{\partial \alpha_g} = (\rho_l + \rho_g) \\ r_g &= \frac{\alpha_g}{\rho_g} e & r_l &= \frac{\alpha_l}{\rho_l} e. \end{aligned} \right\} \quad (15)$$

The "virtual mass" coefficient k can be used to adjust the strength of the interfacial momentum coupling with respect to different flow regimes. For idealized dispersed droplet or bubbly flow, a value of $k = 0.5$ can be derived from classical potential flow analysis whereas for completely separated flows (e.g. stratified flow) it is expected that k tends to zero. In the case of churn-turbulent two-phase flow conditions with strong interfacial momentum coupling, a value of $k \gg 1$ might be appropriate.

It can be shown that for incompressible flow (phasic sound velocities $a_g \rightarrow \infty$, $a_l \rightarrow \infty$) the present approach is equivalent to a two-pressure model where phase-to-interface pressure differences are given by the algebraic equations

$$\left. \begin{aligned} p_g - p^{int} &= \frac{1}{2} \alpha_l \rho_g (\vec{v}_g - \vec{v}_l)^2 \\ p_l - p^{int} &= \frac{1}{2} \alpha_g \rho_l (\vec{v}_g - \vec{v}_l)^2 \end{aligned} \right\}. \quad (16)$$

2.3 ALGEBRAIC SOURCE TERMS

Apart from the non-viscous interfacial forces, all the other source terms on the right-hand side of the balance equations are assumed to be determined by algebraic functions of flow and state parameters.

For the interfacial friction forces, a general resistance law is applied which can be formulated with respect to the square of the "slip velocity" between the gas and liquid phases, and the interfacial area concentration

$$\vec{F}_{g,l}^{fr} = \pm C_w a^{int} \rho_c |\vec{v}_g - \vec{v}_l| (\vec{v}_g - \vec{v}_l) \quad (17)$$

with a Reynolds number-dependent interfacial friction coefficient

$$C_w = f(\text{Re}).$$

Note that only the frictional part of the interfacial forces contributes to the entropy source term in equation (11).

The source terms for heat and mass transfer between the phases are determined by the sum and the difference of the heat fluxes from the bulk of the corresponding phases to the interface, resulting in

$$\left. \begin{aligned} \sigma_{g,l}^M &= \pm \frac{a^{int}}{\Delta h^s} [h_g^q (T_g - T^s) + h_l^q (T_l - T^s)] \\ \sigma_{g,l}^Q &= \pm a^{int} (h_g^q + h_l^q) (T_g - T_l) \end{aligned} \right\}. \quad (18)$$

2.4 CHARACTERISTIC ANALYSIS OF GOVERNING EQUATIONS

The balance equations (6), (7) and (10), together with the interfacial area transport equation (12), represent a system of partial differential equations which can be combined in compact matrix form as

$$\frac{\partial \mathbf{U}}{\partial t} + \mathbf{G} \cdot (\nabla \mathbf{U}) = \mathbf{C} \quad (19)$$

with the vector of "primitive" parameters

$$\mathbf{U} = \{p, \vec{v}_g, \vec{v}_l, \alpha_g, s_g, s_l, a^{int}\}^T. \quad (20)$$

With an appropriate definition for the "non-viscous" part of the interfacial forces as given by equation (14), only real eigenvalues appear in the coefficient matrix for the convective part of the flow equations, \mathbf{G} , together with a corresponding set of linearly-independent eigenvectors (see Stadtke & Holtbecker, 1991, for more details). A specific feature of the present "hyperbolic" model is that all eigenvalues and eigenvectors can be expressed as algebraic functions of the major dependent flow parameters.

The eigenvalues obtained, being characteristic velocities of the various wave propagation processes, are defined in the n -direction (Fig. 2) as

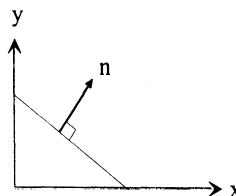


Figure 2: Direction of wave propagation

$$\left. \begin{array}{ll} \lambda_{1,2} = \vec{v}_g \cdot \vec{n}, \vec{v}_l \cdot \vec{n} & \text{void waves} \\ \lambda_{3,4} = \vec{v} \cdot \vec{n} \pm a & \text{pressure/density waves} \\ \lambda_{5,6} = \vec{v}_g \cdot \vec{n}, \vec{v}_l \cdot \vec{n} & \text{shear waves} \\ \lambda_{7,8} = \vec{v}_g \cdot \vec{n}, \vec{v}_l \cdot \vec{n} & \text{temperature/entropy waves} \\ \lambda_9 = \vec{v}^{int} \cdot \vec{n} & \text{interfacial area waves} \end{array} \right\} \quad (21)$$

with the mixture velocity

$$\vec{v} = \frac{\vec{v}_1 + k \frac{\varrho^2}{\varrho_g \varrho_l} \vec{v}_2}{1 + k \frac{\varrho^2}{\varrho_g \varrho_l}},$$

with the volumetric- and mass fraction-averaged velocities

$$\vec{v}_1 = \alpha_g \vec{v}_g + \alpha_l \vec{v}_l, \quad \vec{v}_2 = \frac{\alpha_g \rho_g \vec{v}_g + \alpha_l \rho_l \vec{v}_l}{\alpha_g \rho_g + \alpha_l \rho_l}, \quad (22)$$

and the mixture sound velocity

$$a^2 = \tilde{a}^2 - \Delta a^2 \quad (23)$$

with

$$\begin{aligned} \tilde{a}^2 &= \frac{\alpha_g \varrho_f + \alpha_f \varrho_g}{\frac{\alpha_g \varrho_l}{a_g^2} + \frac{\alpha_f \varrho_g}{a_l^2}} \frac{1 + k \frac{\alpha_g \varrho_g + \alpha_f \varrho_l}{\alpha_g \varrho_f + \alpha_f \varrho_l}}{1 + k \frac{\varrho^2}{\varrho_g \varrho_l}} \\ \Delta a^2 &= \alpha_g \alpha_l \varrho_g \varrho_l (\vec{v}_g - \vec{v}_l)^2 \frac{(\varrho_l + k \varrho)(\varrho_g + k \varrho)}{(\varrho_g \varrho_l + k \varrho^2)^2}. \end{aligned}$$

The strong influence of the virtual mass forces on the sound velocity is shown in Fig. 3 for a water/air mixture at 10 bar pressure. With regard to the virtual mass coefficient k , limiting values are obtained referred to as the "homogeneous equilibrium sound velocity" ($k \rightarrow \infty$) and the "frozen sound velocity" ($k \rightarrow 0$).

Since the convective part of the flow equations (19) is hyperbolic, the coefficient matrix \mathbf{G} can be diagonalized

$$\Lambda = \mathbf{T}^{-1} \mathbf{G} \mathbf{T} \quad (24)$$

to yield the diagonal matrix Λ with eigenvalues λ_k along the principal diagonal, and where the columns of the transformation matrix \mathbf{T} are the right eigenvectors of the coefficient matrix \mathbf{G} .

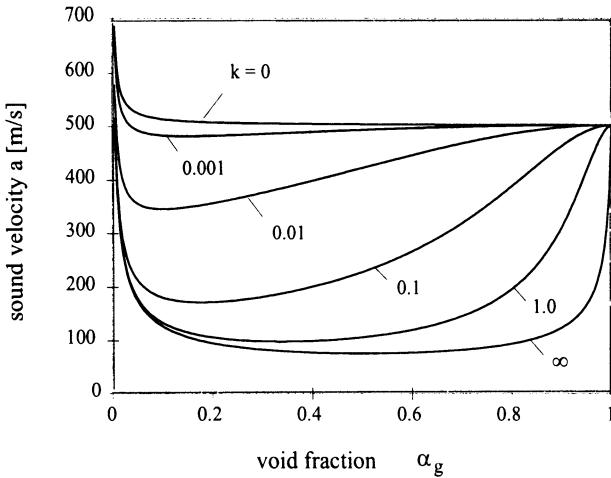


Figure 3: Sound velocity in water/staem mixture at $p = 10$ bar

With respect to the individual characteristic velocities (eigenvalues), the coefficient matrix \mathbf{G} can be split into elementary parts as

$$\mathbf{G} = \sum_{k=1}^9 \mathbf{G}_k$$

with

$$\mathbf{G}_k = \mathbf{T} \boldsymbol{\Lambda}_k \mathbf{T}^{-1}$$

where the diagonal matrix $\boldsymbol{\Lambda}_k$ includes only the k_{th} eigenvalues. Further details of the physical interpretation of the complete eigenspectrum is discussed in (Stadtke & Holtbecker, 1991).

3. Numerical Method

The hyperbolic two fluid model for inhomogeneous, thermal non-equilibrium two-phase flow allows the application of Godunov-type methods which combine finite volume discretization in space with characteristic based upwind techniques. In the present case, a modified Flux Vector Splitting has been used based on the solution of a linearized Riemann problem as is summarized in the following section.

3.1 FINITE VOLUME DISCRETIZATION

For the numerical solution, the governing flow equations are transformed into the conservative form

$$\frac{\partial \mathbf{V}}{\partial t} + \nabla \cdot \mathbf{F} + \mathbf{H}^{nc} \nabla \cdot \mathbf{F} = \mathbf{D} \quad (25)$$

with the state vector of conservative variables, \mathbf{V} , and the corresponding flux vector, \mathbf{F} , given respectively as

$$\mathbf{V} = \begin{pmatrix} \alpha_g \varrho_g \\ \alpha_l \varrho_l \\ \alpha_g \varrho_g \vec{v}_g \\ \alpha_l \varrho_l \vec{v}_l \\ \alpha_g \varrho_g s_g \\ \alpha_l \varrho_l s_l \\ a^{int} \end{pmatrix}, \quad \mathbf{F} = \begin{pmatrix} \alpha_g \varrho_g \vec{v}_g \\ \alpha_l \varrho_l \vec{v}_l \\ \alpha_g \varrho_g \vec{v}_g \vec{v}_g + \alpha_g \mathbf{I}_p \\ \alpha_l \varrho_l \vec{v}_l \vec{v}_l + \alpha_l \mathbf{I}_p \\ \alpha_g \varrho_g s_g \\ \alpha_l \varrho_l s_l \\ a_l^{int} \vec{v}^{int} \end{pmatrix}. \quad (26)$$

The new source term vector, \mathbf{D} , for the conservative variables includes the volume source terms for mass, momentum, energy and interfacial area as already introduced in the general balance equations (6) to (8) and (12).

The 'non-conservative' part of the coefficient matrix \mathbf{H}^{nc} in equation (25), defined as

$$\mathbf{H}^{nc} = (\mathbf{JG} - \mathbf{K}) \mathbf{K}^{-1}, \quad (27)$$

with the Jacobian matrices

$$\mathbf{J} = \frac{\partial \mathbf{V}}{\partial \mathbf{U}}, \quad \mathbf{K} = \frac{\partial \mathbf{F}}{\partial \mathbf{U}}, \quad (28)$$

reflects the fact that the separated momentum equations for the liquid and vapour phases cannot be put into a fully conservative form due to some coupling terms including time and space derivatives.

For the numerical solution scheme, the governing equations (25) are transformed into a finite volume approximation for arbitrary polygon-shaped computational cells i with volume V_i , boundary segment area A_s and perimeter index s (Fig. 4), as given by

$$\begin{aligned} \mathbf{V}_i^{n+1} &= \mathbf{V}_i^n - \frac{\Delta t}{V_i} \sum_s A_s (\hat{\mathbf{F}}_i)_s^{n+1} \\ &\quad - \frac{\Delta t}{V_i} \sum_s A_s (\mathbf{H}_i^{nc})_s^{n+1} (\hat{\mathbf{F}}_i)_s^{n+1} + \mathbf{D}_i^{n+1} \Delta t \end{aligned} \quad (29)$$

where the intercell fluxes $(\hat{\mathbf{F}}_i)_s$, the non-conservative part of the coefficient matrix $(\mathbf{H}_i^{nc})_s$ and source term vector \mathbf{D}_i are evaluated implicitly at the new time level.

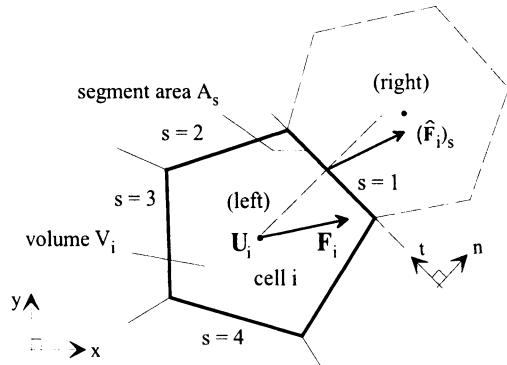


Figure 4: Finite volume discretization

The numerical fluxes at the interfaces between the computational cells,

$$\hat{\mathbf{F}}_s = \vec{F} \cdot \vec{n}_s \quad (30)$$

are calculated from a series of linearized, quasi one-dimensional Riemann problems normal to the specific surface areas of the computational cell boundary segments, as shown schematically in Fig. 5.

This yields, for the "Godunov" fluxes at the cell interfaces, $\hat{\mathbf{F}}_s$, the following:

$$\hat{\mathbf{F}}_s = \sum_{k, \lambda_k \geq 0} [\tilde{\mathbf{R}}_s] \mathbf{F}_l + \sum_{k, \lambda_k \leq 0} [\tilde{\mathbf{R}}_s] \mathbf{F}_r. \quad (31)$$

The "weighting" factors for the left and right fluxes in equation (31) are the sums of the split coefficient matrices for fluxes ordered with respect to the signs of the corresponding eigenvalues λ_k . These matrices can be obtained from the basic system of equations (19) by a parameter transformation

$$\frac{\partial \mathbf{F}}{\partial t} + \mathbf{R} \cdot (\nabla \mathbf{F}) = \mathbf{K} \mathbf{C} \quad (32)$$

with the new coefficient matrix

$$\mathbf{R} = \sum_k \mathbf{R}_k, \quad (33)$$

and

$$\mathbf{R}_k = \mathbf{K} \mathbf{G}_k \mathbf{K}^{-1} \quad (34)$$

where the Jacobian matrix is defined as

$$\mathbf{K} = \frac{\partial \mathbf{F}}{\partial \mathbf{U}}. \quad (35)$$

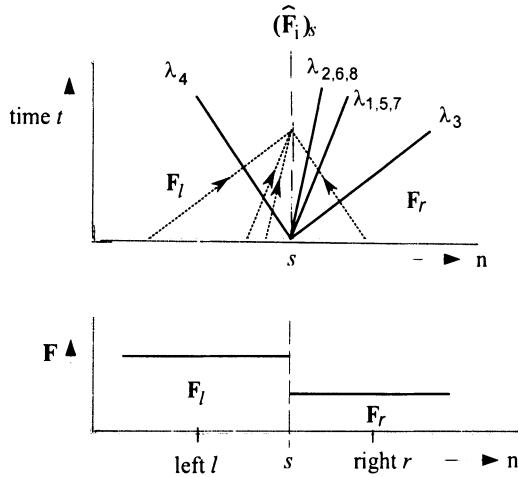


Figure 5: Linearized Riemann problem for two-phase flow

One might note that these transformations do not change the eigenvalues given by equation (21).

If the split matrices defined by equation (33) are divided by the corresponding eigenvalue, one obtains

$$\tilde{\mathbf{R}}_k = \frac{1}{\lambda_k} \mathbf{R}_k, \quad (36)$$

as used in equation (31) with the condition

$$\sum_k \tilde{\mathbf{R}}_k = \mathbf{I}. \quad (37)$$

More details with respect to the numerical method applied can be found in (Stadtke, Franchello & Worth, 1998).

3.3 SECOND-ORDER ACCURACY

A near second-order accuracy is obtained by a linear reconstruction of the solution in all computational cells following the Monotonic Upstream Scheme for Conservation Laws (MUSCL) approach (Van Leer, 1979) as indicated schematically in Fig. 6. New values for the 'primitive' parameters at the left and right side of the cell interfaces are calculated by a linear extrapolation from the adjacent cells as

$$\left. \begin{aligned} \mathbf{U}_{i+1/2}^l &= \mathbf{U}_i + \sigma_i \frac{(\Delta x)_i}{2} \\ \mathbf{U}_{i+1/2}^r &= \mathbf{U}_{i+1} - \sigma_{i+1} \frac{(\Delta x)_{i+1}}{2} \end{aligned} \right\}. \quad (38)$$

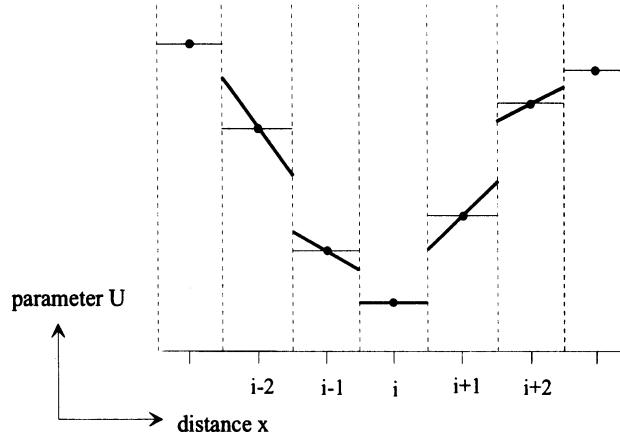


Figure 6: Linear reconstruction of solution

Slope limiter functions are then applied in order to maintain a monotonic behaviour of the solution

$$\sigma_i = f_{\text{lim}} \left[\left(\frac{\partial \mathbf{U}}{\partial x} \right)_{i-1/2}, \left(\frac{\partial \mathbf{U}}{\partial x} \right)_{i+1/2} \right]. \quad (39)$$

The limiter function used in the following numerical examples combine some properties of both 'minmod' and 'superbee' limiters.

4. Numerical Examples

In the following, only a few examples are shown to demonstrate in particular the ability of the present approach for the numerical simulation of strong parameter gradients at low and high Mach-numbers.

4.1 SINGLE PHASE GAS FLOW

Since the numerical technique described above differs from standard Flux Vector Splitting techniques or other Approximate Riemann solvers, the method has been tested first against gas dynamics benchmark test cases.

Fig. (7) compares first- and second-order calculations with the exact solution for a shock-tube problem. The achieved accuracy is close to what has been calculated with a first- and second-order Approximate Riemann Solver based on a Roe average matrix.

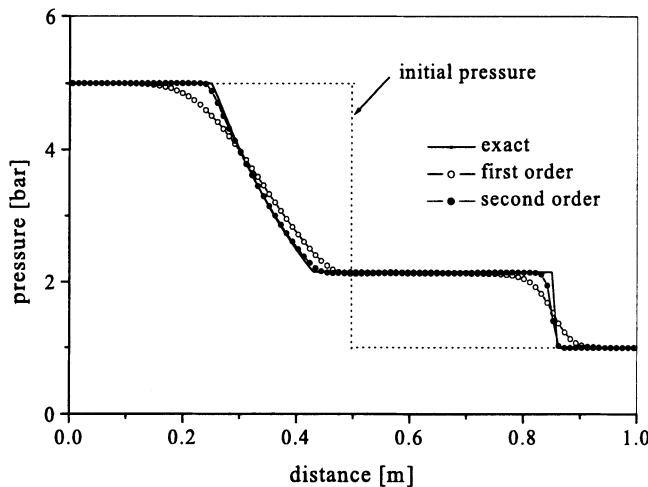


Figure 7: Shock-tube problem: pressure distribution at 6 ms

4.2 OSCILLATING WATER COLUMN IN A U-TUBE MANOMETER

This 1-dimensional benchmark problem was originally proposed by Ransom (DOE/EPRI Workshop, 1987) in order to check whether a particular model can simulate correctly a moving liquid level and to assess the magnitude of any artificial (numerical) viscosity present in the numerical method used. The general problem is as shown in Fig. 8.

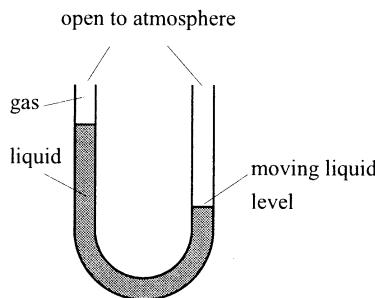


Figure 8: Oscillating water column in a U-tube manometer

The beauty of this test case is that, in the absence of wall friction and fluid viscosity, the analytical solution is known, assuming that the liquid column oscillates as a rigid body. With the present model, the tracking of the liquid level at the liquid/vapour interface (where the void fraction

changes from 0.0 to 1.0) is achieved with a resolution smeared over just two computational cells. As shown in Fig. 9, the predicted frequency is in

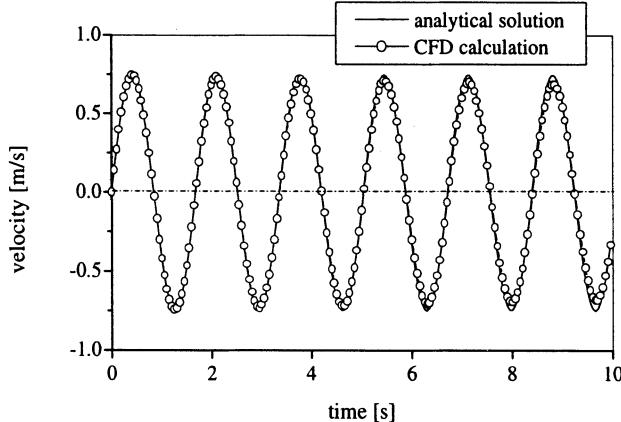


Figure 9: Liquid velocity at bottom of U-tube

excellent agreement with the undamped analytical solution. Any inherent damping is extremely low and may be attributable also to the slight energy dissipation caused by small slip velocities between the phases in the vicinity of the exposed liquid surface.

4.3 EDWARDS' PIPE BLOWDOWN

A classical test case for transient two-phase flow codes has been the prediction of the blowdown of initially subcooled liquid from a pipe of approximately 4 m length, with the initial conditions as given in Fig. 10.

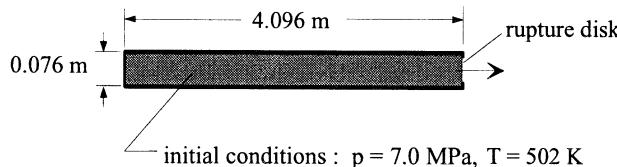


Figure 10: Edwards' pipe blowdown

The transient is initiated by rupturing a bursting disk at time zero. The pressure of the environment is atmospheric pressure. The first 10 ms of the transient is characterized by the propagation of a rarefaction wave from the opening into the pipe and the reflection of the wave at the closed end of the pipe where a distinct undershoot of the pressure occurs.

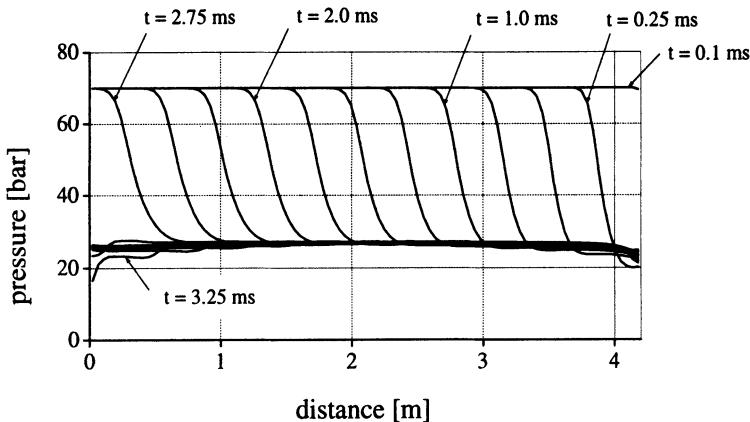


Figure 11: Pressure distribution during first blowdown period

This is clearly indicated in Fig. 11, where predicted pressure distributions along the pipe are shown for different time values. Note that the wave profile does not change before its refection at the closed end. Due to the onset of flashing (fast evaporation) the pressure is maintained during the first 5 ms close to the saturation pressure of the initial temperature of the liquid (e.g. 28.0 bar). This first pressure wave propagation period is followed by a more moderate depressurization and a continuous evaporation process. This is shown in Fig. 12 where measured and predicted values are shown for the pressure at the closed end of the pipe and for the volumetric vapour fraction at the middle of the pipe.

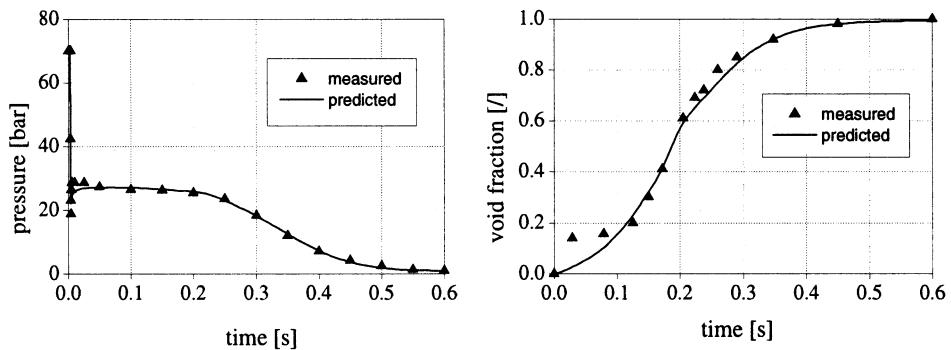


Figure 12: Measured and predicted pressure and void fraction

4.4 SLOSHING IN A CLOSED TANK

This might be seen as a 2-dimensional equivalent to the oscillating water column in a U-tube manometer. It is assumed that there is no phase change (evaporation or condensation effects). The initial conditions used are shown in Fig. 13. For the subsequent movement of the interface between the

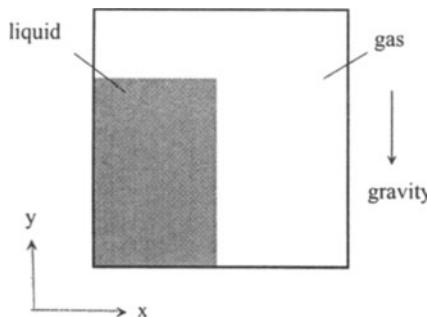


Figure 13: Initial condition for sloshing test case

phases, viscosity effects are taken into account (contrary to the oscillating water column in the preceding U-tube manometer case).

Various time frames with overlays of predicted volumetric concentration and related iso-contour lines are given in Fig. 13. The results indicate the formation of a strong water "wave" which surges to the opposite side of the tank where the wave crest "breaks" on reflection from the right side and top walls of the tank before falling back to the lower part of the vessel.

5. Summary

Based on an extended 'two-fluid' approach, a new model for multi-dimensional inhomogeneous non-equilibrium two-phase flow has been developed using separate balance equations for mass, momentum and energy (entropy) for the two (coupled) phases. By introducing appropriate formulations for the non-viscous interfacial forces, a hyperbolic system of equations has been derived, characterized by the existence of only real eigenvalues and a complete complementary set of independent eigenvectors.

The new model allows the application of Flux Vector Splitting techniques for the numerical integration of the flow equations. This technique combines preservation of the signal propagation along the characteristic wave directions with the property of accurate conservation of mass, momentum and energy.

Results of several test cases are shown to demonstrate the capability of the new models and numerical strategies for the prediction of 1-D and 2-D inhomogeneous, non-equilibrium two-phase flow processes

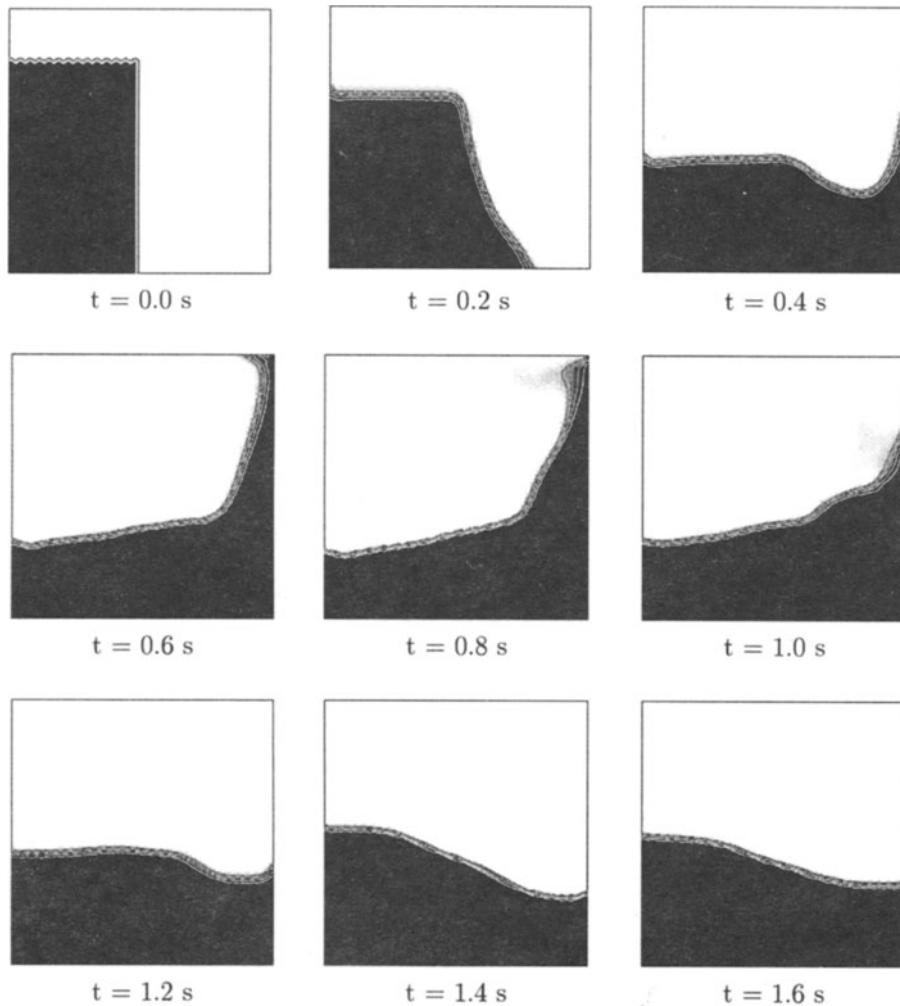


Figure 14: Sloshing of a water column in a closed tank, volumetric concentration for water (dark) and air (light) at different times

Nomenclature

a	sound velocity
A^{int}	interfacial area
a^{int}	interfacial area concentration per unit volume
C_w	interfacial resistance coefficient
F	force per unit volume
k	virtual mass coefficient
p	pressure
Re	Reynolds number
s	entropy
t	time
T	temperature
u	internal energy
v	flow velocity
\mathbf{C}, \mathbf{D}	source term vectors
$\mathbf{F}, \hat{\mathbf{F}}$	flux vector, numerical flux vector
\mathbf{U}, \mathbf{V}	state vectors
$\mathbf{G}, \mathbf{H}, \mathbf{R}$	coefficient matrices
\mathbf{I}	identity matrix
\mathbf{J}, \mathbf{K}	Jacobian matrices
Λ	diagonal matrix of eigenvalues

Greek letters

α	volumetric concentration of phase i
γ_i	distribution function of phase i
λ_k	k^{th} eigenvalue
ρ	density
$\sigma^M, \sigma^Q, \sigma^S$	volumetric source for mass, heat, entropy

Subscripts

g, l	gas, liquid phase
i	phase i , or computational cell i
s	cell boundary segment index

Superscript

<i>ext</i>	external
<i>exc</i>	quantity exchanged between phases
<i>fr</i>	friction
<i>int</i>	interfacial parameter
<i>nc</i>	non-conservative part
<i>nv</i>	non-viscous terms
<i>T</i>	transpose of matrix

References

- Boure, A., On a Unified Presentation of the Non-Equilibrium Two-Phase Flow Models, Proc. of ASME Symp., New York, (1975)
- Delhaye, J.M., Achard, J.L. On the Averaging Operators Introduced in Two-Phase Flow Modelling, Proc. CSNI Specialist Mtg. on Transient Two-Phase Flow, Toronto, (1976)
- Drew, D., Cheng, I., Lahey, R.T., The Analysis of Virtual Mass Effects in Two-Phase Flow, *Int. J. Multiphase Flow*, **5**, 233-242, (1979)
- Drew, D., Lahey, R.T., Application of General Constitutive Principles to the Derivation of the Multidimensional Two-phase Flow Equations, *Int. J. Multiphase Flow*, **5**, 243-264, (1979)
- Gray, W.G., Lee, P.C., On the Theorems for Volume Averaging of Multiphase Flow Systems, *Int. J. Multiphase Flow*, **3**, 333-340, (1977)
- Ishii, M., Thermodynamics of Two-Phase Flow, Eyrolles, Paris, (1975)
- Stadtke, H., Holtbecker, Hyperbolic Model for Inhomogeneous Two-phase Flows, Int. Conf. on Multiphase Flows, '91 TSUKUBA, Univ. of Tsukuba, Japan, (1991)
- Stadtke, H., Franchello, G., Worth, B., Towards a High Resolution Numerical Simulation of Transient Two-phase Flow, Third Int. Conf. on Multiphase Flow, ICMF '98, Lyon, France, (1998)
- Stadtke, H., Blahak, A., Worth, B., Modelling of Interfacial Area Concentration in Two-Phase Flow Systems, 8th Int. Meeting on Nuclear Thermal Hydraulics, Kyoto, Japan, (1997)
- Van Leer, B., Towards the Ultimate Conservative Difference Scheme, Second Order Sequel to Godunov's Method, *J. Computational Phys.* **32**, (1979)
- DOE/EPRI Workshop on Two-phase Flow Fundamentals, Rensselaer Polytechnic Institute, Troy, NY, USA, (1987)

A METHOD OF LINES FLUX-DIFFERENCE SPLITTING FINITE VOLUME APPROACH FOR 1D AND 2D RIVER FLOW PROBLEMS

G. STEINEBACH

*Bundesanstalt für Gewässerkunde
Postfach 200253, 56002 Koblenz, Germany
Email: Steinebach@bafg.de*

AND

A.Q.T. NGO

*SAP AG (Basis 04)
Neurottstrasse 16, 69190 Walldorf, Germany
Email: adrian.ngo@sap.com*

Abstract

The reliable forecasting of water levels is a very important issue. The modelling approach for water level forecasting at the Middle and Lower River Rhine is based on hydrodynamic river flow models coupled with precipitation-runoff models. The hydrodynamic model is defined by a numerical solution of the one-dimensional (1d) shallow water equations. If flood plains or flood risk maps are important, a two-dimensional (2d) model is required.

In this paper the usage of the Alcrudo-Garcia-Navarro scheme (Alcrudo and Garcia-Navarro, 1993) for 1d and 2d problems within the method of lines framework is described. The scheme is slightly modified, in order to allow a more accurate solution of problems with strong variations in the bottom topography. The problem of drying and re-wetting of mesh-cells is not yet fully sufficiently solved. Numerical results for some test problems and an application to a natural river will be presented.

1. Introduction

With the connection of the Central and East European states, the waterways in Germany will increasingly be used for transit navigation. Therefore all possibilities of logistical concepts and information exchange must be used

in order to assist navigation. The reliable forecasting of water levels in free flowing rivers is one contribution to an intelligent waterway (Fröhlich et. al., 1998). The aim is to provide navigation already during loading with enough water level information, to enable the loading of the ship to a maximum level in case of medium or low flow conditions. On the other hand, during floods due warnings are required for a damage limitation for the river-side residents (Steinebach and Wilke, 2000). Therefore, water-level forecasts are daily performed at the German Federal Institute of Hydrology.

Additionally to these forecasts, flood maps are required for the classification of potential damage and flood risk assessment.

The accurate simulation of river flow is essential for water-level forecasting and the calculation of flood maps. Advanced computer models for river flow simulation are based on the hyperbolic shallow water equations. They will be discussed in section 2 with special emphasis to the influence of bottom friction and the treatment of natural river-bed geometries.

In section 3, the numerical solution approach is presented. Finally, in section 4 the practical usage of the method is demonstrated by its application to a natural river stretch.

2. 1d and 2d river flow modelling

River flow modelling in two space dimensions is usually based on the 2d shallow-water equations (2d SWEs) (Vreugdenhil, 1994):

$$\mathbf{q}_t + \mathbf{e}(\mathbf{q})_x + \mathbf{g}(\mathbf{q})_y = \mathbf{s}(\mathbf{q}). \quad (1)$$

$\mathbf{q} = (\mathbf{h}, \mathbf{uh}, \mathbf{vh})^t$, is the vector of states with water depth $h(t, x, y)$ and depth averaged velocities $u(t, x, y), v(t, x, y)$ in x resp. y direction. The flux vectors are given by $\mathbf{e}(\mathbf{q}) = (\mathbf{uh}, \mathbf{u}^2\mathbf{h} + \frac{1}{2}\mathbf{gh}^2, \mathbf{uvh})^t$ and $\mathbf{g}(\mathbf{q}) = (\mathbf{vh}, \mathbf{uvh}, \mathbf{v}^2\mathbf{h} + \frac{1}{2}\mathbf{gh}^2)^t$.

The source term $\mathbf{s}(\mathbf{q}) = (0, gh(S_{0x} - S_{fx}), gh(S_{0y} - S_{fy}))^t$ accounts for bottom friction and bottom slope. The expression S_{fx} resp. S_{fy} is called friction slope and is assumed to be proportional to $-\mathbf{u}|\mathbf{u}|$ (Newton's friction), where \mathbf{u} is the velocity. An empirical formula (by Manning-Strickler) reads

$$S_{fx} = \frac{1}{K_S^2 h^{\frac{4}{3}}} \cdot u \cdot \sqrt{u^2 + v^2}, \quad S_{fy} = \frac{1}{K_S^2 h^{\frac{4}{3}}} \cdot v \cdot \sqrt{u^2 + v^2}, \quad K_S \in \mathbb{R}^+.$$

The constant K_S is called roughness coefficient. The bottom slope is given by $S_{0i} = -\partial_i b$ ($i = x, y$) with the bottom elevation $b(x, y)$.

The Saint-Venant equations are the shallow water equations in one space dimension directed along the river course. The water surface elevation then is given by $z(t, x) = b(x) + h(t, x)$, where $b(x)$ is the bottom level, $S_0 :=$

$-b_x$ the bottom slope and $h(t, x)$ is the water depth. For rectangular cross sections $A(t, x) = B \cdot h(t, x)$ of constant width B the equations read (Stoker, 1957):

$$\partial_t \begin{bmatrix} h \\ uh \end{bmatrix} + \partial_x \begin{bmatrix} uh \\ u^2 h + \frac{1}{2} g h^2 \end{bmatrix} = \begin{bmatrix} 0 \\ gh(S_0 - S_f) \end{bmatrix} \quad (2)$$

The friction slope in 1d is given by $S_f = \frac{1}{K_s^2 R^{\frac{4}{3}}} \cdot |u| \cdot u$ where the hydraulic radius R can be well approximated by $R \approx h$ for wide river cross-sections.

Many numerical schemes have been developed for the solution of the 2d SWEs and 1d Saint-Venant equations. Usually, the considerations are restricted to the homogeneous parts of the equations. Therefore, we give a brief description of the influence of the bottom friction and bottom slope.

The damping influence of the frictional force $-k|u|u$ is illustrated by the Cauchy problem for the Burgers equation with source term:

$$u_t + uu_x = -ku|u| \quad \text{with } k \in \mathbb{R}^+$$

Although this is merely a scalar equation, it embodies elementary properties of the convective part of a nonlinear system of conservation laws.

Given an initial condition $u_0(x) = u(0, x) > 0$ that is differentiable and monotonically decreasing the solution would “break down” (formation of a shock) at time $t_B = -\frac{1}{\min_{x_0} u'_0(x_0)}$ for the inviscid Burgers equation (Whitham, 1974).

By the method of characteristics it can be shown that this shock formation is delayed by the presence of the source term depending on k :

$$t_B(k) = -\frac{1}{\min_{x_0} \{ku_0(x_0) + u'_0(x_0)\}} > t_B(0) .$$

It is obvious that $t_B(k) \rightarrow t_B$ as $k \rightarrow 0$ and $t_B(k) \rightarrow \infty$ as $k \rightarrow -\frac{u'_0(x_0)}{u_0(x_0)}$ from the left. In the latter case the solution smoothens so fast that no shock can occur at all. Therefore, friction adds some smoothness to the solution and no special numerical treatment seems necessary. It should be mentioned, that S_f must be limited in case of very small water depth or even dry river beds.

The influence of a bottom slope $S_0 \neq 0$ is more critical to the numerical behaviour of the discretization scheme. Any discretization that is not compatible with the discretization of the convective terms (flux-terms) might cause difficulties (Nujić, 1996). For further investigations, a frictionless rectangular channel is assumed for $(t, x) \in \mathbb{R}^+ \times [\alpha, \beta]$ with initial condition $h(0, x) = z_0 - b(x)$, $u(0, x) = 0$ and boundary condition $u(t, \alpha) = u(t, \beta) = 0$. If the arbitrary bed profile $b(x) \in C^1[\alpha, \beta]$, then

$h(t, x) = z_0 - b(x)$ and $u(t, x) = 0$ satisfy equations (2) with $S_f = 0$, i.e. the surface does not move at all. A non-smooth bottom elevation $b(x)$ will generate a non-smooth solution component $h(t, x)$. Appropriate numerical schemes must properly handle this difficulty.

3. Numerical solution approach

Forecast systems for large rivers are usually built up by the coupling of several models for the main river and its tributaries (Steinebach, 1998). Each river-reach may be modelled by the 1d Saint-Venant equations or even by a 2d approach. Other network elements are weirs or further river structures (Rentrop et. al., 1999). The coupling is performed through the boundary conditions. A common approach for the numerical solution of the whole network equations is the method of lines (MOL) (Rentrop and Steinebach, 1997). The semi-discretization in space leads to a large system of differential algebraic equations (DAEs). The algebraic equations are defined through the boundary and coupling conditions.

For the solution of the DAEs, a special variant (RODASP) of the well-known fourth-order Rosenbrock-Wanner (ROW) method RODAS is used (Hairer and Wanner, 1991; Steinebach, 1995). The coefficients of RODASP are adapted to avoid order-reduction effects in the MOL framework and the linear algebra routines allow the treatment of sparse matrices. ROW-methods are known to be efficient for moderate accuracy requirements (Verwer et. al., 1998). Due to their semi-implicit structure they allow large time-steps, which are appropriate for the simulation of slowly varying flow problems.

The choice of the space discretization scheme is problem dependent. For many subcritical 1d flow problems standard finite differences are sufficient. In case of supercritical flow with possibly discontinuous solution components, more sophisticated schemes must be applied. Finite difference ENO-schemes were proposed in (Hilden and Steinebach, 1998) for 1d flow and transport problems. The 2d modelling of natural rivers requires the adaptation of the computational mesh to complex geometries (Sleigh et. al., 1996; Sleigh et. al., 1998). Conservative finite volume (FV) schemes are preferable in this case. The further investigations are based on the flux-difference splitting FV scheme described in (Alcrudo et. al., 1992; Alcrudo and Garcia-Navarro, 1993). This scheme was derived for the homogeneous flow equations in one and two space dimensions. A brief review is given below.

We start with the IBVP

$$q_t + f(q)_x = s(q), \quad (3)$$

with $\mathbf{q}(\mathbf{0}, \mathbf{x}) = \mathbf{q}_0(\mathbf{x})$ and $\mathbf{q}(t, \alpha) = \mathbf{q}_\alpha(t)$, $\mathbf{q}(t, \beta) = \mathbf{q}_\beta(t)$. The space-interval $[\alpha, \beta]$ is partitioned in N cells I_1, \dots, I_N through a given set of $N+1$ mesh-points by $\alpha = x_{\frac{1}{2}} < \dots < x_{N+\frac{1}{2}} = \beta$.

Integration of (3) over the control volume $I_j = [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$ with length $\Delta x_j = x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}$ yields

$$\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \partial_t \mathbf{q}(t, \mathbf{x}) d\mathbf{x} = - \left[\mathbf{f}(\mathbf{q}(t, x_{j+\frac{1}{2}})) - \mathbf{f}(\mathbf{q}(t, x_{j-\frac{1}{2}})) \right] + \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathbf{s}(\mathbf{q}(t, \mathbf{x})) d\mathbf{x} \quad (4)$$

Defining the cell-average

$$\mathbf{Q}_j(t) = \frac{1}{\Delta x_j} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathbf{q}(t, \mathbf{x}) d\mathbf{x} \quad (5)$$

on the cell centers $x_j = \frac{1}{2}(x_{j-\frac{1}{2}} + x_{j+\frac{1}{2}})$, $j = 1, \dots, N$, equation (4) can now be written as

$$\mathbf{Q}'_j(t) = - \frac{1}{\Delta x_j} \left[\mathbf{f}(\mathbf{q}(t, x_{j+\frac{1}{2}})) - \mathbf{f}(\mathbf{q}(t, x_{j-\frac{1}{2}})) \right] + \frac{1}{\Delta x_j} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathbf{s}(\mathbf{q}(t, \mathbf{x})) d\mathbf{x}.$$

We now put emphasis on the approximation of the right hand side. We will actually solve a system of ordinary differential equations

$$\mathbf{Q}'_j(t) = \left(- \frac{1}{\Delta x_j} \left[\mathbf{f}_{j+\frac{1}{2}}^* - \mathbf{f}_{j-\frac{1}{2}}^* \right] + \mathbf{s}_j \right) (\mathbf{Q}(j; t)) \quad (6)$$

The computation of $\mathbf{f}_{j+\frac{1}{2}}^*$, $\mathbf{f}_{j-\frac{1}{2}}^*$ requires the solution of local Riemann problems. In practice this is carried out by an approximate Riemann solver. In this approach the well known flux-difference splitting Roe-scheme is applied:

$$\mathbf{f}_{j+\frac{1}{2}}^* = \frac{1}{2} \left(\mathbf{f} \left(\mathbf{w}_{j+\frac{1}{2}}^{\mathbf{L}} \right) + \mathbf{f} \left(\mathbf{w}_{j+\frac{1}{2}}^{\mathbf{R}} \right) - \sum_{p=1}^m |\hat{\lambda}_p| \alpha_p \hat{\mathbf{r}}_p \right) \quad (7)$$

with $\mathbf{w}_{j+\frac{1}{2}}^{\mathbf{L}} = \mathbf{Q}_j(t)$ and $\mathbf{w}_{j+\frac{1}{2}}^{\mathbf{R}} = \mathbf{Q}_{j+1}(t)$ and where $\hat{\lambda}_P$, $\hat{\mathbf{r}}_p$ are the eigenvalues resp. eigenvectors of the Roe-matrix and α_p is given through

$$\mathbf{w}_{j+\frac{1}{2}}^{\mathbf{R}} - \mathbf{w}_{j+\frac{1}{2}}^{\mathbf{L}} = \sum_{p=1}^m \alpha_p \hat{\mathbf{r}}_p \quad (8)$$

The Roe-matrix for problem (2) is given in (Alcrudo et. al., 1992). A second order extension of the scheme is obtained by standard MUSCL extrapolation.

The momentum equation of (2) contains two kinds of source terms. The bottom friction can simply be discretized pointwise. I.e., for (6) we set $s_j(t) \approx (0, gh_j(t) S_f(Q_j(t))^t)$.

The bottom slope (ghb_x) contains a spatial derivative for which a discrete representation has to be established. In order to be compatible with the discretization of the convective terms, the Roe-scheme has to be modified.

First of all, it can be observed that the system (2) with $S_f = 0$ is equivalent to

$$\begin{aligned} z_t &= -(hu)_x \\ (uh)_t &= -(u^2 h)_x - ghz_x \end{aligned} \quad (9)$$

Now, $\mathbf{q} = (\mathbf{z}, \mathbf{uh})^t$ is chosen as the new state vector. Unfortunately, the right-hand side of this system is not in a flux-conservative representation (3). Nevertheless, it can be written in quasi-linear form with $c = \sqrt{gh}$:

$$\begin{pmatrix} z \\ uh \end{pmatrix}_t + \begin{pmatrix} 0 & 1 \\ c^2 - u^2 & 2u \end{pmatrix} \begin{pmatrix} z \\ uh \end{pmatrix}_x = \begin{pmatrix} 0 \\ u^2 b_x \end{pmatrix} \quad (10)$$

Thus, the homogeneous part is equivalent to the quasi-linear representation of (2) and the Roe-matrix is known. Therefore, the flux-difference term in (7) is fully determined by using the new state vector.

For the evaluation of the fluxes $\mathbf{f}(\mathbf{w}_{j+\frac{1}{2}}^L)$, $\mathbf{f}(\mathbf{w}_{j+\frac{1}{2}}^R)$, representation (9) is used. Assuming a piecewise constant function $h(t, x) = h_j(t)$, for $x \in I_j$, (9) is formally in flux-conservative form on each mesh cell and the evaluation of the fluxes in (7) is straightforward.

The extension of this 1d Roe-scheme to the 2d problem (1) is possible due to the rotational invariance of the 2d SWEs (Alcrudo and Garcia-Navarro, 1993; Ngo, 1999).

4. Applications

In a first test case, the modified scheme is compared to the original one for a steady frictionless flow q_0 through a rectangular channel of 1m width and 1000m lenght. From (2) we get: $u(x) = \frac{q_0}{h(x)}$ and $b_x = \frac{q_0^2}{gh^3} h_x - h_x$. An analytical solution is constructed by defining $h(x) := 4 + \frac{1}{2} \sin(\frac{\pi x}{50})$ and $q_0 = 10$. Integrating b_x from $x = 0$ up to $x = 1000$ yields

$$b(x) \approx 1.31855$$

$$-\frac{20.387 + 32 \sin(0.063 x) + 8 \sin(0.063 x)^2 + \frac{1}{2} \sin(0.063 x)^3}{(8 + \sin(0.063 x))^2}$$

In Figure 1 the numerical results of the original first order scheme (Roe(1)) and of the modified scheme are compared to the analytical solution for a mesh-size of $\Delta x = 25$. It can be concluded that the modified scheme gives much better results for strong variations in the bottom elevation.

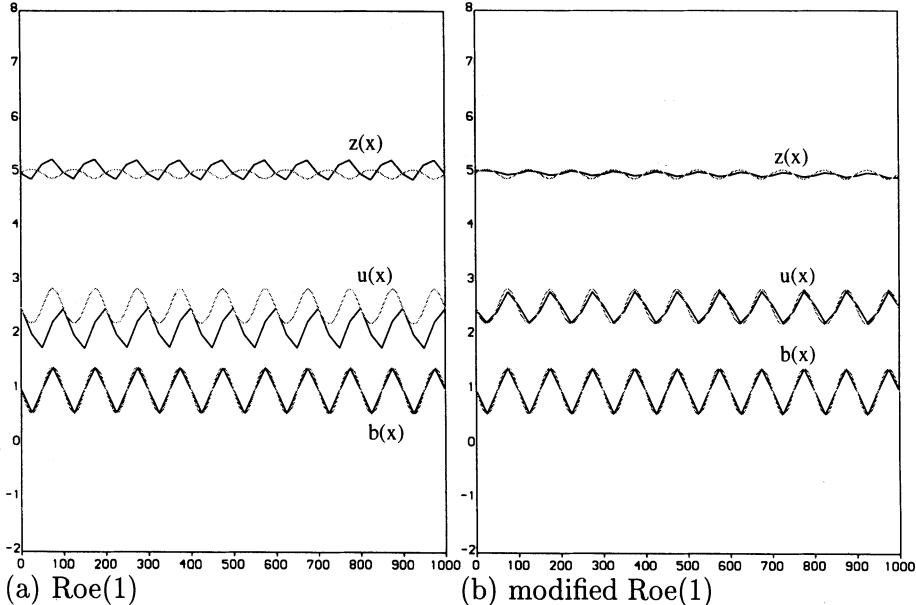


Figure 1: Numerical versus analytical solution for strong variations in the bottom elevation. The dashed lines represent the benchmark solution, the solid lines show the corresponding numerical approximations.

The 1d idealized dam-break problem served as another benchmark. The performance and efficiency of the modified method is comparable with Roe's original scheme (Ngo, 1999). It has to be mentioned that explicit schemes are preferable for those highly dynamical simulations with a strongly varying solution.

The final application (Figure 2) shows a flood map at River Saar during the flood event in December 1993. This map was computed with the 2d variant of the proposed solution approach. The river section including the flood plain was discretized by a 300×75 mesh of quadrilaterals.

The overall performance of the algorithm is severely influenced by the process of drying and re-wetting of some mesh-cells during simulation. If the water-depth in a mesh-cell j is below a certain limit, the equations (6) for this cell are replaced by $\mathbf{Q}_j'(t) = \mathbf{0}$ and new boundary conditions are defined between wet and dry cells. If the water elevation z of a neighbouring cell exceeds $z_j + \epsilon$, the original equations are solved again. This re-wetting process results in severe step-size reductions of the time integration algorithm.



Figure 2: Flood map of the River Saar stretch Wiltinger Bogen

5. Conclusion

The proposed Roe-scheme of Alcrudo et al. is a very attractive candidate for the solution of the 1d and 2d SWEs. A slight modification allows the proper treatment of flow problems with uneven bottom elevations. The consideration of bottom friction should not cause problems since it adds some smoothness to the solution. The method can be used within explicit and implicit time integration schemes. For the modelling of whole river networks, the semi-implicit ROW-scheme RODASP is a robust working horse in the method of lines framework.

The efficient solution of the drying and re-wetting problem remains a great challenge (Molinaro, 1992), especially in the context of implicit time integration for slowly varying flows in natural rivers.

Acknowledgement

The authors like to thank Joachim Strauch for implementing an early version of the 2d code and Karl-Heinz Daamen and Ingo Schnatz for arranging the maps. For many helpful discussions the first author is indebted to Michael Hilden. The work was partially funded by the EU Commission under the NOAH-project.

References

- Alcrudo F, Garcia-Navarro P, Sapiro J M (1992). Flux-difference splitting for 1d open channel flow equations. *Int. J. f. Num. Meth. Fluids*, **14**, pp 1009-1018.
- Alcrudo F, Garcia-Navarro P (1993). A high-resolution Godunov-type scheme in finite volumes for the 2d shallow-water equations. *Int. J. f. Num. Meth. Fluids*, **16**, pp 489-505.
- Fröhlich W, Heinz M, Steinebach G, Wilke K (1998). Water level forecasts for navigation on the river Elbe and river Rhine - A contribution to an intelligent waterway. *Proc. 29th PIANC International Navigation Congress The Hague 1998*, Section I, subject 3, pp 55-60, International Navigation Association.
- Hairer E, Wanner G (1991). Solving ordinary differential equations II, *Springer Series in Computational Mathematics*, Berlin, Heidelberg.
- Hilden M, Steinebach G (1998). ENO-discretizations in MOL-applications: some examples in river hydraulics. *Applied Numerical Mathematics* **28**, pp 293-308.
- Molinaro P (1992). A review of 2d mathematical models for the simulation of flood propagation on dry bed. in *Blain, W.R., Cabrera, E. (eds.): Fluid flow modelling, Compuataional mechanics pub.*, pp 431-443, Elsevier.
- Ngo Q T (1999). Numerical simulation of river flow problems based on a finite volume model. *Diploma thesis*, Univ. Kaiserslautern.
- Discussion by Nujić M (1996). *Journal of Hydraulic Eng.*, January 1996, pp 51-52.
- Rentrop P, Hilden M, Steinebach G (1999). Wissenschaftliches Rechnen. *Der Ingenieur in der Wasser- und Schifffahrtsverwaltung* **4**, pp 19-23.
- Rentrop P, Steinebach G (1997). Model and numerical techniques for the alarm system of river Rhine. *Surveys Math. Industry* **6**, pp 245-265.
- Sleigh P A, Berzins M and Gaskell P H (1996). A Reliable and Accurate Technique for the Modelling of Complex Hydraulic Flows. *Proceedings of the First International Symposium on Finite Volumes for Complex Applications(held at Rouen)*, Hermes, Paris, 635-642.
- Sleigh P A, Berzins M, Gaskell P H and Wright N G (1998). An Unstructured Finite Volume Algorithm for Predicting Flow in Rivers and Estuaries. *Computers and Fluids* **27**,**4**, pp 479-508.
- Steinebach G (1995). Order-reduction of ROW-methods for DAEs and method of lines applications. *Preprint-Nr. 1741*, FB Math., TH Darmstadt.
- Steinebach G (1998). Using hydrodynamic models in forecast systems for large rivers. *Proc. Advances in Hydro-Science and -Engineering*, Vol.3 incl. CD-ROM, Holz, K.P., Bechteler, W., Wang, S.S.Y., Kawahara, M. (ed.), Cottbus.
- Steinebach G, Wilke K (2000). Flood forecasting and warning on the River Rhine. *Water and Environmental Management, J.CIWEM*, **14** No.1, pp 39-44.
- Stoker J J (1957). Water waves, the mathematical theory with applications. *Interscience Publishers Inc.*, New York.
- Verwer J G, Hundsdorfer W H, Blom J G (1998). Numerical time integration for air pollution models. *Modelling, Analysis and Simulation Report MAS-R9825*, 58 p., CWI Amsterdam.
- Vreugdenhil C B (1994). Numerical methods for shallow-water flow. *Kluwer Acad. Pub.*, Dordrecht.
- Whitham G B (1974). Linear and nonlinear waves, *John Wiley and Sons Inc.*

A SIMPLE SMOOTHING TVD SCHEME ON STRUCTURED AND UNSTRUCTURED GRIDS

M. SUN

*Shock Wave Research Center, Institute of Fluid Science,
Tohoku University, Katahira 2-1-1, Aoba, Sendai 980, Japan
e-mail: sun@ceres.ifs.tohoku.ac.jp*

AND

K. TAKAYAMA

*Shock Wave Research Center, Institute of Fluid Science,
Tohoku University, Katahira 2-1-1, Aoba, Sendai 980, Japan
e-mail: takayama@ifs.tohoku.ac.jp*

Abstract. A symmetric TVD Lax-Wendroff scheme is formulated via conservative smoothing or artificial viscosity. The approach is based on the Richtmyer scheme, a predictor-corrector Lax-Wendroff scheme. The predictor step is conducted at interface to determine the states there by advancing the Euler equations by modified half time step. The corrector step sums fluxes given by the Euler equations, and the smoothing flux given by artificial dissipation. The scheme is a simple central difference, and its artificial dissipation can be as less as the first-order upwind scheme around a discontinuity in solving the scalar equation. The scheme is extended to unstructured grid and applied to a variety of gas-dynamic problems by the finite volume method.

1. Introduction

Consider the convection equation

$$u_t + cu_x = 0, \quad (1)$$

where c is a wave speed. A conservative scheme is written as

$$u_i^{n+1} = u_i^n - \lambda(\hat{f}_{i+1/2} - \hat{f}_{i-1/2}), \quad (2)$$

where $\lambda = \Delta t / \Delta x$ is the ratio of time step Δt to grid size Δx . $\hat{f}_{i+1/2}$ is often referred to as *numerical flux*. We propose a central scheme whose numerical flux reads

$$u^* = u_{i+1/2} - \frac{\Delta t}{2}(1 - 2\mu)c\Delta_{i+1/2}u/\Delta x, \quad (3)$$

$$\hat{f}_{i+1/2} = cu^* - \mu\Delta_{i+1/2}u/\lambda, \quad (4)$$

where symbol μ is a coefficient of artificial viscosity. For $\mu = 0$, the scheme degenerates to a predictor-corrector Lax-Wendroff scheme, known as the Richtmyer scheme. A big difference between the present scheme and other schemes with artificial viscosity is that the artificial viscosity is also taken into account in the predictor step (3) as well as in the flux evaluation (4). The scheme has a close relationship to the smoothing method (Sun and Takayama, 1998), then it is referred to as a smoothing scheme. If μ is a constant, it is a linear smoothing scheme; if μ is not constant, a nonlinear smoothing scheme.

This paper summarizes main conclusions concerning the scheme without giving any proof in sections 2 and 3. The extension to system of equations and multi-dimensions is briefly discussed in sections 4 and 5. Detailed proof and discussion of the approach and more numerical applications may be found in thesis (Sun, 1998).

2. Linear smoothing scheme

THEOREM 1. The linear smoothing scheme is monotone under the CFL condition $|\sigma| \leq 1$ if $\frac{|\sigma|}{2(1+|\sigma|)} \leq \mu \leq 1/2$.

Remark. The truncation error of the scheme is

$$u_t + cu_x = \mu \frac{\Delta x^2}{\Delta t} (1 - \sigma^2) u_{xx} + \dots$$

Clearly, parameter μ is a sort of artificial viscosity coefficient. Any central and upwind schemes using a stencil with three nodes can be written in this form, for instance,

μ	schemes
0	the Lax-Wendroff scheme
$ \sigma /[2(1 + \sigma)]$	the First-order upwind scheme
1/2	the Lax-Friedrichs scheme

The Lax-Friedrichs scheme and the first-order upwind scheme give the upper limit and the lower limit respectively. Therefore, the Lax-Friedrichs

scheme is the most diffusive, while the upwind scheme is the least one. Even for $\mu = 1/4$ the scheme is half as diffusive as the Lax-Friedrichs scheme. For convenience, we hereafter use the notation,

$$\mu_{up} \equiv |\sigma|/[2(1 + |\sigma|)], \quad (5)$$

where the subscript “*up*” denotes the viscosity coefficient of the upwind scheme.

3. Second order smoothing scheme

Let's switch the linear smoothing scheme to the Lax-Wendroff scheme, in a nonlinear way, by setting

$$\mu = \begin{cases} 0, & \phi_{i+1/2} \leq \phi_0, \\ \mu_0, & \text{otherwise,} \end{cases} \quad (6)$$

where $\phi_{i+1/2} = \max(\phi_i, \phi_{i+1})$, and

$$\phi_0 \equiv \frac{(2 + 2/|\sigma|)(1 - 2\mu_0) - 1}{(2 + 2/|\sigma|)(1 - 2\mu_0) + 1}.$$

The smoothness indicator ϕ_i is defined as

$$\phi_i \equiv \frac{|u_{i+1} + u_{i-1} - 2u_i|}{\varepsilon_0 + |u_{i+1} - u_{i-1}|}, \quad (7)$$

where ε_0 is an infinitely small positive value used when $u_{i+1} - u_{i-1} = 0$.

THEOREM 2. The smoothing scheme with nonlinear coefficient (6) is monotonicity preserving under the CFL condition $|\sigma| \leq 1$ if $\mu_{up} \leq \mu_0 \leq 1/2$.

Consider a continuous nonlinear coefficient,

$$\mu = \min[\mu_0, \max(0, \mu_t)] \quad (8)$$

where μ_t denotes a continuous function of smoothness indicator in the transition regions from smooth to shock regions.

THEOREM 3. The smoothing scheme with nonlinear coefficient 8 is monotonicity preserving under the CFL condition $|\sigma| \leq 1$ if

$$\mu_0 - (1 - 2\mu_0)r_{i+1/2} \leq \mu_t \leq \mu_0$$

with $\mu_{up} \leq \mu_0 \leq 1/2$ and $r_{i+1/2} = (1 - \phi_{i+1/2})/(1 + \phi_{i+1/2})$.

By choosing the lower limit for μ_t , one obtains a nonlinear coefficient

$$\mu = \min[\mu_0, \max(0, \mu_0 - (1 - 2\mu_0)r_{i+1/2})], \quad (9)$$

with $\mu_{up} \leq \mu_0 \leq 1/2$. Coefficient μ_0 is the largest coefficient used around shock regions.

4. System of equations

We are interested in the solution to the initial value problem of one-dimensional hyperbolic conservation laws, $\mathbf{U}_t + \mathbf{F}_x = 0$, where \mathbf{U} and \mathbf{F} are vectors with equal length. The smoothing scheme is written as

$$\begin{aligned}\mathbf{U}^* &= \mathbf{U} - (1/2 - \mu_{i+1/2})\Delta t \mathbf{F}_x, \\ \hat{\mathbf{F}}_{i+1/2} &= \mathbf{F}(\mathbf{U}^*) - \mu_{i+1/2}(\Delta x)^2 / \Delta t \mathbf{U}_x, \\ \mathbf{U}_i^{n+1} &= \mathbf{U}_i - \lambda(\hat{\mathbf{F}}_{i+1/2} - \hat{\mathbf{F}}_{i-1/2}),\end{aligned}\quad (10)$$

where subscript $i+1/2$ of some variables is removed for clearance. The first two equations in (10) which evaluate the flux through an interface $i+1/2$ completely based on *local* values and gradients, and it is similar in multi-dimensions so that the scheme can be easily extended to any unstructured grids after gradient reconstruction.

We propose a scalar smoothness indicator for the system of equations:

$$\phi_i = \frac{\sum_k \left\{ |U_{i+1}^k + U_{i-1}^k - 2U_i^k| / \bar{U}_i^k \right\}}{\varepsilon_0 + \sum_k \left\{ |U_{i+1}^k - U_{i-1}^k| / \bar{U}_i^k \right\}}. \quad (11)$$

All differences between variables U^k in (11) are normalized by some representative quantities \bar{U}_i^k . We choose \bar{U}_i^k to be the local pressure, density, and sound speed for U^k pressure, density, and velocity respectively. It is easily seen that in the case of single scalar equation (11) becomes (7). The scalar coefficient μ follows (9).

5. Multi-dimensions

We extend the smoothing scheme to unstructured and structured grids by the finite volume method. The method solves the conservation laws by directly applying them to every non-overlapping discrete volume the summation of which covers the whole computational domain. The conservation laws written for a discrete control volume are,

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i - \frac{\Delta t}{\Delta \Omega_i} \sum_{k=1}^4 \hat{\mathbf{F}}_k^{n+1/2}, \quad (12)$$

where $\hat{\mathbf{F}}_k^{n+1/2}$ are fluxes obtained by a predictor step, and locate at the center of the interface. Since the variables are unknown at the center, they

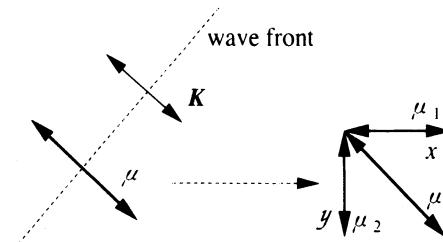


Figure 1. Decomposition of the artificial viscosity coefficient according to the wave direction.

are interpolated from the centroid (cell-centered data are considered here). The predictor step gives the state of variables at modified half time step, $\frac{\Delta t}{2}(1 - 2\mu)$, by locally solving the two-dimensional Lax-Friedrichs scheme at the interface. Both the interpolation and the predictor step necessitate estimation of the gradients at every cell which are given by the least-square method. In solving the compressible Navier-Stokes equations, the fluxes through cell faces, which consist of inviscid fluxes, viscous fluxes and smoothing fluxes, are

$$\hat{\mathbf{F}} = \hat{\mathbf{F}}_{\text{inviscid}} + \hat{\mathbf{F}}_{\text{viscous}} + \hat{\mathbf{F}}_{\text{smoothing}}. \quad (13)$$

The inviscid fluxes are convection terms and pressure surface terms, or those in the Euler equations. The viscous fluxes are used here to describe all terms that are specific to the Navier-Stokes, i.e. viscous dissipation and thermal diffusion terms. The smoothing fluxes added to suppress spurious oscillations, read

$$\hat{\mathbf{F}}_{\text{smoothing}} = \begin{pmatrix} -\rho_x \mu_1 - \rho_y \mu_2 \\ -(\rho u)_x \mu_1 - (\rho u)_y \mu_2 \\ -(\rho v)_x \mu_1 - (\rho v)_y \mu_2 \\ -(\rho E)_x \mu_1 - (\rho E)_y \mu_2 \end{pmatrix}, \quad (14)$$

where u_n is the normal velocity through the interface, μ_1 and μ_2 are two nonlinear coefficients of artificial viscosity. In multi-dimensional flow, the artificial viscosity coefficient is actually a tensor. Since it should be as less diffusive as possible, it is set to be zero in directions other than the normal direction of the wave front. It is then rotated to the Cartesian coordinates, and become

$$\begin{vmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{vmatrix}.$$

This decomposition is shown in Fig. 1. Having obtained the artificial viscosity coefficient in x and y directions, one may easily design a smoothing

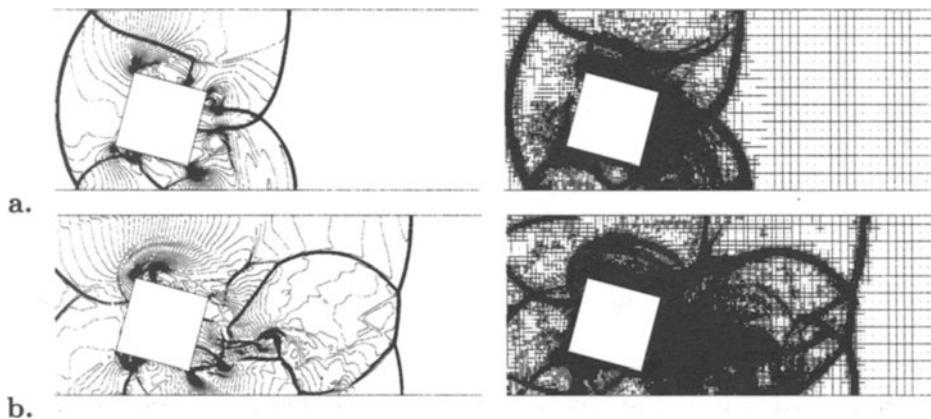


Figure 2. Shock / square-cylinder interaction, isopycnics and adaptive meshes. Incident shock Mach number $M_s = 2$, 3-level refinement, CFL = 0.7. **a.** $t = 1.0$, 22981 cells, CPU time = 29.6s; **b.** $t = 2.0$, 38383 cells, CPU time = 82.0s

flux as that in (14). The direction of the wave front is estimated by velocity difference between two cells.

6. Numerical examples

The present approach has been successfully applied to the solution of a variety of gas-dynamic problems, such as nozzle starting process, interfacial instability, airfoil flow. In this paper only a truly unsteady phenomenon, shock wave interaction with an oblique square, is demonstrated. The distance from the square center to its tip is taken as 1, and the square is oblique to the horizontal line by 30° . The distances from the square center to the upper and lower surface are 2 and 1.3 respectively. The results for incident shock Mach number $M_s = 2$ are shown in Fig. 2. The approach resolves very well all dominant flow features, which actually include most basic phenomena in compressible flows, such as shock diffraction, reflection, vorticity generation, and interactions of shock/expansion waves, shock/vortex.

References

- Davis SF (1987) A simplified TVD finite difference scheme via artificial viscosity, SIAM J. Sci. Stat. Comput. 8 : 1-18.
- Sun M (1998) Numerical and experimental studies of shock wave interaction with bodies, Ph.D. Thesis, Tohoku University, Japan.
<http://ceres.ifs.tohoku.ac.jp/~sun/thesis.html>
- Sun M, Takayama K (1999) Conservative smoothing on an adaptive quadrilateral grid, *J. Comput. Phys.*, **150** : 143-180.
- Yee HC, (1987) Construction of explicit and implicit symmetric TVD schemes and their applications, *J. of Comput. Phys.* **68** : 151-179.

GODUNOV METHODS

PETER K. SWEBY

*Department of Mathematics,
The University of Reading
Whiteknights P.O. Box 220
Reading RG6 6AX
England
Email: P.K.Sweby@rdg.ac.uk*

Abstract. This paper reviews the class of numerical schemes, known as Godunov Methods, used for the solution of hyperbolic conservation laws. Such numerical schemes can be characterised by the solution (exact or approximate) of a Riemann Problem (classical or generalised) within computational cells in order to obtain the numerical fluxes.

Since the original first order scheme, proposed by Godunov in 1959, there has been much development of the idea; for example, the MUSCL scheme of van Leer in 1979, the PPM scheme of Woodward and Colella in 1984 and the Higher Order Godunov schemes of Bell, Colella and Trangenstein (1989).

As well as considering the original scheme and its later variants, we place these developments in historical context, making links with other work in the area.

1. Introduction

The title of this paper begs the question: What is a Godunov method? In 1983, Harten, Lax & van Leer gave a technical definition of Godunov-type schemes in (Harten, Lax and Van Leer, 1983). However, more recently, van Leer (Van Leer, 1997; Van Leer, 1999) has given the succinct definition:

... we define Godunov-type methods as non-oscillatory finite-volume schemes that incorporate the solution (exact or approximate) to Riemann's initial-value problem, or a generalization of it,.....

It is this definition that we will explore.

In 1959 Godunov published his inspirational paper (Godunov, 1959), based on the work of his Ph.D., in which he used the solution of the Riemann problem as a building block for a finite-volume scheme for compressible flow. His scheme, as originally presented, involved a Lagrangian step followed by an Eulerian remapping. However, it may be recast (see for example (Van Leer, 1984)) into a conservative Eulerian framework. Whilst Godunov used the exact solution of the Riemann problem in his work, others (Roe (Roe, 1981), Osher & Solomon (Osher and Solomon, 1982), Harten, Lax & van Leer (Harten, Lax and Van Leer, 1983) and Einfeldt (Einfeldt, 1988) to name just a few) have adopted approximations to its solution and thereby generated variants on the original method.

Godunov's method is only first order accurate but gives solutions which preserve monotonicity of the data. Indeed it was in his paper (Godunov, 1959) that Godunov presented his now famous theorem, which states that monotonicity preserving constant coefficient schemes can be at most first order accurate. It is this theorem which has led to much research in the area of non-linear schemes for hyperbolic conservation laws, since it is through the use of non-linear schemes that both monotonicity and high order accuracy can be achieved. Pioneering work carried out in the 1970s by Boris & Book (Boris and Book, 1973), van Leer (Van Leer, 1973; Van Leer, 1974; Van Leer, 1977; Van Leer, 1977; Van Leer, 1979) and Roe (Roe, 1981a) has been built upon by others (for example (Zalesak, 1979; Sweby, 1984; Colella and Woodward, 1984; Harten, Osher, Engquist and Chakravarthy, 1986; Harten, Engquist, Osher and Chakravarthy, 1987; Bell, Colella and Trangenstein, 1989)) resulting in an abundance of high resolution non-oscillatory schemes, many of which are based on Godunov's method. We now proceed to look at these.

In Section 2 we review Godunov's original scheme and its various formulations/interpretations. We then look at extensions to the scheme, firstly by replacing the exact Riemann Solver with an approximate one in Section 3 and secondly by modifying the data representation to obtain higher order accuracy in Section 4. Finally in Section 5 we look at some other advances of the Godunov methodology.

2. Godunov's Scheme

We are considering the numerical solution of the (system of) hyperbolic conservation laws

$$\mathbf{u}_t + \mathbf{f}(\mathbf{u})_x = \mathbf{0}. \quad (1)$$

Godunov's method considers the numerical values of the solution \mathbf{u}_i^n to

be the cell averages of the analytic solution $\mathbf{u}(x, t)$ at time level n ,

$$\mathbf{u}_i^n = \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} \mathbf{u}(x, n\Delta t) dx, \quad (2)$$

where for simplicity of notation we assume a regular grid. We therefore have a piecewise constant data representation (see Figure 1). At each cell boundary, the resulting Riemann problem is then solved and the union of all Riemann solutions averaged over each cell to give the updated numerical solution values (see Figure 2). Since each Riemann problem is solved in isolation, the need to avoid interaction suggests a CFL limit of $\frac{1}{2}$.

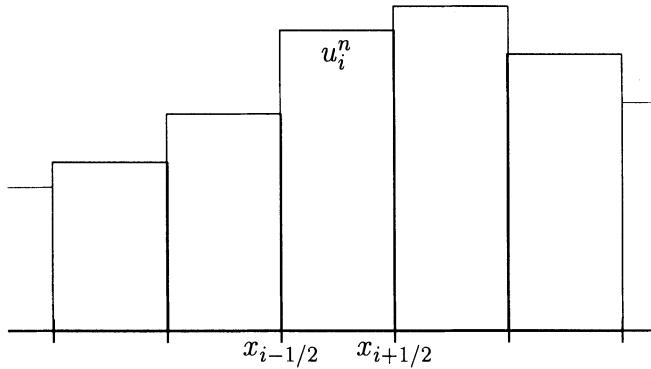


Figure 1. The piecewise constant data representation.

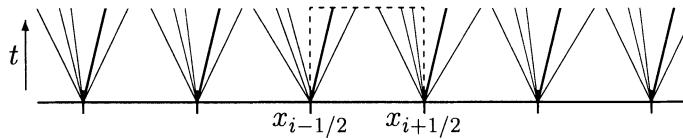


Figure 2. The resulting Riemann problems and their averaging.

For the special case of the scalar linear advection equation the scheme is easily interpreted as advection of the cell averages (Lagrangian stage), which are then remapped back onto the Eulerian grid (see Figure 3).

Godunov's scheme can be recast into Eulerian form by integrating (1) over the cell $[x_{i-1/2}, x_{i+1/2}] \times [n\Delta t, (n + 1)\Delta t]$:

$$\int_{t^n}^{t^{n+1}} \int_{x_{i-1/2}}^{x_{i+1/2}} \mathbf{u}_t(x, t) dx dt = - \int_{t^n}^{t^{n+1}} \int_{x_{i-1/2}}^{x_{i+1/2}} \mathbf{f}_x(\mathbf{u}(x, t)) dx dt,$$

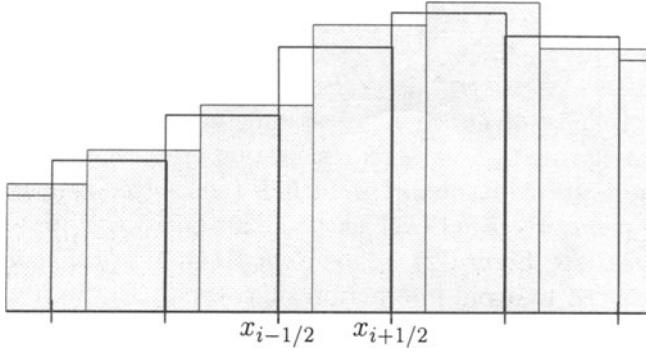


Figure 3. Lagrangian stage plus remap for linear advection.

whence

$$\int_{x_{i-1/2}}^{x_{i+1/2}} \mathbf{u}(x, t^{n+1}) - \mathbf{u}(x, t^n) dx = - \int_{t^n}^{t^{n+1}} \mathbf{f}(\mathbf{u}(x_{i+1/2}, t)) - \mathbf{f}(\mathbf{u}(x_{i-1/2}, t)) dt,$$

or

$$\begin{aligned} \Delta x (\mathbf{u}_i^{n+1} - \mathbf{u}_i^n) &= \\ &- \Delta t \left(\frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} \mathbf{f}(\mathbf{u}(x_{i+1/2}, t)) dt - \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} \mathbf{f}(\mathbf{u}(x_{i-1/2}, t)) dt \right). \end{aligned}$$

If we now define a numerical flux as

$$\mathbf{f}_{i-1/2}^n = \mathbf{f}(\mathbf{u}_{i-1}^n, \mathbf{u}_i^n) = \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} \mathbf{f}(\mathbf{u}(x_{i-1/2}, t)) dt, \quad (3)$$

then we can write Godunov's method in conservation form

$$\mathbf{u}_i^{n+1} = \mathbf{u}_i^n - \frac{\Delta t}{\Delta x} (\mathbf{f}_{i+1/2}^n - \mathbf{f}_{i-1/2}^n). \quad (4)$$

Looking at the numerical flux (3) we see that it does not depend on the whole Riemann solution, but only the flux of the state at $x_{i-1/2}$, and because the solution of the Riemann is self similar along rays $x/t = \text{constant}$ we can rewrite it as

$$\mathbf{f}_{i-1/2}^n = \mathbf{f}(\mathbf{u}_{i-1/2}^n). \quad (5)$$

We now see that since the detailed Riemann problem solutions within the cell are not important for calculation of the flux, they may be allowed

to interact so long as the solution at $x_{i-1/2}$ does not influence the state at $x_{i+1/2}$ and vice-versa. Thus it can be seen that this form of Godunov's scheme has a CFL condition of 1.

Finally we note that using Jensen's inequality (see for example (Harten, Lax and Van Leer, 1983)) it can be shown that Godunov's scheme is entropy satisfying, i.e. all shocks are physically correct.

Godunov's method solves the Riemann problem at each cell boundary exactly. However variants of the method may be generated by utilising an approximate Riemann solver instead. We now look at some possibilities.

3. Riemann Solvers

Whilst for some scalar conservation laws the Riemann problem is easily solved, this is not the case for non-linear systems of conservation laws. Here an iterative procedure is often required which, since this must be used at every cell boundary at every time step, will make it the most computationally expensive task of the whole method. To simplify the process and reduce this overhead approximate Riemann solvers, which do not employ iteration, are often used. This can be achieved either by approximating the Riemann states and applying the physical flux, or by approximating the numerical flux directly. In this paper we look at the latter, and outline the distinguishing features of some of the approximate Riemann solvers used. A full discussion of approximate Riemann solvers can be found, for example, in (Toro, 1997).

3.1. ROE

Perhaps the simplest approximate Riemann solver is that due to Roe (Roe, 1981). The system of conservation laws (1) may be written in quasi-linear form

$$\mathbf{u}_t + A(\mathbf{u})\mathbf{u}_x = \mathbf{0},$$

where $A(\mathbf{u})$ is the Jacobian matrix $\frac{\partial \mathbf{f}}{\partial \mathbf{u}}$. Roe linearises this form of the equations in each interval (x_{i-1}, x_i) by replacing the Jacobian by interval-wise constant matrices $\tilde{A}(\mathbf{u}_{i-1}, \mathbf{u}_i)$ which, for any two adjacent states $\mathbf{u}_L, \mathbf{u}_R$ satisfy

1. $\tilde{A}(\mathbf{u}_L, \mathbf{u}_R)$ is diagonalisable with real eigenvalues (hyperbolicity);
2. $\tilde{A}(\mathbf{u}_L, \mathbf{u}_R) \rightarrow A(\mathbf{u})$ as $\mathbf{u}_L, \mathbf{u}_R \rightarrow \mathbf{u}$ (consistency);
3. $\mathbf{f}(\mathbf{u}_L) - \mathbf{f}(\mathbf{u}_R) = \tilde{A}(\mathbf{u}_L, \mathbf{u}_R)(\mathbf{u}_L - \mathbf{u}_R)$ (conservation).

The first two conditions are readily satisfied if \tilde{A} is taken to be the Jacobian evaluated at an averaged state, i.e. $\tilde{A}(\mathbf{u}_L, \mathbf{u}_R) = A(\bar{\mathbf{u}})$. However a straight arithmetic average will not, in general, satisfy the final condition and instead a geometric average is often used (in the form of the form of

the arithmetic mean of an auxiliary vector, known as the parameter vector — see (Roe, 1981; Roe and Pike, 1984; Glaister, 1988)).

Once an \tilde{A} has been obtained, it is diagonalised $\tilde{X}\tilde{\Lambda}\tilde{X}^{-1}$ which results in a set of decoupled linear advection equations in each interval. The flux differences $\mathbf{f}_R - \mathbf{f}_L$ in each interval are decomposed onto the local eigenvectors

$$\Delta\mathbf{f} = \mathbf{f}_R - \mathbf{f}_L = \sum_{k=1}^n \tilde{\alpha}^{(k)} \tilde{\lambda}^{(k)} \tilde{\mathbf{x}}^{(k)}$$

where $\tilde{\lambda}^{(k)}$, $\tilde{\mathbf{x}}^{(k)}$ and $\tilde{\alpha}^{(k)}$ are the eigenvalue, eigenvector and coefficient for $\Delta\mathbf{u}$, respectively, corresponding to the k th characteristic field of \tilde{A} .

Whilst Roe's original scheme updated the solution by upwinding and directly adding these flux difference components, it may be placed in the framework of intercell fluxes by integration around the half cell $(x_{i-1/2}, x_i) \times (t^n, t^{n+1})$ (see (Harten, Lax and Van Leer, 1983)) resulting in the flux

$$\mathbf{f}_{i-1/2} = \frac{1}{2}(\mathbf{f}_{i-1} + \mathbf{f}_i) - \frac{1}{2} \sum_{k=1}^n \tilde{\alpha}_{i-1/2}^{(k)} |\tilde{\lambda}_{i-1/2}^{(k)}| \tilde{\mathbf{x}}_{i-1/2}^{(k)}.$$

In this formulation the \tilde{A} can be seen to be identified with the cell interfaces.

Because the resulting individual approximate Riemann problems are linear, their solutions contain only discontinuities and not expansion fans. For this reason Roe's original method is not entropy satisfying, although a number of *entropy fixes* have since been proposed (see for example (Harten and Hyman, 1983; Roe, 1992; Roe and Pike, 1984)).

3.2. OSHER

Whereas Roe's approximate Riemann solver approximates the solution of the Riemann problem using only discontinuities, Osher (Engquist & Osher (Engquist and Osher, 1981), Osher & Solomon (Osher and Solomon, 1982)) approximates the Riemann solution using only simple waves. The numerical flux for Osher's scheme may be written as

$$\mathbf{f}_{i-1/2} = \frac{1}{2}(\mathbf{f}_{i-1} + \mathbf{f}_i) - \frac{1}{2} \int_{\mathbf{u}_{i-1}}^{\mathbf{u}_i} |A(\mathbf{u})| d\mathbf{u},$$

where $|A(\mathbf{u})|$ is defined to be $X|\Lambda|X^{-1}$, (Λ is the matrix of eigenvalues of A and X the corresponding matrix of eigenvectors) and the integration path between \mathbf{u}_{i-1} and \mathbf{u}_i is taken in phase space over simple wave solutions. That is, for a system of n conservation laws, $n - 1$ intermediate states are found, connected via n simple waves. In his original scheme Osher took the paths to be in order of decreasing eigenvalue, however the reverse order

may also be selected (see (Hemker and Spekreijse, 1986)). Since the approximation uses only simple waves, not discontinuities, it can be shown to be entropy satisfying.

3.3. HARTEN, LAX & VAN LEER (HLL)

For a system of n conservation laws, both Roe's and Osher's approximate Riemann solvers use n intermediate states, connected in the case of Roe by discontinuities and in the case of Osher by simple waves. In 1983 Harten Lax & van Leer (Harten, Lax and Van Leer, 1983) proposed a much simpler approximation which assumes a Riemann solution consisting of just two waves separating three constant states. If $s_{i-1/2}^R$ and $s_{i-1/2}^L$ are upper and lower bounds, respectively, for the largest and smallest signal velocities resulting from the solution of the Riemann problem centred at $x_{i-1/2}$, then the approximate solution is taken to be

$$\tilde{\mathbf{u}}(x, t) = \begin{cases} \mathbf{u}_{i-1} & \text{if } \frac{x}{t} \leq s_{i-1/2}^L \\ \mathbf{u}_{i-1/2}^{HLL} & \text{if } s_{i-1/2}^L \leq \frac{x}{t} \leq s_{i-1/2}^R \\ \mathbf{u}_i & \text{if } s_{i-1/2}^R \leq \frac{x}{t} \end{cases},$$

where the intermediate state is obtained from conservation to be

$$\mathbf{u}_{i-1/2}^{HLL} = \frac{s_{i-1/2}^R \mathbf{u}_i - s_{i-1/2}^L \mathbf{u}_{i-1}}{s_{i-1/2}^R - s_{i-1/2}^L} - \frac{\mathbf{f}_i - \mathbf{f}_{i-1}}{s_{i-1/2}^R - s_{i-1/2}^L}. \quad (6)$$

Integrating this solution substituted into the conservation law over the half cell $(x_{i-1/2}, x_i) \times (t^n, t^{n+1})$ results in the HLL flux

$$\mathbf{f}_{i-1/2} = \begin{cases} \mathbf{f}_{i-1} & \text{if } \frac{x}{t} \leq s_{i-1/2}^L \\ \mathbf{f}_{i-1/2}^{HLL} & \text{if } s_{i-1/2}^L \leq \frac{x}{t} \leq s_{i-1/2}^R \\ \mathbf{f}_i & \text{if } s_{i-1/2}^R \leq \frac{x}{t} \end{cases},$$

where

$$\mathbf{f}_{i-1/2}^{HLL} = \frac{s_{i-1/2}^R \mathbf{f}_{i-1} - s_{i-1/2}^L \mathbf{f}_i + s_{i-1/2}^L s_{i-1/2}^R (\mathbf{u}_i - \mathbf{u}_{i-1})}{s_{i-1/2}^R - s_{i-1/2}^L}.$$

Harten *et al.* noted that the intermediate state, as given by (6), is a mean value of the exact Riemann solution, and it therefore follows from Jensen's inequality that the scheme is entropy satisfying. They also note that if the two states \mathbf{u}_{i-1} and \mathbf{u}_i can be connected by a shock of the first or n th family, then the correct shock speed is obtained and the solution is exact. They then proceed to derive a scheme with two intermediate states

which has the additional property that if the end states can be connected by a shock or contact discontinuity of *any* family, then the approximate Riemann solver does so. Toro, Spruce & Speares (Toro, Spruce and Speares, 1994) also modified the original HLL approximate solver restoring missing contact or shear waves, naming it the HLLC solver.

It now remains to specify the upper and lower bounds $s_{i-1/2}^R$ and $s_{i-1/2}^L$. One possibility is to evaluate them directly, or, as suggested by Davis (Davis, 1988) to use the maximum eigenvalue evaluated at the right state and the minimum eigenvalue evaluated at the left state respectively, or to take the maximum of the largest eigenvalue at either state, and the minimum of the smallest. Another alternative, also suggested by Davis (Davis, 1988) and Einfeldt (Einfeldt, 1988) is to use Roe's averaged eigenvalues as estimates.

4. Higher Order Extensions

First order methods such as Godunov's tend to be very diffusive, smearing the discontinuities that often arise in the solution of conservation laws. However, classical higher order methods (for example Lax–Wendroff (Lax and Wendroff, 1960)) whilst giving sharper features also produce spurious oscillations around discontinuities, possibly resulting in unphysical values (for example negative density) and/or violation of stability bounds thus causing a breakdown of the solution. Godunov (Godunov, 1959) proved that this was inevitable for constant coefficient schemes, which could not be both monotonicity preserving and higher than first order accurate. This led to much work on non-linear schemes, both practical and theoretical.

One of the theoretical advances was the adoption of *total variation* as a monitor of spurious oscillations. This stems from the fact that, for the scalar conservation law, the analytic total variation,

$$\text{TV}(u) = \int |u_x| dx,$$

does not increase (and only decreases across shocks) (Lax, 1973). Harten (Harten, 1983) proposed that schemes should mimic this behaviour with their discrete total variation,

$$\text{TV}(u^n) = \sum_i |u_i - u_{i-1}|,$$

i.e.

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

and christened such schemes Total Variation Diminishing (TVD). He also gave a set of algebraic criteria for the (non-linear) coefficients of a scheme to

satisfy which are sufficient to show the scheme to be TVD. These algebraic conditions are more commonly used for non Godunov-type schemes, with a geometric approach being used for Godunov-type methods. Conservative schemes which are TVD can be shown to converge to a weak solution of the conservation law.

Some of the non-linear schemes developed, such as Flux Corrected Transport (FCT) (Boris and Book, 1973; Zalesak, 1979) and Flux Limiters (Sweby, 1984), can not be considered as Godunov-type schemes except possibly for the linear case. Others however are either direct extensions of Godunov's method or at least use the same methodology in their construction.

We now look at a set of modifications to Godunov's method which result in higher order accuracy.

As already mentioned, Godunov's method assumes the data to consist of piecewise constant cell averages which are advanced in time by solving the Riemann problems at each cell interface and then re-averaging. This results in a first order method. To achieve higher order accuracy we must change the data representation. For clarity we describe here the data representations for the scalar non-linear conservation law. The methodology is usually extended to systems via characteristic decomposition of the states.

4.1. MUSCL AND VARIANTS

Van Leer (Van Leer, 1977; Van Leer, 1979) in his MUSCL¹ scheme replaced Godunov's piecewise constant representation with a piecewise linear one (see Figure 4).

This piecewise linear representation was constructed to maintain conservation by defining the cell representation to be

$$u_i(x) = u_i^n + \frac{\Delta_i u}{\Delta x}(x - x_i),$$

where u_i^n is the Godunov cell average (2) and the slope $\frac{\Delta_i u}{\Delta x}$ must be defined. Van Leer (Van Leer, 1977) gave three possibilities for $\Delta_i u$,

1. centred differencing of the piecewise constant cell averages $\Delta_i u = \frac{1}{2}(u_{i+1} - u_{i-1})$;
2. differencing of the underlying continuous function $\Delta_i u = (u(x_{i+1/2}, t^n) - u(x_{i-1/2}, t^n))$, i.e. a difference of the unaveraged values – this leads to the necessity to evolve $\Delta_i u$ as well as u ;

¹MUSCL - Monotonic Upstream-centred Scheme for Conservation Laws - was actually the name of Paul Woodward's computer code incorporating van Leer's ideas, but the name has stuck with this type of scheme

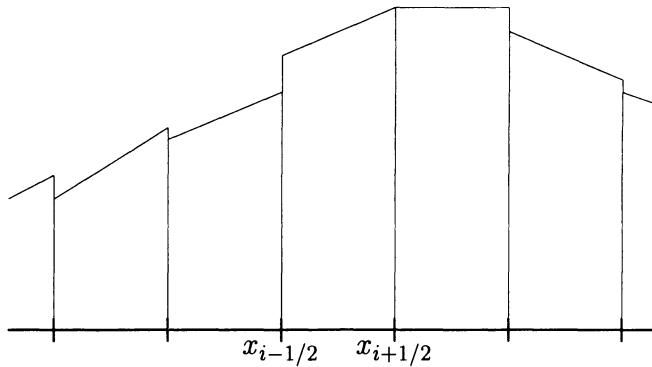


Figure 4. The piecewise linear data representation.

3. maintaining the first moment of the underlying analytical solution $\Delta_i u = \frac{12}{(\Delta x)^2} \int_{x_{i-1/2}}^{x_{i+1/2}} u(x, t^n) (x - x_i) dx$. Again this leads to independent updates of the slope and cell average.

If we calculate the slopes in any of these fashions there is the potential of producing a data representation which has a higher maximum or lower minimum than the piecewise constant average (see Figure 5), i.e. increasing the total variation of the data representation. Viewing the Lagrangian plus remap interpretation of Godunov's scheme, it can easily be seen that this increase will be maintained, and spurious oscillations may ensue.

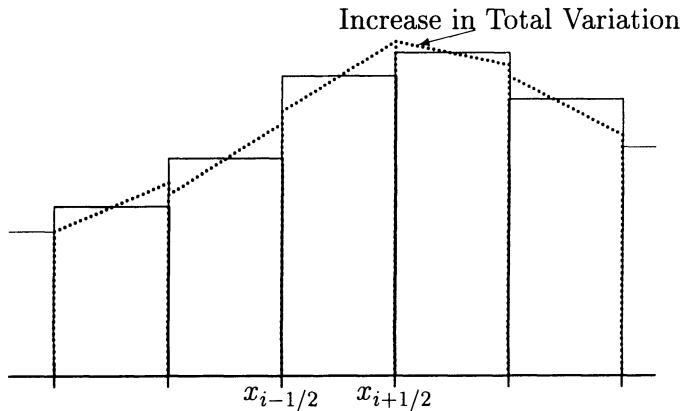


Figure 5. An increase in total variation.

To avoid this increase in total variation, van Leer limited the gradient of his slopes. He achieved this by defining a monotonised slope as

$$(\Delta_i u)_{\text{mono}} = \begin{cases} \min\{2|\Delta u_{i-1/2}|, |\Delta_i u|, 2|\Delta u_{i+1/2}|\} \operatorname{sgn} \Delta_i u & \text{if } \operatorname{sgn} \Delta u_{i-1/2} = \operatorname{sgn} \Delta u_{i+1/2} = \operatorname{sgn} \Delta_i u \\ 0 & \text{otherwise} \end{cases},$$

which may be applied to any definition of $\Delta_i u$ and where $\Delta u_{i-1/2} = u_i - u_{i-1}$. This prevents the linear function taking values outside of the range of the neighbouring mesh averages and reduces the slope to zero, i.e. reverting to piecewise constant, where there is an extremum of the data. For his first choice of slope, van Leer also gave an improved limiting of

$$(\Delta_i u)_{\text{mono}} = \begin{cases} \frac{2\Delta u_{i-1/2}\Delta u_{i+1/2}}{\Delta u_{i-1/2} + \Delta u_{i+1/2}} & \text{if } \operatorname{sgn} \Delta u_{i-1/2} = \operatorname{sgn} \Delta u_{i+1/2} \\ 0 & \text{otherwise} \end{cases}, \quad (7)$$

which has the effect of taking the harmonic mean of $\Delta u_{i-1/2}$ and $\Delta u_{i+1/2}$ instead of their algebraic mean as in the non-monotonised slope.

This slope limiting has much in common with flux limiters (Sweby, 1984), except that here it is the slope which is being limited. If we define $r_{i-1/2} = \Delta u_{i-1/2}/\Delta u_{i+1/2}$ it can be seen that (7) can be written as

$$(\Delta_i u)_{\text{mono}} = \frac{r_{i-1/2} + |r_{i-1/2}|}{1 + |r_{i-1/2}|} = \phi_{VL}(r_{i-1/2}),$$

where $\phi_{VL}(r)$ is van Leer's flux limiter (Van Leer, 1974; Sweby, 1984). Indeed other flux limiters can be used as slope limiters, for example Goodman & Le Veque (Goodman and Le Veque, 1988) use Roe's minmod flux limiter (Roe, 1981a; Sweby, 1984). It is important to note however that, even though slope limiter and flux limiter schemes are equivalent for linear scalar equations, they are two distinct types of method, with only the former fitting into the Godunov-type framework.

At the cell interfaces of the piecewise linear data representation we now have a set of so-called generalised Riemann problems, i.e. a discontinuity separating two linear states. These are not as easily solved as the basic Riemann problem and the wave paths are now curves rather than straight lines in $x-t$ space. This means that the Lagrangian advection step is not so readily achieved except for linear problems, and in conservation form (4) the numerical flux (3) no longer reduces to (5). Whilst some methods are based on the generalised Riemann problem (Ben-Artzi and Falcovitz, 1984; Ben-Artzi and Falcovitz, 1985), more often an approximate Riemann solution is used, for example Goodman & Le Veque (Goodman and Le Veque, 1988) approximate the flux by a linear function near the cell interfaces.

An alternative approach, due to Hancock (Hancock, 1980) was noted by van Leer (Van Leer, 1984) whereby a second order accurate scheme is obtained by evolving the (extrapolated) cell boundary values

$$u_i^{n,L} = u_i - \frac{1}{2}(\Delta_i u)_{\text{mono}}, \quad \text{and} \quad u_i^{n,R} = u_i + \frac{1}{2}(\Delta_i u)_{\text{mono}},$$

obtained by the piecewise linear data representation, by a time $\frac{1}{2}\Delta t$. This is achieved via a Taylor series expansion

$$\begin{aligned} u_i^{n+1/2,L} &= u_i^{n,L} + \frac{\Delta t}{2} u_t, i \\ &= u_i^{n,L} - \frac{\Delta t}{2} f_x, i \quad \text{using (1)} \\ &\approx u_i^{n,L} - \frac{1}{2} \frac{\Delta t}{\Delta x} (f(u_i^{n,R}) - f(u_i^{n,L})), \end{aligned}$$

and similarly

$$u_i^{n+1/2,R} \approx u_i^{n,R} - \frac{1}{2} \frac{\Delta t}{\Delta x} (f(u_i^{n,R}) - f(u_i^{n,L})).$$

These advanced states are then used as piecewise constant data for a conventional Riemann problem at the intercell boundary,

$$u_t + f(u)_x = 0 \quad u(x, 0) = \begin{cases} u_{i-1}^{n+1/2,R}, & x < 0, \\ u_i^{n+1/2,L}, & x > 0, \end{cases},$$

to obtain the similarity solution $u_{i-1/2}(x/t)$ which is in turn used to obtain the intercell numerical flux via (5) to be used in (4).

In their Higher Order Godunov method, Bell, Colella & Trangenstein (Bell, Colella and Trangenstein, 1989) extended this technique to general systems of hyperbolic conservation laws, using Osher's approximate Riemann solver, with additional modifications needed to treat loss of strict hyperbolicity which arise due to eigenvector deficiencies that may arise, for example in the black–oil model of petroleum engineering.

4.2. PPM AND ENO

Woodward & Colella (Woodward and Colella, 1984; Colella and Woodward, 1984) extended the idea of MUSCL further by constructing a piecewise parabolic data representation. This representation is limited so as to avoid overshoots and undershoots and incorporates a discontinuity detection mechanism so as to sharpen any discontinuities of the data. This piecewise parabolic representation is then advanced using either a Lagrangian

step followed by a remap, or in conservation form, resulting in the third order Piecewise Parabolic Method (PPM). (For simple problems the method is third order in space and time; however for more complicated situations approximations in the Riemann solution can degrade the time accuracy to second order.)

Harten, Osher, Chakravarthy and Engquist (Harten, Osher, Engquist and Chakravarthy, 1986; Harten and Osher, 1987; Harten, Engquist, Osher and Chakravarthy, 1987) and later Osher & Shu (Shu and Osher, 1988; Shu and Osher, 1989) extended the idea of polynomial data representation even further. The technique used is similar to that employed in the MUSCL and PPM schemes, except that the data representation constructed from the cell averages $\{u_i^n\}$ does not damp the values of local extrema, as in the aforementioned schemes, and is even allowed occasionally to accentuate these local features.

The Essentially Non-Oscillatory (ENO) scheme, as it is called, starts from the cell averages $\bar{u}^n = \{u_i^n\}$ and constructs the approximate function $u_{\Delta x}(x; t^n) = R(x; \bar{u}^n)$, where $R(x; \bar{u}^n)$ is a piecewise polynomial in x of degree $p - 1$ satisfying:

1. $R(x; \bar{u}^n) = u(x, t^n) + O(\Delta x^p)$ where the functions are smooth;
2. $R(x; \bar{u}^n)$ is conservative, i.e. $\frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} R(x; \bar{u}^n) dx = u_i^n$;
3. $R(x; \bar{u}^n)$ is essentially non-oscillatory,
i.e. $\text{TV}(R(\cdot; \bar{u}^n)) \leq \text{TV}(u(\cdot, t^n)) + O(\Delta x^p)$.

Both MUSCL ($p = 2$) and PPM ($p = 3$) fit into this framework, except that they have the more restrictive condition of $\text{TV}(R(\cdot; \bar{u}^n)) \leq \text{TV}(u(\cdot, t^n))$, i.e. TVD.

Once the data has been reconstructed, the solution of the conservation law (1) with initial data $u_{\Delta x}(\cdot, t^n)$ is calculated and the solution re-averaged to obtain updated cell averages u_i^{n+1} .

The key step of ENO is in the reconstruction, the essence of which is as follows. The interpolant $R(x; \bar{u}^n)$ is built up in stages using Newton interpolation. Initially we may construct a local linear interpolant in the cell $(x_{i-1/2}, x_{i+1/2})$ either using u_{i+1} and u_i or u_i and u_{i-1} . The pair with smallest difference is chosen; this process being repeated for each cell. Next a quadratic interpolant for each cell is constructed by adding an additional interpolation point — this can be either the value to the left or right of the previous stencil. For example, if u_{i+1} and u_i had been chosen to form the linear interpolation for our cell, then we can add in either u_{i+2} or u_{i-1} . The one which gives the smoothest interpolant (as monitored by comparison of divided differences) is chosen. This is done for each cell and the method applied recursively until the desired degree of interpolation is reached. A variant of ENO is Weighted ENO (WENO) (Liu, Osher and Chan, 1994) where a linear combination of the candidate stencils for interpolation is

taken. Subcell resolution has also been used (Harten, 1989) in order to sharpen contact discontinuities.

ENO can be shown to be Total Variation Bounded (TVB), i.e. $\text{TV}(u^n) \leq C\text{TV}(u^0)$ which means theoretically that solutions still converge as for TVD schemes, and practically that small oscillations on the scale of truncation error may appear but usually vanish if the solution is adequately resolved.

4.3. WAF

Toro (Toro, 1989; Toro, 1991; Toro, 1997) achieves second order accuracy in a different way. Instead of enhancing the data representation, as in the methods previously described, he exploits more of the information provided by the full solution of the conventional Riemann problem. Whilst Godunov's method evaluates the intercell flux only along $x = x_{i-1/2}$ (see (5)), Toro takes a weighted average of the flux vector across the whole wave structure of the Riemann solution at time $t^n + \frac{\Delta t}{2}$. His intercell flux is defined to be

$$\mathbf{f}_{i-1/2} = \frac{1}{\Delta x} \int_{x_{i-1}}^{x_i} \mathbf{f}(\mathbf{u}_{i-1/2}(x, t^n + \frac{\Delta t}{2})) dx, \quad (8)$$

where the integration is depicted in Figure 6.

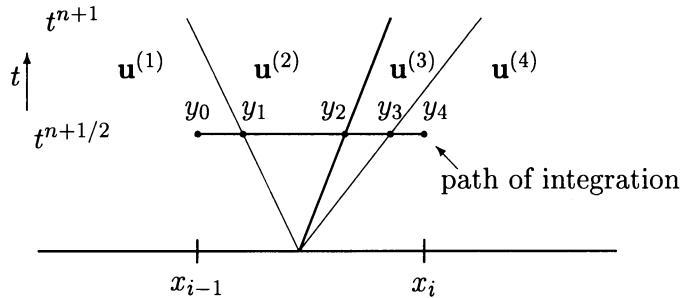


Figure 6. Calculation of the WAF flux.

If we consider first the case where there is no expansion wave (as shown in Figure 6) we see that the integral in (8) may be evaluated to give

$$\mathbf{f}_{i-1/2} = \sum_{k=1}^{n+1} \beta_k \mathbf{f}_{i-1/2}^{(k)}, \quad (9)$$

where the fluxes $\mathbf{f}_{i-1/2}^{(k)}$ are evaluated at the constant states, i.e. $\mathbf{f}_{i-1/2}^{(k)} = \mathbf{f}(\mathbf{u}^{(k)})$ and the weights β_k are the normalised lengths

$$\beta_k = \frac{y_k - y_{k-1}}{\Delta x}.$$

The weighted sum (9) leads to the name of the Weighted Average Flux (WAF) method. If there is an expansion wave present in the Riemann solution then Toro (Toro, 1997) suggests that the integral across it could be evaluated exactly (giving an extra weight) or the expansion could be combined with the closest constant state (taking the value of the state to be $\mathbf{u}_{i-1/2}(0)$ for a sonic rarefaction). In order to solve the Riemann problem to obtain the solution at $t^{n+1/2}$ either the exact solution can be used or an approximate Riemann solver employed.

As it stands the method is not TVD, however Toro applies a flux limiter type device to the flux differences across waves of the Riemann solution - see (Toro, 1997) for details.

5. And there's more...

In the previous sections we have only been able to give a brief description of just some of the Godunov methods in the literature. Any literature search will reveal a wealth of material on the subject, some of it extensions and new applications of established techniques and some of it novel in its own right. We conclude by looking at a couple of these developments.

5.1. SOURCE TERMS

In the exposition of the various Godunov methods, we have restricted our attention to homogeneous conservation laws. However, many models of physical situations involve source terms, giving the inhomogeneous equation

$$\mathbf{u}_t + \mathbf{f}_x(\mathbf{u}) = \mathbf{S}(x, \mathbf{u}). \quad (10)$$

Although much work has been carried out in the investigation of how to incorporate source terms into numerical methods for conservation laws, this is still an open issue. The simplest approach is to use fractional steps, splitting the non-homogeneous equation (10) into the homogeneous equation (1) supplemented by the ordinary differential equation

$$\mathbf{u}_t = \mathbf{S}(x, \mathbf{u}),$$

solving them alternatively for each timestep. This avoids incorporating the source term directly into the numerical solver for the conservation law.

However this approach can perform badly when solving near-steady state flows, since the contributions from the two separate stages can be large but non-cancelling. The most obvious inclusion of a source term into a Godunov method would add extra complication to the Riemann solution and calculation of the intercell fluxes since the Riemann solution is no longer constant on rays $x/t = \text{constant}$.

Le Veque (Le Veque, 1998), however, takes a novel approach. He introduces a new discontinuity in the centre of each cell by decomposing the piecewise constant data representation into two different states \mathbf{u}_i^- and \mathbf{u}_i^+ , (see Figure 7). This is done in such a manner so as to be conservative, and if possible so that

$$\mathbf{f}(\mathbf{u}_i^+) - \mathbf{f}(\mathbf{u}_i^-) = \mathbf{S}(x_i, \mathbf{u}_i)\Delta x. \quad (11)$$

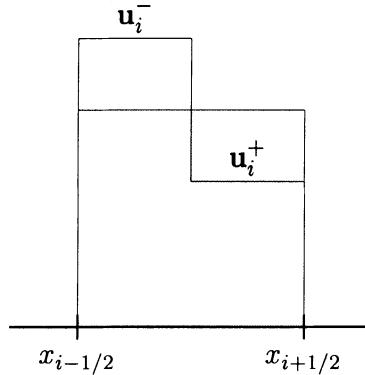


Figure 7. The additional Riemann problem.

This should then ensure that the effect of the source term in the cell is exactly cancelled by the waves resulting from this new Riemann problem. This has the implication that this new Riemann problem need not be solved, nor the source term explicitly included in the scheme, except for choosing the new states \mathbf{u}_i^- and \mathbf{u}_i^+ , via (11), which are then used with the standard Godunov approach or its high order extensions. If it is not possible to choose the new states to satisfy (11) then an additional term accounting for the discrepancy is added to the update for the cell (see (Le Veque, 1998)).

5.2. CENTRAL SCHEMES

In (Nessyahu and Tadmor, 1990) Nessyahu & Tadmor introduced a central scheme based on the Godunov philosophy, although it avoids explicit solution of Riemann problems. The first order case is a staggered grid version

of the Lax–Friedrichs scheme, and is constructed in the following manner. Like Godunov's scheme the data is taken to be the set of cell averages, resulting in a set of Riemann problems at their interfaces (Figures 1 and 2). However, instead of re-averaging the solution of adjacent Riemann problems over the cell $(x_{i-1/2}, x_{i+1/2})$, the solution of just a single Riemann problem is averaged over the staggered cell (x_{i-1}, x_i) , see Figure 8.

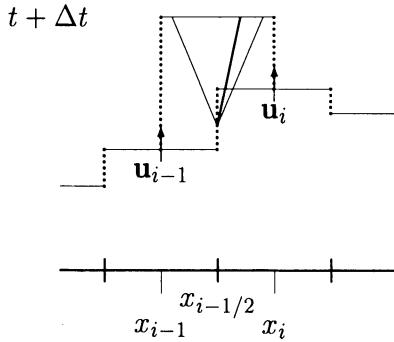


Figure 8. Central differencing by Godunov type scheme.

Equations (4) and (3) now become

$$\mathbf{u}_{i-1/2}^{n+1} = \mathbf{u}_{i-1/2}^n - \frac{\Delta t}{\Delta x} (\mathbf{f}_i^n - \mathbf{f}_{i-1}^n),$$

and

$$\mathbf{f}_i^n = \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} \mathbf{f}(\mathbf{u}(x_i, t)) dt \quad (12)$$

where

$$\mathbf{u}_{i-1/2}^n = \frac{1}{\Delta x} \left[\int_{x_{i-1}}^{x_{i-1/2}} \mathbf{u}_{i-1}(x) dx + \int_{x_{i-1/2}}^{x_i} \mathbf{u}_i(x) dx \right]. \quad (13)$$

Of course, for the piecewise constant representation we are considering at present, (13) reduces to $\mathbf{u}_{i-1/2}^n = \frac{1}{2}(\mathbf{u}_{i-1}^n + \mathbf{u}_i^n)$. The mid-points are evolved according to the conservation law (1) with initial data $\mathbf{u}_i^n(x)$ and integrals in the flux definition (12) are evaluated using suitable quadrature. (The evolved mid-cell values will remain continuous for small enough time.) For the piecewise constant case, the mid-cell values remain constant, giving a flux $\mathbf{f}_i^n = \mathbf{f}(\mathbf{u}_i^n)$ and the resulting scheme becomes

$$\mathbf{u}_{i-1/2}^{n+1} = \frac{1}{2}(\mathbf{u}_{i-1}^n + \mathbf{u}_i^n) - \frac{\Delta t}{\Delta x} (\mathbf{f}(\mathbf{u}_i^n) - \mathbf{f}(\mathbf{u}_{i-1}^n)),$$

which is a staggered Lax–Friedrichs scheme.

To extend this scheme to higher order accuracy, the MUSCL approach is used, replacing the piecewise constant data by a piecewise linear representation, suitably limited to ensure TVD. For the higher order case, the integrals in (13) are evaluated for the piecewise linear representation, which will evolve in time, and the second order accurate mid-point quadrature rule is used to evaluate the integral in (12).

The method is extended to non-staggered grids in (Jiang, Levy, Lin, Osher and Tadmor, 1998).

References

- J. B. Bell, P. Colella, and J. A. Trangenstein. Higher-order Godunov methods for general systems of hyperbolic conservation-laws. *J. Computational Phys.*, **82**, pp 362–397, (1989).
- M. Ben-Artzi and J. Falcovitz. A 2nd-order Godunov-type scheme for compressible fluid-dynamics. *J. Computational Phys.*, **55**, pp 1–32, (1984).
- M. Ben-Artzi and J. Falcovitz. GRP - An analytic approach to high-resolution upwind schemes for compressible fluid-flow. *Lecture Notes Phys.*, **218**, pp 87–91, (1985).
- J.P. Boris and D.L. Book. Flux-corrected transport. I. SHASTA, a fluid-transport algorithm that works. *J. Computational Phys.*, **11**, pp 38–69, (1973).
- P. Colella and P. R. Woodward. The piecewise parabolic method (PPM) for gas-dynamical simulations. *J. Computational Phys.*, **54**, pp 174–201, (1984).
- S. F. Davis. Simplified 2nd-order Godunov-type methods. *Siam J. On Scientific Statistical Computing*, **9**, pp 445–473, (1988).
- B. Einfeldt. On Godunov-type methods for gas-dynamics. *Siam J. On Numerical Analysis*, **25**, pp 294–318, (1988).
- B. Engquist and S. Osher. One-sided difference approximations for non-linear conservation-laws. *Mathematics Computation*, **36**, pp 321–351, (1981).
- P. Glaister. Flux difference splitting for the Euler equations with axial symmetry. *J. Engineering Mathematics*, **22**, pp 107–121, (1988).
- S. K. Godunov. A difference scheme for numerical computation of discontinuous solutions of equations of fluid dynamics. *Math. Sbornik*, **47**, pp 271–306, (1959).
- J. B. Goodman and R. J. Le Veque. A geometric approach to high-resolution TVD schemes. *Siam J. On Numerical Analysis*, **25**, pp 268–284, (1988).
- S. Hancock. Physics International, San Leandro, California. Unpublished Private Communication to van Leer, (1980).
- A. Harten. High-resolution schemes for hyperbolic conservation-laws. *J. Computational Phys.*, **49**, pp 357–393, (1983).
- A. Harten. ENO schemes with subcell resolution. *J. Computational Phys.*, **83**, pp 148–184, (1989).
- A. Harten, B. Engquist, S. Osher, and S. R. Chakravarthy. Uniformly high-order accurate essentially nonoscillatory schemes, III. *J. Computational Phys.*, **71**, pp 231–303, (1987).
- A. Harten and J. M. Hyman. Self-adjusting grid methods for one-dimensional hyperbolic conservation-laws. *J. Computational Phys.*, **50**, pp 235–269, (1983).
- A. Harten, P. D. Lax, and B. Van Leer. On upstream differencing and Godunov-type schemes for hyperbolic conservation-laws. *Siam Review*, **25**, pp 35–61, (1983).
- A. Harten and S. Osher. Uniformly high-order accurate nonoscillatory schemes .I. *Siam J. On Numerical Analysis*, **24**, pp 279–309, (1987).
- A. Harten, S. Osher, B. Engquist, and S. R. Chakravarthy. Some results on uniformly high-order accurate essentially nonoscillatory schemes. *Applied Numerical Mathematics*,

- atics*, **2**, pp 347–377, (1986).
- P.W. Hemker and S.P. Spekreijse. Multiple grid and Osher's scheme for the efficient solution of the steady Euler equations. *Applied Numerical Mathematics*, **2**, pp 475–493, (1986).
- G. S. Jiang, D. Levy, C. T. Lin, S. Osher, and E. Tadmor. High-resolution nonoscillatory central schemes with nonstaggered grids for hyperbolic conservation laws. *Siam J. On Numerical Analysis*, **35**, pp 2147–2168, (1998).
- P. D. Lax. *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves*, volume 11 of *Regional Conference Series in Applied Mathematics*. SIAM, (1973).
- P. D. Lax and B. Wendroff. Systems of conservation laws. *Comm. Pure & Appl. Math.*, **13**, pp 217–237, (1960).
- R. J. Le Veque. Balancing source terms and flux gradients in high-resolution Godunov methods: The quasi-steady wave-propagation algorithm. *J. Computational Phys.*, **146**, pp 346–365, (1998).
- X. D. Liu, S. Osher, and T. Chan. Weighted essentially nonoscillatory schemes. *J. Computational Phys.*, **115**, pp 200–212, (1994).
- H. Nessyahu and E. Tadmor. Non-oscillatory central differencing for hyperbolic conservation-laws. *J. Computational Phys.*, **87**, pp 408–463, (1990).
- S. Osher and F. Solomon. Upwind difference-schemes for hyperbolic systems of conservation-laws. *Mathematics Computation*, **38**, pp 339–374, (1982).
- P. L. Roe. Approximate Riemann solvers, parameter vectors, and difference-schemes. *J. Computational Phys.*, **43**, pp 357–372, (1981).
- P. L. Roe. Sonic flux formulas. *Siam J. On Scientific Statistical Computing*, **13**, pp 611–630, (1992).
- P.L. Roe. Numerical algorithms for the linear wave equation. Technical Report 81047, Royal Aircraft Establishment, (1981).
- P.L. Roe and J. Pike. Efficient construction and utilisation of approximate Riemann solutions. In *Computing Methods in Applied Science and Engineering*. North Holland, (1984).
- C. W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *J. Computational Phys.*, **77**, pp 439–471, (1988).
- C. W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes .2. *J. Computational Phys.*, **83**, pp 32–78, (1989).
- P. K. Sweby. High-resolution schemes using flux limiters for hyperbolic conservation-laws. *Siam J. On Numerical Analysis*, **21**, pp 995–1011, (1984).
- E. F. Toro. A weighted average flux method for hyperbolic conservation laws. *Proceedings Royal Soc. London Series A-Mathematical Phys. Engineering Sciences*, **423**, pp 401–418, (1989).
- E. F. Toro. A linearized Riemann solver for the time-dependent Euler equations of gas-dynamics. *Proceedings Royal Soc. London Series A-Mathematical Phys. Engineering Sciences*, **434**, pp 683–693, (1991).
- E. F. Toro. *Riemann Solvers and Numerical Methods for Fluid Dynamics. A Practical Introduction*. Springer, (1997).
- E.F. Toro, M. Spruce, and W. Speares. Restoration of the contact surface in the HLL-Riemann solver. *Shock Waves*, **4**, pp 25–34, (1994).
- B. Van Leer. Towards the ultimate conservative difference scheme I. The quest of monotonicity. *Springer Lecture Notes in Physics*, **18**, pp 163–168, (1973).
- B. Van Leer. Towards the ultimate conservative difference scheme II. Monotonicity and conservation combined in a second order scheme. *J. Computational Phys.*, **14**, pp 361–370, (1974).
- B. Van Leer. Towards the ultimate conservative difference scheme III. Upstream-centered finite-difference schemes for ideal compressible flow. *J. Computational Phys.*, **23**, pp 263–275, (1977).
- B. Van Leer. Towards the ultimate conservative difference scheme IV. A new approach

- to numerical convection. *J. Computational Phys.*, **23**, pp 276–299, (1977).
- B. Van Leer. Towards the ultimate conservative difference scheme V. A second order sequel to Godunov's method. *J. Computational Phys.*, **32**, pp 101–136, (1979).
- B. Van Leer. On the relation between the upwind-differencing schemes of Godunov, Engquist-Osher and Roe. *Siam J. On Scientific Statistical Computing*, **5**, pp 1–20, (1984).
- B. Van Leer. Godunov's method for gas-dynamics: Current applications and future developments. *J. Computational Phys.*, **132**, p 1, (1997).
- B. Van Leer. An introduction to the article "Reminiscences about difference schemes" by S. K. Godunov. *J. Computational Phys.*, **153**, pp 1–5, (1999).
- P. Woodward and P. Colella. The numerical-simulation of two-dimensional fluid-flow with strong shocks. *J. Computational Phys.*, **54**, pp 115–173, (1984).
- S.T. Zalesak. Fully dimensional flux corrected transport algorithms for fluids. *J. Computational Phys.*, **31**, pp 335–362, (1979).

CENTRED UNSPLIT FINITE VOLUME SCHEMES FOR MULTI-DIMENSIONAL HYPERBOLIC CONSERVATION LAWS

E. F. TORO

Department of Computing and Mathematics,

Manchester Metropolitan University,

Chester Street, Manchester, M1 5GD, U.K.

Email: E.F.Toro@doc.mmu.ac.uk

Webpage: www.doc.mmu.ac.uk/STAFF/E.F.Toro

AND

W. HU

Department of Computing and Mathematics,

Manchester Metropolitan University,

Chester Street, Manchester, M1 5GD, U.K.

Email: W.Hu@doc.mmu.ac.uk

Abstract. New unsplit finite volume centred schemes are presented. The construction of the schemes relies on the finite-volume frame work suggested by Billett and Toro, and on two existent one-dimensional centred schemes, namely the FORCE and the SLIC schemes. First and second order accurate schemes are constructed. These are found to possess improved stability properties as compared to existent finite volume methods. An application to the two-dimensional shallow water equations shows that the proposed schemes are accurate, robust and efficient.

1. Introduction

Over the past five decades the number and quality of numerical methods to solve hyperbolic conservation laws have increased dramatically. Amongst them are finite difference methods, finite element methods and finite volume methods. In this paper, we utilise finite volume methods, which combine

the simplicity of finite-difference methods with the geometric flexibility of finite element methods. Calculating the intercell flux is a critical step in finite volume schemes. Several approaches are available for obtaining the numerical flux, one of which being upwind-type schemes that have become a mature class of numerical methods for solving hyperbolic conservation laws (Hirsch, 1988), (Leveque, 1992), (Toro, 1999a). But for some systems the solution of Riemann problems needed in upwind approaches is cumbersome and sometimes is impossible. In such cases, central schemes are a valuable contribution (Nessyahu and Tadmor, 1990; Toro and Billett, 2000).

In this paper we use the framework of the upwind finite volume WAF-type scheme of Billett and Toro (Billett and Toro, 1997a) to design new centred finite volume methods by combining one first order (First Order Centred Scheme, FORCE) and one second order (Slope Limiter Centred Scheme, SLIC) centred schemes. In section 2, the two dimensional WAF-type finite volume scheme is reviewed. Our new centred finite volume approach is introduced in section 3. The analysis of accuracy and stability of the centred schemes is carried out in section 4. Finally, in section 5, the reflection of an oblique bore wave occurring in shallow water flows is simulated by the new centred finite volume scheme. Both analysis and numerical experiments confirm the improved stability and high resolution properties of the scheme. As the schemes do not require the (explicit) use of upwind directions, their implementation is very simple, which makes the schemes attractive for practical applications, particularly for very complex problems.

2. Review of 2D WAF-type Schemes

A WAF-type (Weighted Average Flux) unsplit approach (Billett and Toro, 1997a) for multidimensional hyperbolic conservation laws is the framework of this paper. It encompasses multidimensional upwinding aspects and good stability properties by combining one first order and one second order upwind-type scheme. Consider a hyperbolic system of equations,

$$\mathbf{U}_t + \mathbf{F}(\mathbf{U})_x + \mathbf{G}(\mathbf{U})_y = \mathbf{0}, \quad (1)$$

where \mathbf{U} is the vector of conservative variables and \mathbf{F} and \mathbf{G} are flux vectors. On a Cartesian mesh, the solution can be obtained by an explicit unsplit finite volume method,

$$\mathbf{U}_{i,j}^{n+1} = \mathbf{U}_{i,j}^n - \frac{\Delta t}{\Delta x} [\mathbf{F}_{i+1/2,j} - \mathbf{F}_{i-1/2,j}] - \frac{\Delta t}{\Delta y} [\mathbf{G}_{i,j+1/2} - \mathbf{G}_{i,j-1/2}] \quad (2)$$

where $\mathbf{U}_{i,j}^n$ is the integral average of \mathbf{U} in cell (i, j) at time level n and \mathbf{F} and \mathbf{G} are intercell fluxes. WAF-type unsplit approach of Billett and Toro (Billett and Toro, 1997a) can be described in the following operator form:

$$\mathbf{F}_{i+1/2,j} = \mathbf{L}_{x,\Delta t/2}^{WAF} (\mathbf{L}_{y,\Delta t/2}^{GOD}(\mathbf{U}_{i,j}), \mathbf{L}_{y,\Delta t/2}^{GOD}(\mathbf{U}_{i+1,j})) \quad (3)$$

$$\mathbf{G}_{i,j+1/2} = \mathbf{L}_{y,\Delta t/2}^{WAF} (\mathbf{L}_{x,\Delta t/2}^{GOD}(\mathbf{U}_{i,j}), \mathbf{L}_{x,\Delta t/2}^{GOD}(\mathbf{U}_{i,j+1})) \quad (4)$$

Here $L_{y,\Delta t/2}^{GOD}$ means that the solution is evolved by half time step along the y direction and the needed flux is calculated by the Godunov first order scheme, and $L_{x,\Delta t/2}^{WAF}$ means that the flux along the x -direction is obtained by the WAF-type method making use of the predictor value. Likewise for $L_{x,\Delta t/2}^{GOD}$ and $L_{y,\Delta t/2}^{WAF}$.

3. Centred Schemes for 2D Hyperbolic Conservation Laws

In upwind-type schemes, the computation of the numerical fluxes in (2) requires us to identify *the direction of the wind*, which in some cases is complicated or impossible. Instead, in this section we propose a high-resolution unsplit finite volume centred scheme to solve the hyperbolic system (1). The scheme is analogous to (3)-(4), namely

$$\begin{aligned} \mathbf{F}_{i+\frac{1}{2},j}^n &= \mathbf{L}_{x,\Delta t/2}^{SLIC} \left(\mathbf{L}_{y,\Delta t/2}^{FORCE}(\mathbf{U}_{i-1,j}^n), \mathbf{L}_{y,\Delta t/2}^{FORCE}(\mathbf{U}_{i,j}^n), \right. \\ &\quad \left. \mathbf{L}_{y,\Delta t/2}^{FORCE}(\mathbf{U}_{i+1,j}^n), \mathbf{L}_{y,\Delta t/2}^{FORCE}(\mathbf{U}_{i+2,j}^n) \right) \end{aligned} \quad (5)$$

$$\begin{aligned} \mathbf{G}_{i,j+\frac{1}{2}}^n &= \mathbf{L}_{y,\Delta t/2}^{SLIC} \left(\mathbf{L}_{x,\Delta t/2}^{FORCE}(\mathbf{U}_{i,j-1}^n), \mathbf{L}_{x,\Delta t/2}^{FORCE}(\mathbf{U}_{i,j}^n), \right. \\ &\quad \left. \mathbf{L}_{x,\Delta t/2}^{FORCE}(\mathbf{U}_{i,j+1}^n), \mathbf{L}_{x,\Delta t/2}^{FORCE}(\mathbf{U}_{i,j+2}^n) \right) \end{aligned} \quad (6)$$

where *FORCE* represents the FORCE scheme (Toro, 1999a) with flux

$$\mathbf{F}_{i+\frac{1}{2}}^{FORCE} = \frac{1}{2} \left[\mathbf{F}_{i+\frac{1}{2}}^{LF}(\mathbf{U}_L, \mathbf{U}_R) + \mathbf{F}_{i+\frac{1}{2}}^{RI}(\mathbf{U}_L, \mathbf{U}_R) \right] \quad (7)$$

in which LF denotes the Lax-Friedrichs flux and RI denotes Richtmyer or two step Lax-Wendroff flux (Toro, 1999a); $SLIC$ represents the SLIC scheme which is a MUSCL-based second order extension of the FORCE scheme, the details of which can be found in (Toro and Billett, 2000).

In equation (5), $L_{y,\Delta t/2}^{FORCE}$ means that the solution is evolved by $\frac{1}{2}\Delta t$ along the y -direction with the flux being calculated by the $FORCE$ scheme, and $L_{x,\Delta t/2}^{SLIC}$ means that the flux along the x -direction is calculated by the $SLIC$ scheme. Note that the $SLIC$ operator requires four arguments, each of which is a predicted value by the application of the $FORCE$ operator in the transverse direction. Likewise for $L_{x,\Delta t/2}^{FORCE}$ and $L_{y,\Delta t/2}^{SLIC}$ in equation (6).

4. Accuracy and Stability

Accuracy and stability are two factors to be considered and analysed when a new numerical scheme is developed. Here we prove that centred unsplit finite volume schemes of first- and second-order of accuracy in both space and time can be constructed via (5)-(6). We also show that the linear stability is optimal, the Courant number c in each direction satisfies $|c| \leq 1$.

4.1. A FIRST ORDER CENTRED UNSPLIT FINITE VOLUME SCHEME

Consider the two dimensional linear advection equation

$$u_t + (au)_x + (bu)_y = 0 \quad (8)$$

with positive constant wave speeds a and b . The fluxes $f = au$ and $g = bu$ are calculated according to (5) and (6) with the SLIC scheme being substituted with the FORCE scheme. A scheme of 9-point support is obtained. See Fig. 1, which includes the *upwind corner point* $u_{i-1,j-1}$. Recall that the straight finite volume extension of Godunov's method does not include this crucial point. Conventional truncation error analysis shows that the scheme is first order accurate in both space and time. In order to check the stability, a simple method (Billett and Toro, 1997b) is used in this paper. The stability plots of the first-order centred scheme is depicted on the upper left of Fig. 2, from which we can see that the stability range satisfies,

$$|c_x| \leq 1; \quad |c_y| \leq 1, \quad (9)$$

where c_x and c_y are the directional Courant numbers.

4.2. A SECOND-ORDER CENTRED UNSPLIT FINITE VOLUME SCHEME

It is well known that first-order schemes give very diffusive numerical results, which make them less attractive than higher-order schemes as far as practical engineering problems are concerned. Application of (5), (6) produces a scheme with a stencil of 21 points, see Fig. 1. The stencil includes the complete support of the first order scheme.

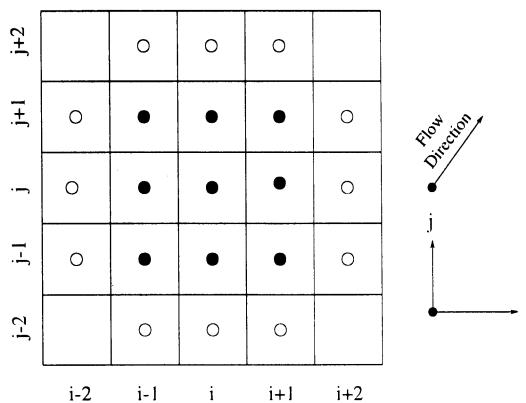


Figure 1. Stencil for the second order centred scheme. Full circles denote the points used in the first order centred scheme; empty circles denote the added points in order to increase the accuracy to second order.

The derived scheme is second-order accurate in space and time for any ω_x and ω_y satisfying

$$\omega_x \in [-1, 1], \quad \omega_y \in [-1, 1], \quad (10)$$

where ω_x and ω_y are parameters in the MUSCL data reconstruction stage in the SLIC scheme.

Stability plots of the first- and second-order centred schemes are shown in Fig. 2, which suggest that the stability range for the centred unsplit schemes is the same as for the first-order scheme, namely $|c_x| \leq 1$; $|c_y| \leq 1$. We remark that the conventional unsplit version of Godunov's method has half the stability range of our scheme. The same applies to the MUSCL-Hancock unsplit scheme.

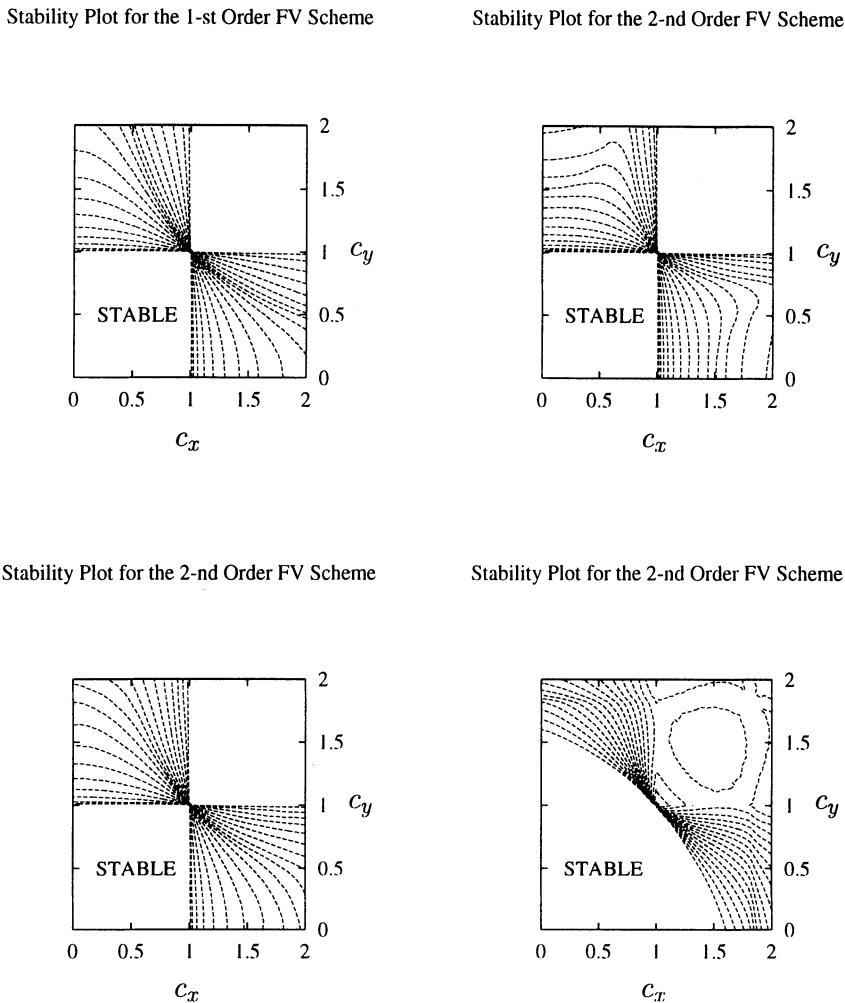


Figure 2. Stability Plots for the 1st and 2nd Order FV Centred Schemes. Upper left: the 1st order scheme. For the 2nd order scheme: the upper-right plot has $\omega_x = \omega_y = -1$; the bottom-left plot has $\omega_x = \omega_y = 0$; the bottom-right plot has $\omega_x = \omega_y = 1$.

5. Application to the 2D Shallow Water Equations

In this section the scheme is applied to a nonlinear system, namely the two dimensional shallow water equations. The same test problem as suggested by Toro (Toro, 1999b) is used. A right travelling bore of Froude number $Frs = 2.0$ starting at $x = 0.25m$ travels down a channel of length 1m and

interacts with a vertical wall inclined at an angle 25° to the direction of flow at $x = 0.3m$. The initial conditions are

$$\begin{cases} h_L = 0.059307033m \\ u_L = 0.57294306m/s \\ v_L = 0.0m/s \end{cases}; \quad \begin{cases} h_R = 0.025m \\ u_R = 0.0m/s \\ v_R = 0.0m/s \end{cases} \quad (11)$$

Here subscript L denotes the post-shock value and R denotes initial values ahead of the bore. In the computations the Courant number coefficient was set at 0.9 and a SUPERBEE-type slope limiter function (Toro, 1999a) was used. Computations were carried out on a mesh of 500×250 cells.

A contour plot of the height of water h is shown in Fig. 3, which depicts the occurrence of Mach reflection. The numerical results are satisfactory

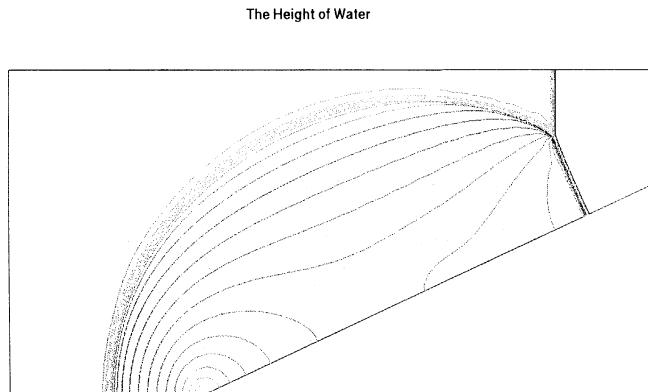


Figure 3. the Contour Plot of Mach Reflection of Bore in Shallow Water.

and reflect the main physical features of Mach reflection (Ben-Dor, 1992). On the whole, our results compare well with those obtained with WAF-type Godunov schemes (Billett and Toro, 1997a).

6. Conclusions

New centred unsplit first-order and second-order finite volume schemes for hyperbolic conservation laws have been presented. Analysis of the schemes

applied to the two dimensional linear advection equation shows that the schemes have an improved stability range over established methods such as the Lax-Wendroff scheme, the Godunov scheme, and the MUSCL-Hancock scheme. The second-order centred scheme has been extended to the two dimensional shallow water equations in curvilinear coordinates and the numerical results are almost as good as those obtained by upwind-based schemes but with increased simplicity and efficiency. On the whole, the numerical experiment shows that the derived centred finite volume scheme is robust, efficient and of high-resolution.

The great attraction of the newly derived centred schemes is their simplicity and efficiency. For engineering problems which are too complicated to be solved by upwind-based schemes, our centred scheme can be a first choice. For some scientists who do not want to be involved with the intricacies of Riemann solvers, our centred scheme can be an alternative.

References

- Toro E F(1999a). Riemann Solvers and Numerical Methods for Fluid Dynamics. Second Edition. Springer-Verlag, New York.
- Billett S J and Toro E F (1997a). On WAF-Type Scheme for Multidimensional Hyperbolic Conservation Laws. *Journal of Computational Physics* **130**: 1.
- Toro E F and Billett S J (2000). Centred TVD Scheme for Hyperbolic Conservation Laws. *IMA J. Numerical Analysis* 47.
- Nessyahu H and Tadmor E (1990). Non-oscillatory Central Differencing for Hyperbolic Conservation Laws. *Journal of Computational Physics* **87**:408.
- Billett S J and Toro E F (1997b). On the Accuracy and Stability of Explicit Schemes for Multidimensional Linear Homogeneous Advection Equations. *Journal of Computational Physics* **131**:247
- Ben-Dor G (1992). Shock Wave Reflection Phenomena. Springer-Verlag.
- Toro E F and Olim M and Takayama K(1999b). Unusual Increase in Tsunami Wave Amplitude at the Okushiri Island: Mach Reflection of Shallow Water Waves. *Proceedings of the 22nd International Symposium on Shock Waves*. University of Southampton, UK.
- Leveque R J (1992). Numerical Methods for Conservation Laws. Birkhauser Basel.
- Hirsch C (1988). Numerical Computation of Internal and External Flows. Wiley, New York.

TOWARDS VERY HIGH ORDER GODUNOV SCHEMES

E. F. TORO, R. C. MILLINGTON AND L. A. M. NEJAD

*Department of Computing and Mathematics,
Manchester Metropolitan University,
Chester Street, Manchester, M1 5GD, U.K.
Emails: E.F.Toro@doc.mmu.ac.uk*

SUMMARY. We present an approach, called ADER, for constructing non-oscillatory advection schemes of very high order of accuracy in space and time; the schemes are explicit, one step and have optimal stability condition for one and multiple space dimensions. The approach relies on essentially non-oscillatory reconstructions of the data and the solution of a generalised Riemann problem via solutions of derivative Riemann problems. The schemes may thus be viewed as Godunov methods of very high order of accuracy. We present the ADER formulation for the linear advection equation with constant coefficients, in one and multiple space dimensions. Some preliminary ideas for extending the approach to non-linear problems are also discussed. Numerical results for one and two-dimensional problems using schemes of upto 10-th order accuracy are presented.

1. Introduction

We are concerned with the construction of non-oscillatory numerical schemes of very high-order of accuracy for hyperbolic conservation laws. Here we present a new approach, called ADER, for constructing such schemes. The ADER formulation is presented for the linear advection equation with constant coefficients in one, two and three space dimensions; schemes for linear hyperbolic systems with constant coefficients in one and multiple space dimensions can then be easily constructed. As is well-known from Godunov's

theorem (Godunov, 1959) high accuracy and absence of spurious oscillations near discontinuities are contradictory requirements on numerical methods; the theorem says that any (linear) scheme of accuracy greater than *one* will be oscillatory near discontinuities. A classical way of circumventing Godunov's theorem is to construct non-linear schemes, even when applied to linear problems. A successful class of non-linear schemes are the so-called total Variation Diminishing Methods, or TVD methods, (Harten, 1983), (Sweby, 1984) developed over the last two decades or so. Such schemes provide today the basis of a mature numerical technology suitable for industrial and scientific applications; for upto date presentations of these methods see for example (LeVeque, 1992), (Toro, 1997), (Toro, 1999), (Kröner, 1997), (Godlewski and Raviart, 1996) and (Hirsch, 1988). These methods are second-order accurate almost everywhere, reduce locally to first-order of accuracy and are known to be unsuitable for special application areas such compressible turbulence and for wave propagation problems involving long-time evolution; extrema are clipped and numerical dissipation may become dominant. The aim of this paper is to present an approach capable of producing schemes of uniform very high accuracy for smooth solutions and non-oscillatory near discontinuities, without imposing TVD-like constraints.

There are at present several approaches for constructing high-order methods that are applicable to hyperbolic conservations laws. Examples include spectral methods (Canuto and Quarteroni, 1990), discontinuous Galerkin finite element methods (Cockburn and Shu, 1998), (van der Vegt, 2001), the class of compact difference schemes of Tolstykh and co-workers (Tolstykh, 1994) and the UNO and ENO/WENO schemes of Harten and co-workers (Harten and Osher, 1987), (Harten *et al.*, 1987), (Shu, 1997). The requirement of high-order of accuracy is met by all of the above approaches, although there may be problems in preserving time accuracy, particularly for multiple space dimensions and problems involving source terms. The ENO schemes enjoy an extra advantage; they fulfil the requirement of absence of spurious oscillations near discontinuities in a way that is described as *essentially non-oscillatory*. This means that solutions to model problems are, *to the eye*, free from spurious oscillations and that such oscillations can be shown to decay as the mesh is refined. We note that this

is not a property enjoyed by conventional, linear schemes.

In this paper we present a new approach for constructing very high-order non-oscillatory schemes for hyperbolic conservation laws. The approach is related to the ENO/WENO methods and is so far applicable to linear hyperbolic systems with constant coefficients in one and multiple space dimensions. The schemes are conservative, one-step, explicit and fully discrete, requiring only the computation of the inter-cell fluxes to advance the solution by a full time step. For a one-dimensional problem for example, to compute the inter-cell numerical flux we solve a sequence of m Riemann problems for the k -th spatial derivatives of the solution, with $k = 0, 1, \dots, m - 1$, where m is arbitrary and is the order of accuracy of the resulting scheme. The case $m = 1$ reproduces the basic Godunov first-order upwind method. The non-linear (non-oscillatory) version of the schemes results from applying ENO polynomial reconstructions, in much the same way as in the ENO schemes.

The rest of this paper is organised as follows. In section 2 we review the GRP and Modified GRP schemes, whose philosophy serves as the inspiration for the ADER approach; in section 3 we present the basis of the ADER approach for the one-dimensional case. In section 4 we present ADER for two and three-dimensional linear problems; in section 5 we discuss some preliminary ideas for extending ADER to non-linear problems. In section 6 we present some numerical experiments for one and two-dimensional examples. Section 7 summarises the paper and discusses possible future developments.

2. Review of the GRP Scheme

The Generalised Riemann problem (GRP) scheme of Ben-Artzi and Falcovitz (Ben-Artzi and Falcovitz, 1984) and a modification to it proposed by Toro in 1992 (unpublished) serve as the inspiration for the construction of the ADER schemes presented in this paper. We therefore review this approach and the appropriate modifications that are necessary for the construction of very high order schemes.

2.1. BACKGROUND

Consider the scalar conservation law

$$\partial_t q + \partial_x f(q) = 0 , \quad (1)$$

where $q(x, t)$ is the conserved variable and $f(q)$ is the physical flux function. The integral form of (1), which admits discontinuous solutions, is

$$\oint (q dx - f(q) dt) = 0 , \quad (2)$$

which under suitable smoothness assumptions reproduces the differential form (1) of the conservation law. Consider now a control volume in $x-t$ space of dimensions $\Delta x \times \Delta t$. Evaluation of the integral form (2) in this control volume produces the conservative formula

$$q_i^{n+1} = q_i^n - \frac{\Delta t}{\Delta x} \left[f_{i+\frac{1}{2}} - f_{i-\frac{1}{2}} \right] , \quad (3)$$

where q_i^n is the integral average

$$q_i^n = \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} q(x, t^n) dx \quad (4)$$

at time $t = t^n$ in the *volume*

$$I_i = \left[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}} \right] \quad (5)$$

of width

$$\Delta x = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}} . \quad (6)$$

The *fluxes* in (3) are interpreted as time averages of the physical flux, namely

$$f_{i+\frac{1}{2}} = \frac{1}{\Delta t} \int_0^{\Delta t} f(q(x_{i+\frac{1}{2}}, t)) dt . \quad (7)$$

Conservative numerical methods for (1) are based on (3), whereby an expression for the *numerical flux* $f_{i+\frac{1}{2}}$ is provided. Godunov (Godunov, 1959) introduced the idea of computing the numerical flux in (3) by evaluating (7) in terms of the solution $q^*(x/t)$ of a local initial value problem (IVP)

called the Riemann problem

$$\left. \begin{aligned} \partial_t q + \partial_x f(q) &= 0 \\ q(x, 0) &= \begin{cases} q_i^n & \text{if } x < 0 \\ q_{i+1}^n & \text{if } x > 0 \end{cases} \end{aligned} \right\} \quad (8)$$

evaluated at the *cell interface* $x_{i+\frac{1}{2}}$. Conventionally, the initial data $q(x, 0)$ in the Riemann problem is *piece-wise constant* and consists of two constant states of the form (4) separated by a discontinuity. Such distribution of the initial data is traditionally associated with a first-order accurate Godunov-type scheme with numerical flux

$$f_{i+\frac{1}{2}} = f(q^*(0)) \quad (9)$$

obtained by evaluating (7) with integrand $f(q^*(x/t))$. Second-order Godunov-type schemes can be constructed in essentially two ways. One approach is provided by the the Weighted Average Flux Method of Toro (Toro, 1989), (Billett and Toro, 1997) whereby the same *piece-wise constant* data Riemann problem of the first-order Godunov method is solved but this time its solution is utilised more fully by taking a space/time integral of it in an appropriate control volume. Another approach for constructing second-order Godunov methods relies on piece-wise linear reconstruction of the data and the solution of a so called generalised Riemann problem at each inter-cell position; examples of schemes in this class are the piece-wise linear method (PLM) of Colella (Colella, 1985) and the MUSCL-Hancock scheme (van Leer, 1985). In this case the Riemann problem reads

$$\left. \begin{aligned} \partial_t q + \partial_x f(q) &= 0 \\ q(x, 0) &= \begin{cases} q_i(x) & \text{if } x < 0 \\ q_{i+1}(x) & \text{if } x > 0 \end{cases} \end{aligned} \right\} \quad (10)$$

where $q_i(x)$ and $q_{i+1}(x)$ are *linear* functions in x . IVP (10) is called the *Generalised Riemann Problem*, or GRP. In both the Colella and MUSCL-Hancock schemes one resorts to simplifications of the solution of this GRP, in which the so called boundary extrapolated values

$$q_L = \lim_{x \rightarrow x_{i+\frac{1}{2}}^-} q_i(x), \quad q_R = \lim_{x \rightarrow x_{i+\frac{1}{2}}^+} q_{i+1}(x) \quad (11)$$

play a crucial role. In the PLM scheme for instance, these values constitute the initial condition for a piece-wise constant data Riemann problem, such as (8) with initial data (11), to provide one term of the flux, with the rest of the information computed from tracing characteristics back to the reconstructed data. We note here that using only the first term for the flux leads to an unstable scheme. In the MUSCL-Hancock scheme the boundary extrapolated values (11) are evolved by half the time step before solving a conventional piece-wise constant data Riemann problem such as (8) with initial data given by evolved boundary extrapolated values. See Toro (Toro, 1998), (Toro, 1999) for details.

2.2. CONVENTIONAL AND MODIFIED GRP

Another approach for constructing second-order Godunov-type methods is the GRP approach of Ben-Artzi and Falcovitz (Ben-Artzi and Falcovitz, 1984), in which the generalised Riemann problem (10) is solved more directly than in the PLM and MUSCL-Hancock schemes. Assuming the reconstruction in (10) is linear, Ben-Artzi and Falcovitz produce an expression for the solution of the generalised Riemann problem at the interface $x = x_{i+\frac{1}{2}}$ and time $t = \frac{1}{2}\Delta t$; they do so by use of a time Taylor series expansion at $x = x_{i+\frac{1}{2}}$ around $t = 0$, namely

$$q_{i+\frac{1}{2}}^{n+\frac{1}{2}} = q_{i+\frac{1}{2}}^{(0)} + \frac{1}{2} \Delta t (\partial_t q)_{i+\frac{1}{2}} + O(\Delta t^2), \quad (12)$$

where the term $q_{i+\frac{1}{2}}^{(0)}$ at time $t = 0$ accounts for the very first effect of the interaction of the two piece-wise linear states in (10). Such first interaction is solely determined by the boundary extrapolated values q_L and q_R in (11). Thus the first term $q_{i+\frac{1}{2}}^{(0)}$ in the Taylor series expansion (12) is obtained by solving the conventional Riemann problem (8) with initial data q_L and q_R given by (11); this part is common to the PLM and MUSCL-Hancock schemes. Computing the second term in (12) in the conventional GRP method can be labourious. Ben-Artzi and Falcovitz provided an expression for this term for the non-linear system of compressible Euler equations, thus obtaining a very robust second order Godunov scheme.

A simplification to the GRP scheme of Ben-Artzi and Falcovitz that is central to the development of ADER was suggested by Toro in the early 90's, whereby the computation of the second, difficult, term is replaced by that of solving a *linear* Riemann problem for the *gradient* of the solution, as follows. First we re-write the conservation law (1) in quasi-linear form

$$\partial_t q + \lambda(q) \partial_x q = 0 , \quad (13)$$

where

$$\lambda(q) = \frac{df}{dq} \quad (14)$$

is the characteristic speed; for systems this is the Jacobian matrix. Then, by use of (13) we replace the time derivative in (12) by a space derivative so that the expansion (12), to second order, reads

$$q_{i+\frac{1}{2}}^{n+\frac{1}{2}} = q_{i+\frac{1}{2}}^{(0)} - \frac{1}{2} \Delta t (\lambda(q))_{i+\frac{1}{2}} (\partial_x q)_{i+\frac{1}{2}} . \quad (15)$$

The next problem is the computation of $(\lambda(q))_{i+\frac{1}{2}} (\partial_x q)_{i+\frac{1}{2}}$; one possibility is the following. First linearise (13) around the state $q_{i+\frac{1}{2}}^{(0)}$ in (12), which is the leading term in the expansion and results from solving a conventional non-linear Riemann problem. The linearised problem reads

$$\partial_t q + \hat{\lambda} \partial_x q = 0 , \quad (16)$$

where $\hat{\lambda} = \lambda(q_{i+\frac{1}{2}}^{(0)})$. Then, it is easily seen that $v \equiv \partial_x q$ obeys the linearised evolution equation (16) identically. Moreover, the following more general result holds:

THEOREM 1: Let $v \equiv \partial_x^{(k)} q$ be the k -th order spatial derivative of $q(x, t)$. Then v obeys the linearised evolution equation (16) exactly.

The proof is trivial and is thus omitted. We note that the same result is also valid for all time derivatives $v \equiv \partial_t^{(k)} q$.

By virtue of the above theorem we can pose the following Riemann problem for the *gradient* of the solution, namely

$$v(x, 0) = \left\{ \begin{array}{ll} q_L^{(1)} \equiv \partial_x^{(1)} q_i(x) & \text{if } x < 0 \\ q_R^{(1)} \equiv \partial_x^{(1)} q_{i+1}(x) & \text{if } x > 0 \end{array} \right\} \quad (17)$$

$$\partial_t v + \hat{\lambda} \partial_x v = 0$$

the solution of which is

$$q_{i+\frac{1}{2}}^{(1)}(x/t) = \begin{cases} q_L^{(1)} & \text{if } x - \hat{\lambda}t < 0 \\ q_R^{(1)} & \text{if } x - \hat{\lambda}t > 0 \end{cases} \quad (18)$$

Finally, the inter-cell flux for the Modified GRP scheme (MGRP) is given by

$$f_{i+\frac{1}{2}}^{(mgrp)} = f\left(q_{i+\frac{1}{2}}^{(mgrp)}\right), \quad q_{i+\frac{1}{2}}^{(mgrp)} = q_{i+\frac{1}{2}}^{(0)} - \frac{1}{2}\Delta t \hat{\lambda} q_{i+\frac{1}{2}}^{(1)}, \quad (19)$$

where $q_{i+\frac{1}{2}}^{(0)}$ is the solution of the conventional Riemann problem (8) with initial data (11) and $q_{i+\frac{1}{2}}^{(1)}$ is the solution (18) of the *first-derivative* Riemann problem (17), evaluated at $x/t = 0$, the inter-cell position.

The above Modified GRP scheme was extended by MSc student Cheney (Cheney, 1994) to solve advection-diffusion problems such as

$$\partial_t q + \partial_x f(q) = \alpha \partial_x^{(2)} q, \quad (20)$$

where the viscous flux $-\alpha \partial_x^{(1)} q$ was evaluated using the solution (18) of the gradient Riemann problem (17). A benefit of so doing is an enlarged stability region, in comparison to simple central differencing for the viscous term. The Modified GRP scheme was also extended by Boden (1993, unpublished) to solve the non-linear two-dimensional shallow water equations. Full details of Modified GRP second-order schemes for non-linear systems are presented in (Toro, 1998). MSc student Cáceres (Cáceres, 1993) attempted an extension of the modified GRP scheme to third order of accuracy. It was thought at the time that it was a simple matter of using quadratic reconstructions of the data in the generalised Riemann problem and retention of three terms in the Taylor series expansion (12). The derived scheme by so doing is only $O(\Delta x^3, \Delta t^2)$ and the linearised stability condition is not the *optimal condition* $|c| \leq 1$ but

$$|c| \leq \frac{1}{3}(-1 + \sqrt{13}) \approx 0.87, \quad (21)$$

where c is the Courant number $c = \lambda \Delta t / \Delta x$. Given the loss of time accuracy and the reduction in the stability range for the attempted *third* order scheme, it became clear that a different expression for the inter-cell

flux expansion was required, if successful very high order schemes could be constructed, still using the MGRP idea.

3. ADER for One-Dimensional Problems

The ADER approach is a successful attempt to exploit the Modified GRP scheme to construct schemes of very high order of accuracy and has the following main ingredients. We first produce high-order reconstructions of the initial data and pose a generalised Riemann problem in which the initial data is piece-wise smooth, possibly with a discontinuity at the inter-cell edge. The solution of this generalised Riemann problem is then Taylor expanded in time, at the interface position, to any order of accuracy. We then obtain an average ADER state by time-integrating the Taylor series expansion. A crucial step then follows, which consists of replacing time derivatives by space derivatives, as done for the Modified GRP scheme presented in the previous section; for the linear case this is a trivial step. The main task left is the evaluation of all the spatial derivatives in the expansion; this is achieved by an evolution step based on Theorem 1. Evolution equations for all spatial derivatives are first identified; for the linear case these happen to be homogeneous linear advection equations. To complete the evolution process we pose and solve Riemann problems for *derivatives*, and thus all terms in the expansion for the time average ADER state are determined. This state is then used to compute the ADER numerical flux. We now discuss details of the above ingredients.

3.1. RECONSTRUCTION AND GENERALISED RIEMANN PROBLEM

The data in the form of cell averages is reconstructed via high-order polynomials. Use of a fixed stencil for the reconstruction in the ADER approach leads to *linear* ADER schemes. This means that the schemes have constant coefficients when applied to a linear equation or to a linear system with constant coefficients. In accordance with Godunov's theorem, these *linear schemes* of accuracy greater than one will produce spurious oscillations in the vicinity of discontinuities. To resolve this difficulty we use the ENO (essentially non-oscillatory) interpolation procedure, first presented by Harten

and collaborators (Harten and Osher, 1987), (Harten *et al.*, 1987), (Shu, 1997) in the context of ENO schemes. This interpolation theory has by now reached a stage of maturity and it is proving useful in various areas of application; useful background on the ENO interpolation theory is found in (Laney, 1998). In the ENO procedure the stencil of the interpolation is *adaptive* and leads to *non-linear* ADER schemes; the schemes are non-linear even when applied to linear problems. Fig. 1 depicts higher-order polynomial reconstructions $q_i(x)$ and $q_{i+1}(x)$ of the data for cells I_i and I_{i+1} respectively.

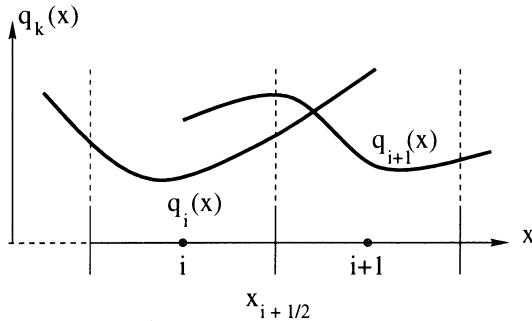


Figure 1. Illustration of high-order polynomial reconstructions $q_i(x)$ and $q_{i+1}(x)$ in cells I_i and I_{i+1} . Functions $q_i(x)$ and $q_{i+1}(x)$ form initial condition for a generalised Riemann problem at $x = x_{i+\frac{1}{2}}$.

In order to compute the ADER numerical flux at the inter-cell position $x_{i+\frac{1}{2}}$ for a scheme of m -th order of accuracy in space and time we solve the generalised Riemann problem

$$\left. \begin{aligned} \partial_t q + \partial_x f(q) &= 0 \\ q(x, 0) &= \begin{cases} q_i(x) & \text{if } x < 0 \\ q_{i+1}(x) & \text{if } x > 0 \end{cases} \end{aligned} \right\} \quad (22)$$

where the initial data are the $(m - 1)$ -th order polynomial functions $q_i(x)$ and $q_{i+1}(x)$, as depicted in Fig. 1. We want to find an expression for the numerical flux, or for a state, at the inter-cell position; this will involve some kind of series expansion. Next we study three possible approaches for finding such expansion.

3.2. STATE SERIES EXPANSION: APPROACH I

Consider the general initial value problem for the linear advection equation

$$\left. \begin{array}{l} PDE : \quad \partial_t q + \lambda \partial_x q = 0 \\ IC : \quad q(x, 0) \equiv q_0(x) \end{array} \right\} \quad (23)$$

where the characteristic speed λ is constant and the initial condition $q_0(x)$ is an arbitrary function of distance x .

THEOREM 2: For a sufficiently smooth function $q_0(x)$ in IVP (23), one may express the solution at a time $t = \Delta t$ as given by the *finite volume* type formula

$$\bar{q}(x, \Delta t) = \bar{q}(x, 0) + \frac{\Delta t}{\Delta x} \left[\lambda \tilde{q}(x - \frac{1}{2}\Delta x) - \lambda \tilde{q}(x + \frac{1}{2}\Delta x) \right], \quad (24)$$

where

$$\bar{q}(x, t) = \frac{1}{\Delta x} \int_{x - \frac{1}{2}\Delta x}^{x + \frac{1}{2}\Delta x} q(x, t) dx \quad (25)$$

is a *cell average* in the interval $[x - \frac{1}{2}\Delta x, x + \frac{1}{2}\Delta x]$,

$$\tilde{q}(x) = q_0^{(0)}(x) + \sum_{k=1}^{\infty} \frac{(-\lambda \Delta t)^k}{(k+1)!} q_0^{(k)}(x) \quad (26)$$

and $q_0^{(k)}(x)$ is the k -th order spatial derivative of $q_0(x)$, namely

$$q_0^{(k)}(x) \equiv \partial_x^{(k)} q_0(x). \quad (27)$$

Before proving the above statement, the following remarks are in order. First we note that no discretisation is involved here, the statements are exact. Expression (24) looks like a finite volume conservative scheme, of infinite accuracy, in which $\bar{q}(x, \Delta t)$ is an integral average of the solution over the interval $[x - \frac{1}{2}\Delta x, x + \frac{1}{2}\Delta x]$, just as in finite volume numerical methods; the sought state is (26) and the *numerical flux* will have the form $\tilde{f} \equiv \lambda \tilde{q}$. Manipulation of the exact solution has given us the correct expansion (26) for the sought state. We also note that (26) coincides with a Taylor expansion at $t = \frac{1}{2}\Delta t$ only for the first two terms.

Proof: The analytical solution of IVP (23) at a time Δt is given by

$$q(x, \Delta t) = q_0(x - \lambda \Delta t) . \quad (28)$$

Now assume that $q_0(x)$ is sufficiently smooth so that all its spatial derivatives are defined, then a Taylor series expansion of (28) reads

$$q(x, \Delta t) = q_0(x) + \sum_{k=1}^{\infty} \frac{(-\lambda \Delta t)^k}{k!} q_0^{(k)}(x) , \quad (29)$$

with $q_0^{(k)}(x)$ as given by (27). The infinite summation in (29) is now regarded as a function of x , for a fixed time step Δt , and may be expressed as

$$\sum_{k=1}^{\infty} \frac{(-\lambda \Delta t)^k}{k!} q_0^{(k)}(x) = -\lambda \Delta t \frac{d}{dx} \tilde{q}(x) , \quad (30)$$

where

$$\tilde{q}(x) = q_0^{(0)}(x) + \sum_{k=1}^{\infty} \frac{(-\lambda \Delta t)^k}{(k+1)!} q_0^{(k)}(x) . \quad (31)$$

Use of (30) and (31) in (29) gives

$$q(x, \Delta t) = q(x, 0) - \lambda \Delta t \frac{d}{dx} \tilde{q}(x) . \quad (32)$$

We now integrate (32) with respect to x in the interval $[x - \frac{1}{2} \Delta x, x + \frac{1}{2} \Delta x]$ and divide the result by Δx to produce (24), with definitions (25)-(27), and the theorem is thus proved.

3.3. STATE SERIES EXPANSION: APPROACH II

A clearer and entirely equivalent procedure to obtain an expansion for an inter-cell state and thus the numerical flux is as follows. We first express the solution of the generalised Riemann problem (22) at time $t = \tau$ as a time Taylor series expansion at the cell inter-cell position $x = x_{i+\frac{1}{2}}$ around the initial time $t = 0$ (in local coordinates)

$$q(x_{i+\frac{1}{2}}, \tau) = q(x_{i+\frac{1}{2}}, 0) + \sum_{k=1}^{\infty} \frac{\tau^k}{k!} \partial_t^{(k)} q(x_{i+\frac{1}{2}}, 0) . \quad (33)$$

We then obtain a time-average state $q_{i+\frac{1}{2}}^{ader}$ by time-integrating (33) in the interval $[0, \Delta t]$ as

$$q_{i+\frac{1}{2}}^{ader} = \frac{1}{\Delta t} \int_0^{\Delta t} q(x_{i+\frac{1}{2}}, t) dt. \quad (34)$$

and use the Lax and Wendroff procedure (Lax and Wendroff, 1960), (Shu, 1997) to substitute all time-derivatives by space-derivatives using the differential equation; for the one-dimensional linear advection equation with constant speed the time derivatives are

$$\partial_t^{(k)} q = (-\lambda)^k \partial_x^{(k)} q. \quad (35)$$

By retaining a finite number m of terms in the above procedure one obtains the average state expansion

$$q_{i+\frac{1}{2}}^{ader} = q_{i+\frac{1}{2}}^{(0)} + \sum_{k=1}^{m-1} \frac{(-\lambda \Delta t)^k}{(k+1)!} q_{i+\frac{1}{2}}^{(k)}. \quad (36)$$

Then, the ADER numerical flux follows as

$$f_{i+\frac{1}{2}}^{ader} = f(q_{i+\frac{1}{2}}^{ader}), \quad (37)$$

which if substituted in the finite volume one-step conservative formula (3) leads to a scheme of m -th order accuracy in space and time for the linear advection equation with constant coefficient.

3.4. DIRECT FLUX EXPANSION

Approach II above can be applied to derive the ADER flux more directly; this is possible by using a remarkable property of the flux (Ghidaglia *et al*, 1999) of a conservation law such as (1), which we now state.

THEOREM 3: For any conservation law (1) the flux function $f(q)$ obeys the evolution equation

$$\partial_t f(q) + \lambda(q) \partial_x f(q) = 0, \quad (38)$$

where $\lambda(q)$ is the characteristic speed.

Proof: The proof is straightforward. Multiply (1) by $\lambda(q)$ and use the chain rule to obtain $\partial_t f(q) = \lambda(q) \partial_t q$ and the result follows.

The above property is also valid for non-linear systems, as is trivially seen by replacing the characteristic speed by the Jacobian matrix. For the linear scalar case, application of above property (38) leads to a direct expression for the ADER flux, namely

$$f_{i+\frac{1}{2}}^{ader} = f_{i+\frac{1}{2}}^{(0)} + \sum_{k=1}^{m-1} \frac{(-\lambda \Delta t)^k}{(k+1)!} f_{i+\frac{1}{2}}^{(k)}. \quad (39)$$

In this expression one could in principle use approximate Riemann solvers that give direct expressions for the flux rather than for a state. We remark here that we have performed numerical experiments in which the Godunov flux based on the Riemann problem solution has been replaced by a centred flux, such as the FORCE flux (Toro, 1996), (Toro and Billett, 2000). The results from using upwind and centred fluxes in the ADER approach look in general comparable and for very high orders of accuracy they are indistinguishable. However these observations are only provisional.

3.5. EVOLUTION VIA DERIVATIVE RIEMANN PROBLEMS

We now have an expression for the sought average state to be utilised in the evaluation of a numerical flux. We utilise the analytical developments of approaches I and II above in a numerical set up. Assume polynomial reconstructions $q_i(x)$ and $q_{i+1}(x)$ of the initial data in cells I_i and I_{i+1} respectively and formulate the generalised Riemann problem as in (22). The ADER state generated by this GRP is denoted by $q_{i+\frac{1}{2}}^{ader}$, with corresponding numerical flux

$$f_{i+\frac{1}{2}}^{ader} = f(q_{i+\frac{1}{2}}^{ader}); \quad q_{i+\frac{1}{2}}^{ader} = q_{i+\frac{1}{2}}^{(0)} + \sum_{k=1}^{m-1} \frac{(-\lambda \Delta t)^k}{(k+1)!} q_{i+\frac{1}{2}}^{(k)}. \quad (40)$$

Here $r_{i+\frac{1}{2}}^{(k)}$ is the solution $r_{i+\frac{1}{2}}^{(k)}(x, t)$ of the k -th order derivative Riemann problem evaluated at $x/t = 0$, as we now explain. By virtue of Theorem 1 we know that for the linear advection equation with constant coefficients all spacial derivatives satisfy an evolution equation, namely the homogeneous

linear advection equation. We then define m derivative Riemann problems

$$\partial_t v + \lambda \partial_x v = 0 \\ v(x, 0) = \begin{cases} q_i^{(k)} & \text{if } x < 0 \\ q_{i+1}^{(k)} & \text{if } x > 0 \end{cases} \quad (41)$$

where the initial condition for the k -th Riemann problem is given by

$$\begin{aligned} q_i^{(k)} &\equiv \lim_{x \rightarrow x_{i+\frac{1}{2}}^-} \partial_x^{(k)} q(x, 0) \\ q_{i+1}^{(k)} &\equiv \lim_{x \rightarrow x_{i+\frac{1}{2}}^+} \partial_x^{(k)} q(x, 0) \end{aligned} \quad (42)$$

and $k = 0, 1, \dots, m - 1$. Note that all derivative Riemann problems are defined even if there are discontinuities at inter-cell boundaries. Thus the average ADER state (40) is completely defined and the ADER numerical flux is given as in (40).

The average state in (40) has leading term $q_{i+\frac{1}{2}}^{(0)}$, which corresponds to the Godunov first-order upwind method. Second and higher order schemes ($m \geq 2$) include a function evaluated at times

$$t_k = \alpha_k \Delta t, \quad \text{with weight } \alpha_k \equiv \frac{1}{(k+1)^{\frac{1}{k}}}, \quad k \geq 1. \quad (43)$$

It is easy to see that the sequence of *time weights* α_k lie in the interval $[\frac{1}{2}, 1]$, as $\lim_{k \rightarrow \infty} \alpha_k = 1$. For all times t_k we then have $t_k \in [\frac{1}{2}\Delta t, \Delta t]$. Fig. 2 shows a graph of the time weights α_k for schemes of accuracy $m = 2, \dots, 100$.

3.6. RE-INTERPRETATION OF ADER SCHEMES

Substitution of the numerical fluxes (40) into the conservative formula (3) produces

$$q_i^{n+1} = q_i^n - \sum_{k=0}^{m-1} \frac{(\Delta t)_k}{\Delta x} \left[\lambda p_{i+\frac{1}{2}}^{(k)} - \lambda p_{i-\frac{1}{2}}^{(k)} \right], \quad (44)$$

with

$$p_{i+\frac{1}{2}}^{(k)} = (-\lambda)^k q_{i+\frac{1}{2}}^{(k)}, \quad (\Delta t)_k = \frac{(\Delta t)^{k+1}}{(k+1)!}. \quad (45)$$

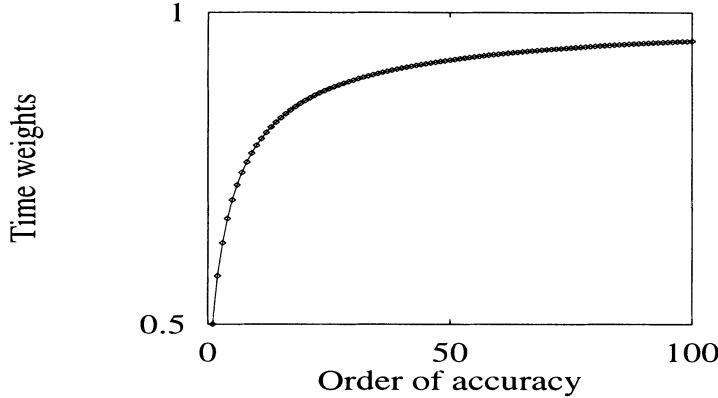


Figure 2. Variation of the time weights α_k with the order of accuracy m , for $m = 2, \dots, 100$.

For $m = 1$ we have a single correction to the initial data q_i^n given by the Godunov first-order upwind method with a time step $(\Delta t)_1 = \Delta t$. For $m = 2$ we add a *flux correction* of gradients with a time step $(\Delta t)_2 = \frac{1}{2}(\Delta t)^2$ to the first-order update. In general, the correction term

$$\frac{(\Delta t)_l}{\Delta x} \left[\lambda p_{i+\frac{1}{2}}^{(l)} - \lambda p_{i-\frac{1}{2}}^{(l)} \right] \quad (46)$$

increases the accuracy of the l -th order update of q_i^n by one.

4. ADER for Multiple Space Dimensions

We first consider the scalar two-dimensional linear advection equation

$$\partial_t q + \lambda_1 \partial_x q + \lambda_2 \partial_y q = 0, \quad (47)$$

for $q(x, y, t)$, where λ_1 and λ_2 are constant components of velocity. We are interested in un-split, explicit one-step finite volume schemes of the form

$$q_{i,j}^{n+1} = q_{i,j}^n - \frac{\Delta t}{\Delta x} \left[f_{i+\frac{1}{2},j} - f_{i-\frac{1}{2},j} \right] - \frac{\Delta t}{\Delta y} \left[g_{i,j+\frac{1}{2}} - g_{i,j-\frac{1}{2}} \right], \quad (48)$$

where $q_{i,j}^n$ is the integral average in a volume $I_{i,j}$ and $f_{i+\frac{1}{2},j}$, $g_{i,j+\frac{1}{2}}$ are numerical fluxes. The ADER approach seeks a time average state at each interface to compute the corresponding numerical flux. Consider the truncated Taylor expansion in time about $t = 0$ is

$$q(x, y, \tau) = \sum_{k=0}^{m-1} \frac{\tau^k}{k!} \partial_t^{(k)} q(x, y, 0). \quad (49)$$

The time average of this between $t = 0$ and $t = \Delta t$ is given by

$$\tilde{q} = \frac{1}{\Delta t} \int_0^{\Delta t} \left\{ \sum_{k=0}^{m-1} \frac{(\Delta t)^k}{k!} \partial_t^{(k)} q(x, y, 0) \right\} dt , \quad (50)$$

which produces

$$\tilde{q} = \sum_{k=0}^{m-1} \frac{(\Delta t)^k}{(k+1)!} \partial_t^{(k)} q(x, y, 0) . \quad (51)$$

But the time derivatives are

$$\partial_t^{(k)} = (-\lambda_1 \partial_x^{(1)} - \lambda_2 \partial_y^{(1)})^k = \sum_{A_k} \binom{k}{\xi} (-\lambda_1)^\xi (-\lambda_2)^\eta \partial_x^{(\xi)} \partial_y^{(\eta)} , \quad (52)$$

which if used in (51) produce

$$\tilde{q} = \sum_{k=0}^{m-1} \frac{(\Delta t)^k}{(k+1)!} \sum_{A_k} \binom{k}{\xi} (-\lambda_1)^\xi (-\lambda_2)^\eta q^{(\xi\eta)}(x, y, 0) , \quad (53)$$

where

$$A_k = \{(\xi, \eta) : \xi + \eta = k, \xi, \eta \geq 0\} . \quad (54)$$

As we are concerned with the average flux value over a cell boundary, at the cell inter-cell position $x = x_{i+\frac{1}{2}}$ for example, we have

$$\tilde{q}_{i+\frac{1}{2},j} = \sum_{k=0}^{m-1} \frac{\Delta t^k}{(k+1)} \sum_{A_k} \frac{(-\lambda_1)^\xi (-\lambda_2)^\eta}{\xi! \eta! \Delta y} \int_j q^{(\xi\eta)}(x_{i+1/2}, y, 0) dy , \quad (55)$$

where $q^{(\xi\eta)}$ denotes the mixed partial derivatives

$$q^{(\xi\eta)} = \partial_x^{(\xi)} \partial_y^{(\eta)} q . \quad (56)$$

The numerical flux $f_{i+\frac{1}{2},j}$ at the inter-cell position $x = x_{i+\frac{1}{2}}$ is then given by

$$f_{i+\frac{1}{2},j} = f(\tilde{q}_{i+\frac{1}{2},j}) . \quad (57)$$

The three-dimensional case follows in an entirely analogous manner. The equation

$$\partial_t q + \lambda_1 \partial_x q + \lambda_2 \partial_y q + \lambda_3 \partial_z q = 0 \quad (58)$$

for $q(x, y, z, t)$, where λ_1, λ_2 and λ_3 are constant components of velocity, is solved by the numerical scheme

$$\left. \begin{aligned} q_{i,j,k}^{n+1} = q_{i,j,k}^n & - \frac{\Delta t}{\Delta x} \left[f_{i+\frac{1}{2},j,k} - f_{i-\frac{1}{2},j,k} \right] - \frac{\Delta t}{\Delta y} \left[g_{i,j+\frac{1}{2},k} - g_{i,j-\frac{1}{2},k} \right] \\ & - \frac{\Delta t}{\Delta z} \left[h_{i,j,k+\frac{1}{2}} - h_{i,j,k-\frac{1}{2}} \right] . \end{aligned} \right\} \quad (59)$$

The numerical flux $f_{i+\frac{1}{2},j,k}$ at the inter-cell position $x = x_{i+\frac{1}{2}}$ is given by

$$f_{i+\frac{1}{2},j,k}^{ader} = f(q_{i+\frac{1}{2},j,k}^{ader}) , \quad (60)$$

with

$$q_{i+\frac{1}{2},j,k}^{ader} = \sum_{k=0}^{m-1} \frac{(\Delta t)^k}{(k+1)} \sum_{A_k} \frac{(-\lambda_1)^\xi (-\lambda_2)^\eta (-\lambda_3)^\zeta}{\xi! \eta! \zeta! \Delta y \Delta z} \int_j \int_k Q \, dy \, dz . \quad (61)$$

where the integrand Q involves mixed partial derivatives, namely

$$Q \equiv q^{(\xi\eta\zeta)}(x_{i+1/2}, y, z, 0) = \partial_x^{(\xi)} \partial_y^{(\eta)} \partial_z^{(\zeta)} q(x_{i+1/2}, y, z, 0) \quad (62)$$

and

$$A_k = \{(\xi, \eta, \zeta) : \xi + \eta + \zeta = k, \xi, \eta, \zeta \geq 0\} . \quad (63)$$

The inter-cell fluxes $g_{i,j+\frac{1}{2},k}$ and $h_{i,j,k+\frac{1}{2}}$ are entirely analogous to (60)-(61). We note here that the computation of the intercell fluxes in multiple space dimensions requires the solution of *mixed derivative Riemann problems*. Fortunately, Theorem 1 for space derivatives in one space dimension extends to multiple space dimensions. The proof is simple and is therefore omitted here.

We have presented the ADER approach for solving the linear advection equation with constant coefficients in one, two and three space dimensions. The schemes can thus be applied directly to linear systems with constant coefficients in one and multiple space dimensions. A pending problem is how to extend ADER to non-linear problems. In the next section we discuss some preliminary ideas.

5. ADER for Non-linear Scalar Equations

Consider the non-linear scalar conservation law

$$\partial_t q + \partial_x f(q) = 0 , \quad (64)$$

which in quasi-linear form reads

$$\partial_t q + \lambda(q) \partial_x f(q) = 0 , \quad (65)$$

where

$$\lambda(q) = \frac{df(q)}{dq} \quad (66)$$

is the *characteristic speed*, which in general is a function of the unknown q . Any extension of the ADER approach requires (i) replacement of the time derivatives $\partial_t^{(k)} q$ by space derivatives $\partial_x^{(k)} q$ via the differential equation and (ii) identification of appropriate evolution equations for the space derivatives. In the linear case step (i) is trivial, see equation (35); step (ii) is also trivial, as the evolution equation for space derivatives is the original linear advection equation, for all derivatives, see Theorem 1. Here we examine three possible approaches for extending ADER to non-linear problems such as (64).

5.1. METHOD 1: SIMPLE LINEARISATION

The approach relies on linearising the equations about the state $q_{i+\frac{1}{2}}^{(0)}$ in (40) at the inter-cell in the flux expansion, as done for the Modified GRP scheme (Toro, 1998). This term corresponds to the Godunov first-order upwind method and results from solving a non-linear Riemann problem. Then the characteristic speed is set to

$$\hat{\lambda}_{i+\frac{1}{2}}^{(0)} = \lambda(q_{i+\frac{1}{2}}^{(0)}) . \quad (67)$$

Then time derivatives are

$$\partial_t^{(k)} q = (-1)^k \hat{\lambda}_{i+\frac{1}{2}}^k \partial_x^{(k)} q \quad (68)$$

and the flux expansion is

$$f_{i+\frac{1}{2}}^{ader} = f(q_{i+\frac{1}{2}}^{ader}) ; \quad q_{i+\frac{1}{2}}^{ader} = q_{i+\frac{1}{2}}^{(0)} + \sum_{k=1}^{m-1} \frac{(-\hat{\lambda}_{i+\frac{1}{2}}^{(0)} \Delta t)^k}{(k+1)!} q_{i+\frac{1}{2}}^{(k)} . \quad (69)$$

The scheme is then identical to that for the linear advection equation. Preliminary results obtained by Millington *et al.* (Millington *et al.*, 1999)

and by Schwartzkopff (Schwartzkopff, 1999) are available but we are not yet certain that this procedure will retain the sought accuracy. A systematic assessment of this method is a pending task.

5.2. METHOD 2: RECURSIVE LINEARISATION

A subtle modification of method 1 results from a kind of recursive linearisation. Based on the time Taylor expansion (33) for the solution of the generalised Riemann problem we define a sequence of approximations in the following way

$$\left. \begin{aligned} s^{(0)} &= q_{i+\frac{1}{2}}^{(0)} \\ s^{(1)} &= q_{i+\frac{1}{2}}^{(0)} + \tau \partial_t^{(1)} q \\ s^{(l)} &= q_{i+\frac{1}{2}}^{(0)} + \sum_{k=1}^l \frac{\tau^k}{k!} \partial_t^{(k)} q \end{aligned} \right\} \quad (70)$$

Clearly

$$s^{(l)} = s^{(l-1)} + \frac{\tau^l}{l!} \partial_t^{(l)} q . \quad (71)$$

Now define wave speeds

$$\lambda_{i+\frac{1}{2}}^{(k)} = \lambda(s^{(k)}) \quad (72)$$

and a corresponding sequence of advection equations

$$\partial_t v + \lambda_{i+\frac{1}{2}}^{(k)} \partial_x v = 0 , \quad (73)$$

where $v \equiv \partial_x^{(k)} q$. To estimate the time derivative $\partial_t^{(k)}$ we use the above sequence to obtain

$$\partial_t^{(k)} q = (-1)^k \left(\prod_{l=0}^{k-1} \lambda_{i+\frac{1}{2}}^{(l)} \right) \partial_x^{(k)} q . \quad (74)$$

In order to evolve the spatial derivatives $\partial_x^{(k)}$ we use the *latest* available characteristic speed, that is the evolution equation (73) with speed $\lambda_{i+\frac{1}{2}}^{(k-1)}$. The final expression of the ADER flux is

$$f_{i+\frac{1}{2}}^{ader} = f(q_{i+\frac{1}{2}}^{ader}) ; \quad q_{i+\frac{1}{2}}^{ader} = q_{i+\frac{1}{2}}^{(0)} + \sum_{k=1}^{m-1} \frac{(-\Delta t)^k}{(k+1)!} \left(\prod_{l=0}^{k-1} \lambda_{i+\frac{1}{2}}^{(l)} \right) q_{i+\frac{1}{2}}^{(k)} . \quad (75)$$

We have not implemented this approach yet, but its simplicity looks very attractive.

5.3. METHOD 3: THE EXACT PROBLEM

Another approach is to effectively solve the full problem. Here the k -th time derivatives are replaced by k -th space derivatives plus all extra terms involving lower order derivatives of q , the characteristic speed λ and mixed derivatives. Consider the exact problem (64)-(66). For a third-order method, for example, we need

$$\left. \begin{aligned} \partial_t^{(1)} q &= -\lambda(q) \partial_x^{(1)} q \\ \partial_t^{(2)} q &= \lambda^2(q) \partial_x^{(2)} q + 2 \frac{d\lambda}{dq} \times \lambda(q) \times (\partial_x^{(1)} q)^2 \end{aligned} \right\} \quad (76)$$

We also need to identify the evolution equations for the space derivatives. For a third-order method these are

$$\left. \begin{aligned} v \equiv \partial_x^{(1)} q, \quad \partial_t^{(1)} v + \lambda \partial_x^{(1)} v &= -v^2 \times \frac{d\lambda}{dq} \\ v \equiv \partial_x^{(2)} q, \quad \partial_t^{(1)} v + \lambda \partial_x^{(1)} v &= -3v(\partial_x^{(1)} q)(\frac{d\lambda}{dq}) - (\partial_x^{(1)} q)^3 (\frac{d^2\lambda}{dq^2}) \end{aligned} \right\} \quad (77)$$

We are fortunate in that all the evolution equations have the same form; however they now have variable coefficients and source terms; these depend on the unknown of the problem and on lower-order derivatives. Applying the recursive procedure of method 2, the source terms are assumed to be known from the previous steps and thus can easily be evaluated. A systematic assessment of this approach is a pending task.

6. Numerical Experiments

In this section we present some illustrative numerical results for the one and two dimensional linear advection equation with constant coefficients. First we solve the linear advection equation in (23) with speed $\lambda = 1$. The computational domain is $[0, 1]$ is discretised into N equally-spaced cells of

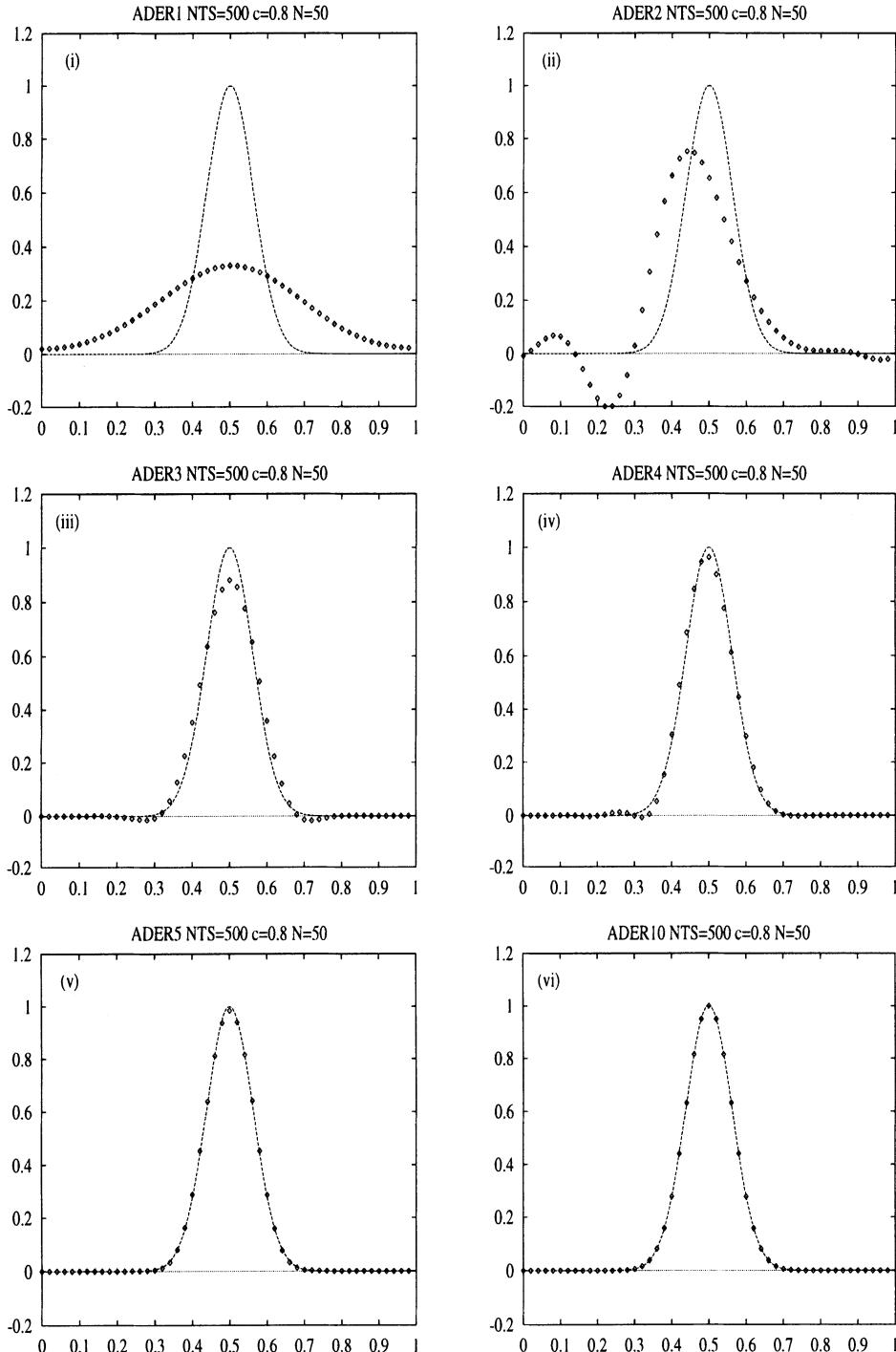


Figure 3. ADER computed results (symbol) for the linear advection equation after $NTS = 500$ time steps for smooth initial condition IC1 in (78) using $N = 50$ cells and CFL number $C = 0.8$. ADER k denotes the linear (fixed stencil) ADER scheme of k -th order of accuracy in space and time.

length Δx . Two initial conditions are considered, namely

$$\left. \begin{array}{ll} \text{IC1: Smooth} & q(x, 0) = e^{-128(x-\frac{1}{2})^2} \\ \text{IC2: Discontinuous} & q(x, 0) = \Pi_{0.2}(x - \frac{1}{2}) \end{array} \right\} \quad (78)$$

The computations reported were performed on an SGI machine to double precision, on a mesh of $N = 50$ cells and a CFL coefficient $c = 0.8$. We applied periodic boundary conditions $q(x, t) = q(x+1, t) \forall t$. The solution was evolved by $NTS = 500$ time steps, which corresponds to a time of $t = 8.0$ units. Results are shown in Figs. 3 to 7, where the numerical computations, shown by symbols, are compared with the exact solution, shown by a full line. Fig. 3 shows computed results for the smooth initial condition in (78). The numerical results were obtained from linear ADER schemes (fixed stencil) of orders of accuracy 1, 2, 3, 4, 5 and 10; see labels $\text{ADER}k$, where k corresponds to the order of accuracy. The numerical results, shown by symbols, are compared with the exact solution, shown by full line. $\text{ADER}1$ corresponds to the Godunov first-order upwind method. The scheme is monotone but is clearly too inaccurate, the extremum has been severely clipped to about 30% of its exact value. The second-order method is also very inaccurate, there is an obvious dispersive error and spurious oscillations are present, even for this problem with smooth solution. The third and fourth order methods give fairly accurate results, but it is only when we use the fifth order scheme that we compute a satisfactory result, at least for this output time and mesh. The 10th order method gives very accurate results. In Fig. 4 we show some results for the same problem comparing the non-linear ADER (adaptive stencil) schemes (ADER-LOS) with the corresponding ENO schemes used in conjunction with TVD Runge-Kutta ODE solvers. Figures on the left show ENO results and figures on the right show ADER-LOS results. Both second order non-linear methods give similar results and are both very inaccurate. Compare these second order results with the corresponding linear ADER result of Fig. 3. The spurious oscillations have been eliminated but the maximum has an error comparable to that of the first-order method. The third order schemes show a clear difference, with ADER-LOS giving a very accurate solution; compare $\text{ADER}3\text{-LOS}$ with $\text{ADER}3$ of Fig. 3. The non-linear version of ADER has eliminated

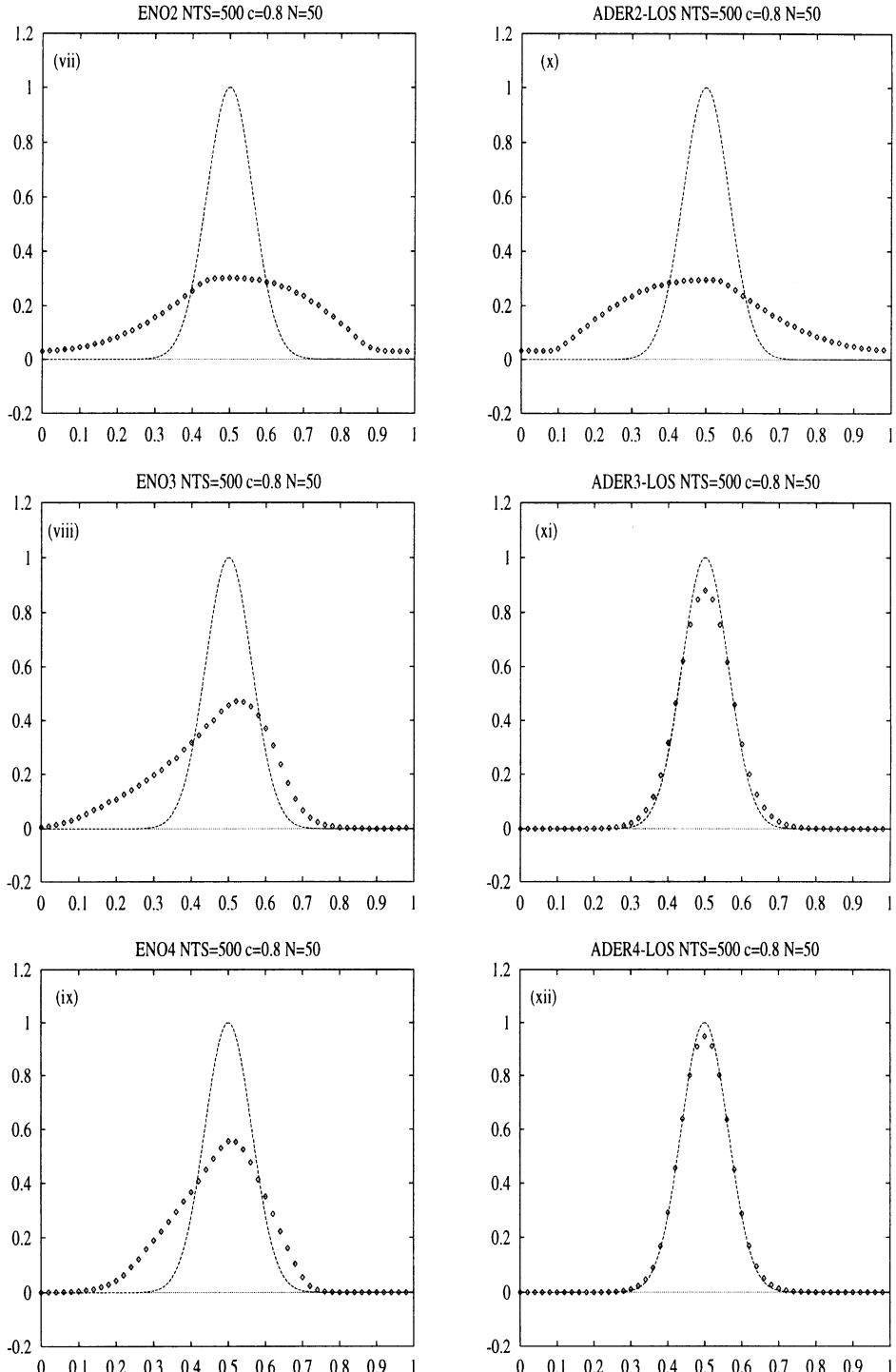


Figure 4. Comparison of nonlinear ADER (denoted by ADER-LOS) with ENO schemes for smooth initial condition IC1 in (78) using $N = 50$ cells and CFL number $C = 0.8$. ENOk and ADERk denote the k -th order accurate ENO and ADER (non-linear) schemes.

the spurious oscillations of the linear ADER3 while preserving, to a good degree, the maximum value at the peak. The fourth-order ADER-LOS is again significantly more accurate than the fourth order ENO and again it preserves virtually the same value of the maximum computed by the linear ADER4 scheme, while eliminating the spurious oscillations. Figs. 5 and 6 show linear ADER results for the case of discontinuous initial condition in (78). As expected, all schemes of order greater than one produce spurious oscillations near discontinuities. Increasing the order of accuracy does not help significantly. Fig. 6 shows results from non-linear schemes, comparing ENO and ADER. All results are oscillation free; note that both second-order methods give similar but very inaccurate results. As seen for the smooth solution case, the third and fourth order ADER schemes are more accurate than the corresponding ENO schemes.

We have implemented the weighted ENO (WENO) idea in the context of ADER schemes to produce weighted ADER schemes (ADER-WLOS); here one takes a polynomial that is a convex combination of polynomials on all candidate stencils. For ENO schemes this procedure is known to increase the accuracy significantly. The same beneficial effect is observed in the ADER schemes, see results of Fig. 7. Again, the weighted ADER schemes are more accurate than the corresponding weighted ENO schemes. Millington (Millington, 2001) has carried out a systematic comparison between ADER and ENO schemes under a wide range of computational conditions and produced tabulated results of convergence rates and CPU times; these results will be published elsewhere. For the very simple problems solved here ADER gives superior results to ENO/WENO. The ADER schemes are also more efficient, by a factor of about 3, and we are confident that this computational cost difference will increase when solving more complex problems.

As a second example we solved the two-dimensional linear advection equation (47) with $\lambda_1 = 1$ and $\lambda_2 = 1$ in a square domain $[0, 1] \times [0, 1]$ with smooth initial condition

$$q(x, y, 0) = e^{-128[(x - \frac{1}{2})^2 + (y - \frac{1}{2})^2]} \quad (79)$$

and periodic boundary conditions.

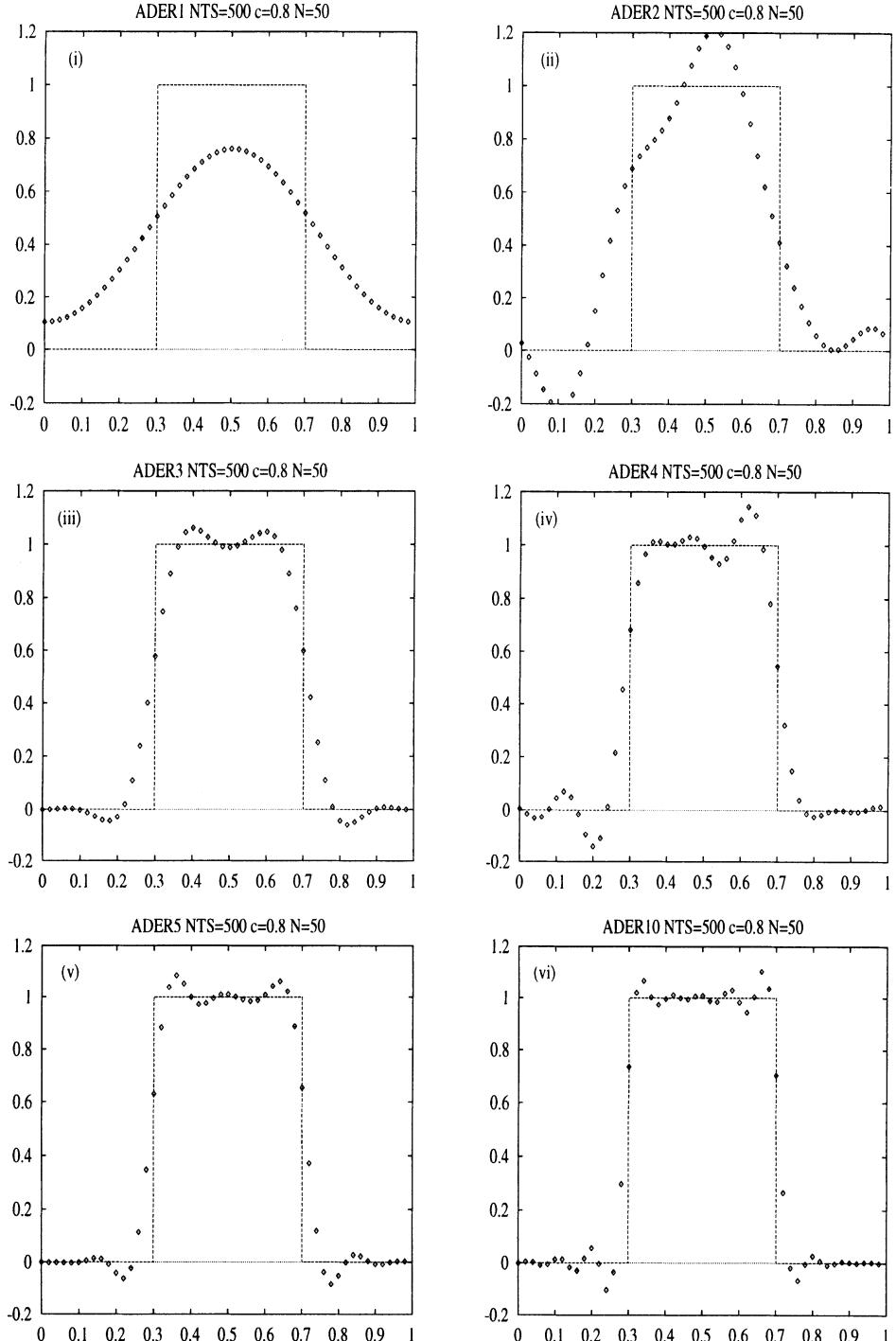


Figure 5. ADER computed results (symbol) for the linear advection equation after $NTS = 500$ time steps for discontinuous initial condition IC2 in (78) using $N = 50$ cells and CFL number $C = 0.8$. ADER k denotes the linear (fixed stencil) ADER scheme of k -th order of accuracy in space and time.

For the computations reported here we used a coarse regular mesh of 25×25 cells, a CFL coefficient $c = 0.9$ and evolved the solution for $NTS = 231$ time steps, which corresponds to an output time of $t = 8$ units. The results shown in Figs. 8 and 9 were obtained from the linear (oscillatory, fixed stencil interpolation) version of ADER. In the exact solution the maximum value is $q_{max} = 1$ and the minimum is $q_{min} = 0$. The numerical values for maximum and minimum give a fair idea of the accuracy of the schemes. Fig. 8 shows the computed results for first, second and third order schemes (ADER1, ADER2 and ADER3). The result from the first order scheme (ADER1) can hardly be seen, the maximum has been very severely clipped to 10% of the correct value, $q_{max} = 0.102$. The results from the second and third order schemes (ADER2 and ADER3) can be seen but are still very inaccurate; they also exhibit spurious oscillations and their maxima are $q_{max} = 0.381$ and $q_{max} = 0.568$ respectively.

Fig. 9 shows results obtained from the linear (oscillatory, fixed stencil interpolation) version of ADER for orders of accuracy 4, 5 and 10. The respective maxima are $q_{max} = 0.715$, $q_{max} = 0.816$ and $q_{max} = 0.965$. Of these, one may say that it is only the tenth-order result that is sufficiently accurate. This two-dimensional example on a coarse mesh shows quite dramatically that on the coarse meshes that are likely to be of practical use in real applications, low-order methods are simply too inaccurate to be used.

7. Conclusions and Further Developments

The ADER approach for constructing non-oscillatory advection schemes of very high order of accuracy in space and time has been presented. The ADER formulation for linear advection with constant coefficients for one, two and three dimensions have been given in complete detail, along with some numerical results. Comparison with the state-of-the art ENO/WENO schemes has been made. The comparisons reveal that the ADER schemes are more accurate; they are also simpler and more efficient. Another attractive property of the ADER schemes, as compared with their closest

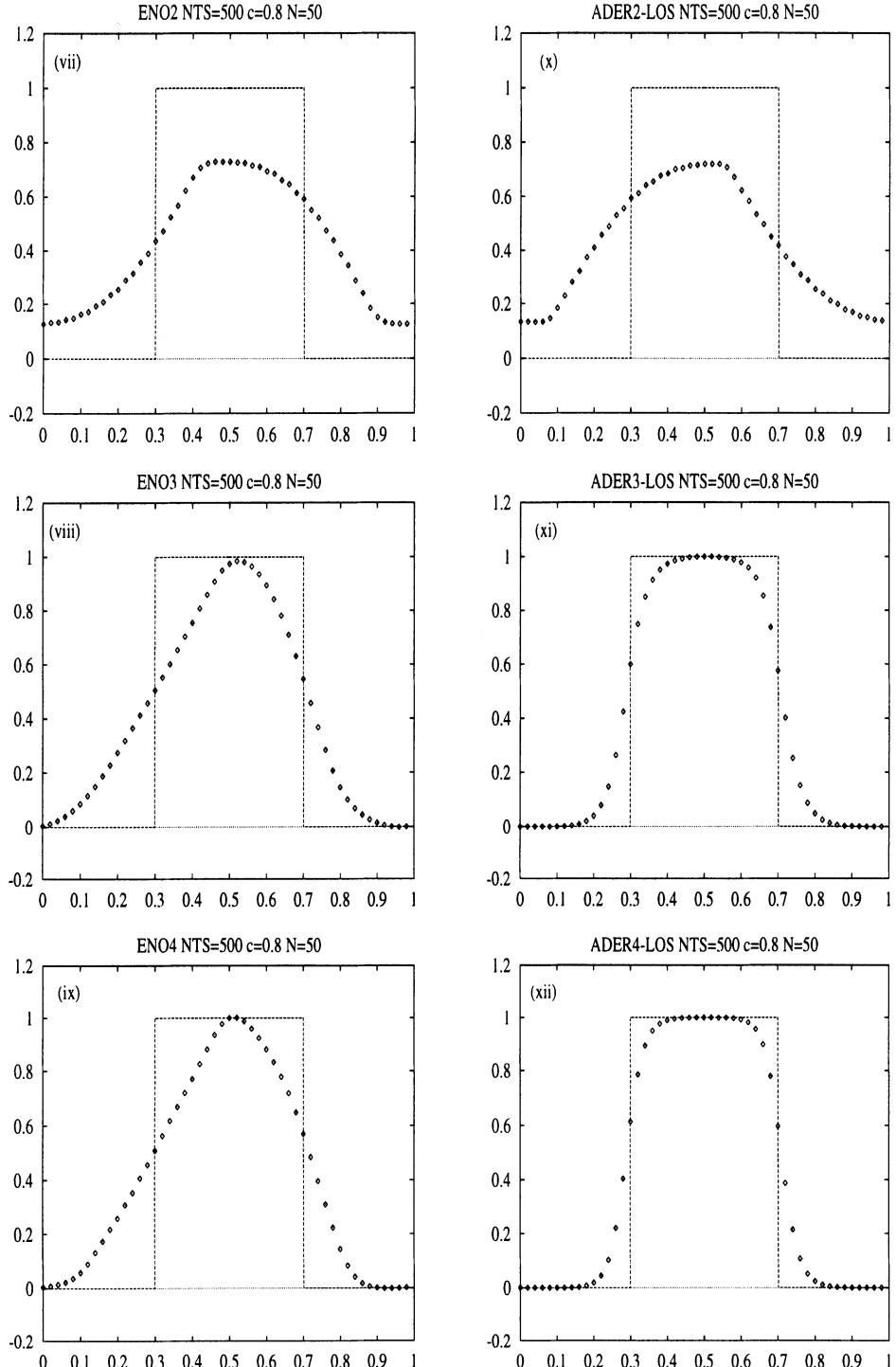


Figure 6. Comparison of nonlinear ADER (denoted by ADER-LOS) with ENO schemes for discontinuous initial condition IC2 in (78) using $N = 50$ cells and CFL number $C = 0.8$. ENOk and ADERk denote the k -th order accurate ENO and ADER (non-linear) schemes.

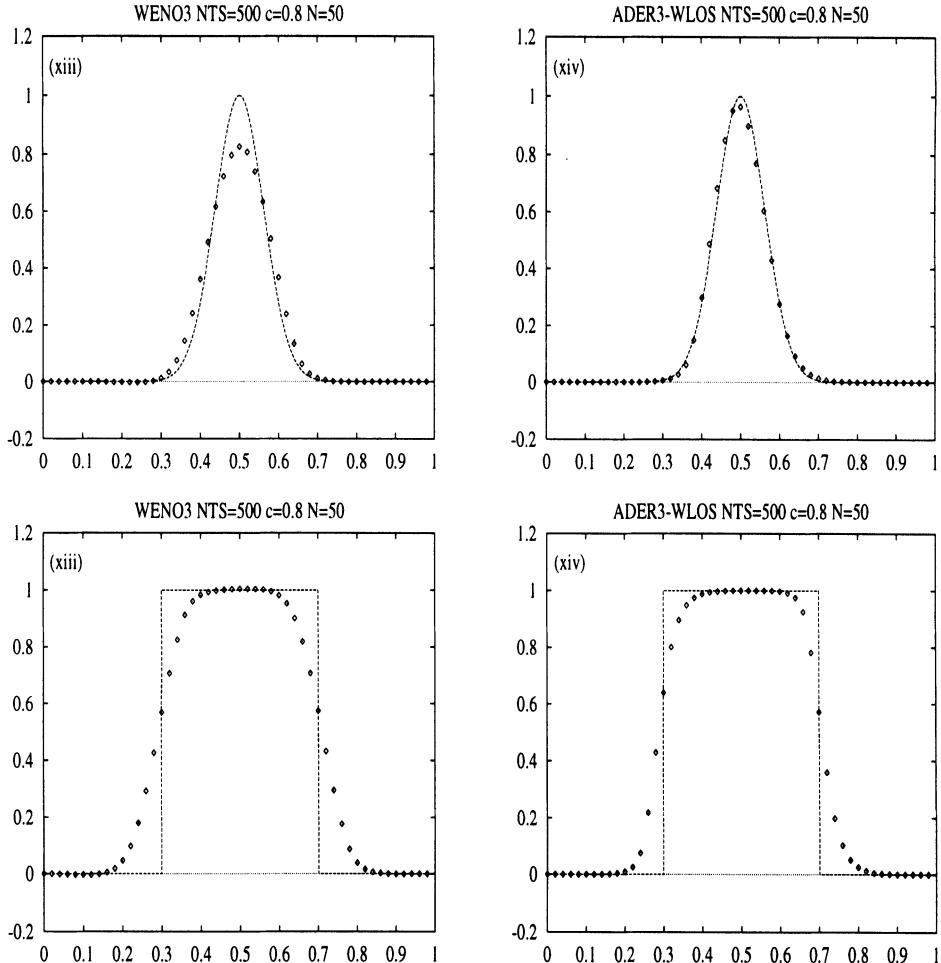
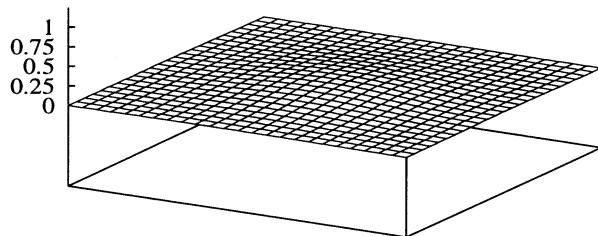


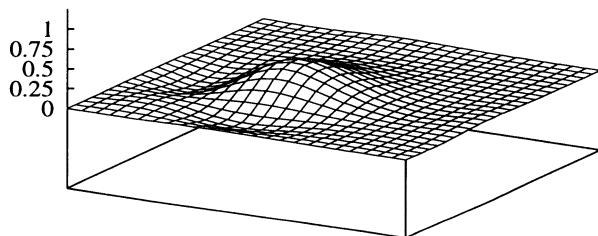
Figure 7. Comparison of nonlinear Weighted ADER (denoted by ADER-WLOS) with Weighted ENO schemes for *smooth* initial condition IC1 in (78) using $N = 50$ cells and CFL number $C = 0.8$. ENOk and ADERk denote the k -th order accurate ENO and ADER (non-linear) schemes.

competitors, the ENO/WENO schemes, is that for ADER we have not yet found accuracy barriers, at least for linear systems with constant coefficients. The ENO/WENO schemes rely on TVD solvers for the associated ordinary differential equations in time; so far the most accurate TVD ODE solver developed for use with ENO/WENO schemes is of fifth order of accuracy. As they stand, the ADER schemes can be extended to linear systems with constant coefficients. Schwartzkopff (Schwartzkopff, 1999) has already applied the ADER approach to the linearised Euler equations of gas dynamics with very encouraging results. Munz and Schneider (Munz

Ader1



Ader2



Ader3

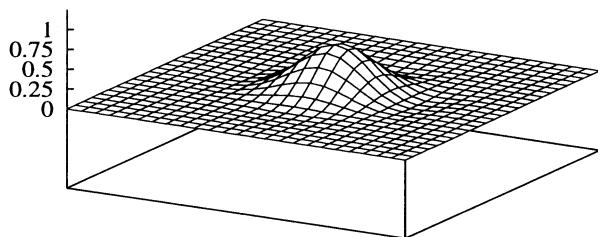
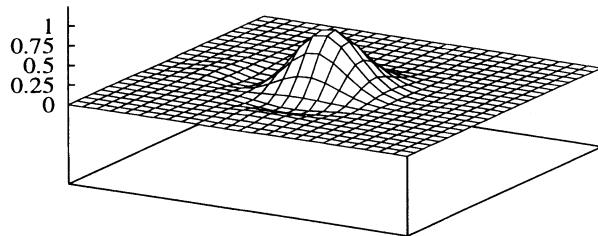
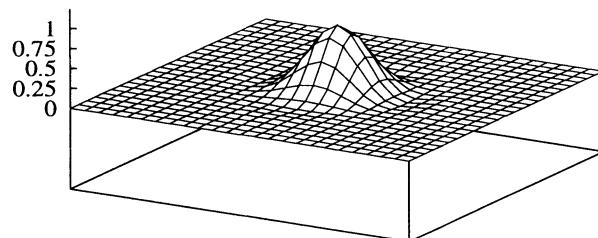


Figure 8. Two-dimensional ADER results for linear advection equation with constant coefficients $\lambda_1 = \lambda_2 = 1$. Results shown are for the linear (fixed stencil) ADER schemes of first, second and third order accuracy.

Ader4



Ader5



Ader10

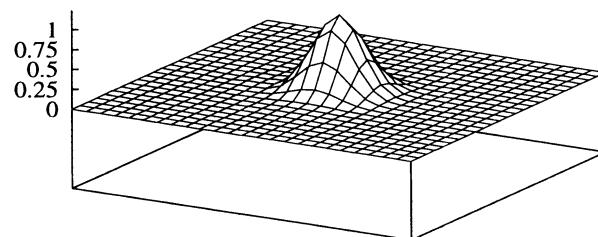


Figure 9. Two-dimensional ADER results for linear advection equation with constant coefficients $\lambda_1 = \lambda_2 = 1$. Results shown are for the linear (fixed stencil) ADER schemes of fourth, fifth and tenth order accuracy.

and R. Schneider, 2000) have extended the ADER approach to unstructured meshes to solve the two-dimensional linearised Maxwell equations. In spite of these two-dimensional applications the practical experience on implementation issues is still very limited; there is so far no experience on three-dimensional implementations. Preliminary results on the application of centred fluxes, rather than upwind, in the context of ADER show that for the low order range upwind schemes are more accurate but for high order of accuracy the results are virtually indistinguishable. This is an area that requires more work. It may well be that one can use approximate Riemann solvers for the first few terms in the flux expansion, say two, and centred schemes for the remaining terms.

The next important development of the ADER approach is its extension to non-linear problems. Here we have presented three potential approaches, two of them relying heavily on local linearisations. Limited practical experience is available with one of the approaches, the simplest, but it is so far uncertain as to whether such simple approach is capable of retaining the expected high accuracy. We are confident that the other two approaches or some combination of the two will be productive, but this is speculative at this stage. Another possibility is the use of the relaxation approach of Shi and (Shi and X. Zhouping, 1995). Schwartzkopff and Schroll (private communication) are already working in this direction.

The ADER schemes for linear systems with constant coefficients in one and multiple space dimensions are so far reasonably well understood, although one would like to be able to prove rigourously the stability limit of the schemes, for example. The indications so far are that the schemes have stability unity for all orders in one and multiple space dimensions. Analysis of ADER schemes for non-linear schemes, if successful, is bound to be simpler than for multi-level schemes.

References

- M. Ben-Artzi and J. Falcovitz (1984). A Second Order Godunov-Type Scheme for Compressible Fluid Dynamics. *J. Comput. Phys.*, **55**, pp 1–32.
- S. J. Billett and E. F. Toro. WAF-Type Schemes for Multidimensional Hyperbolic Conservation Laws. *J. Comp. Phys.*, **130**, pp 1–24, 1997.
- M. Cáceres. Development of a Third-Order Accurate Scheme of the MUSCL Type for the Time-Dependent One-Dimensional Euler Equations. MSc. Thesis, Department of Aerospace Science, Cranfield University, UK, 1993.
- C. Canuto and A. Quarteroni. *Spectral and Higher Order Methods for Partial Differential Equations*. North-Holland, 1990.
- D. N. Cheney. Upwinding Convective and Viscous Terms via a Modified GRP Approach. MSc. Thesis, Department of Aerospace Science, Cranfield University, UK, 1994.
- B. Cockburn and C. W. Shu. The Runge-Kutta Discontinuous Galerkin Method for Conservation Laws. *J. Comput. Phys.*, **141**, pp 199–, 1998.
- P. Colella. A Direct Eulerian MUSCL Scheme for Gas Dynamics. *SIAM J. Sci. Stat. Comput.*, **6**, pp 104–117, 1985.
- J. M. Ghidaglia, G. LeCoq, and I. Toumi. Two Flux Schemes for Computing Two Phases Flows Through Multidimensional Finite Volume Methods. In *Proceedings of the NURETH-9 Conference*. American Nuclear Society, 1999.
- E. Godlewski and P. A. Raviart. *Numerical Approximation of Hyperbolic Systems of Conservation Laws*. Springer, 1996.
- S. K. Godunov. Finite Difference Methods for the Computation of Discontinuous Solutions of the Equations of Fluid Dynamics. *Mat. Sb.*, **47**, pp 271–306, 1959.
- A. Harten. High Resolution Schemes for Hyperbolic Conservation Laws. *J. Comput. Phys.*, **49**, pp 357–393, 1983.
- A. Harten, B. Engquist, S. Osher, and S. R. Chakravarthy. Uniformly High Order Accuracy Essentially Non-oscillatory Schemes III. *J. Comput. Phys.*, **71**, pp 231–303, 1987.
- A. Harten and S. Osher. Uniformly High-Order Accurate Nonoscillatory Schemes I. *SIAM J. Numer. Anal.*, **24**, No 2, pp 279–309, 1987.
- C. Hirsch. *Numerical Computation of Internal and External Flows, Vol. I: Fundamentals of Numerical Discretization*. Wiley, 1988.
- D. Kröner. *Numerical Schemes for Conservation Laws*. Wiley Teubner, 1997.
- C. B. Laney. *Computational Gasdynamics*. Cambridge University Press, 1998.
- P. Lax and B. Wendroff. Systems of Conservation Laws. *Comm. Pure Appl. Math.*, **13**, pp 217–237, 1960.
- R. J. LeVeque. *Numerical Methods for Conservation Laws*. Birkhäuser, 1992.
- R. C. Millington. *Approaches for Constructing Very High-Order Non-Oscillatory Advection Schemes*. PhD thesis, Department of Computing and Mathematics, Manchester Metropolitan University, UK, 2001 (to appear).
- R. C. Millington, E. F. Toro, and L. A. M. Nejad. Arbitrary High-Order Methods for Conservation Laws I: The One-Dimensional Scalar Case. Technical report, Department of Computing and Mathematics, Manchester Metropolitan University, UK, June, 1999.
- C. D. Munz and R. Schneider. An Arbitrary High Order Accurate Finite Volume Scheme for the Maxwell Equations in Two Dimensions on Unstructured Meshes. Technical report, Forschungszentrum Karlsruhe, Germany, 2000.
- T. Schwartzkopff. The ADER Approach for Linear and Non-linear Advection Diffusion Problems. Technical report, Institut für Aero- und Gasdynamik, Universität Stuttgart, Germany, 1999.
- J. Shi and X. Zhouping. The Relaxation Scheme for Systems of Conservation Laws in Arbitrary Dimensions. *Comm. Pure Appl. Math.*, **48**, pp 235–276, 1995.
- C. W. Shu. Essentially Non-Oscillatory and Weighted Essentially Non-Oscillatory Schemes for Hyperbolic Conservation Laws. Technical report, NASA/CR-97-206253,

- ICASE Number 97-65, November 1997.
- P. K. Sweby. High Resolution Schemes Using Flux Limiters for Hyperbolic Conservation Laws. *SIAM J. Numer. Anal.*, **21**, pp 995–1011, 1984.
- A. I. Tolstykh. *High Accuracy Non-Centred Compact Difference Schemes for Fluid Dynamics Applications*. World Scientific Publishing, 1994.
- E. F. Toro. A Weighted Average Flux Method for Hyperbolic Conservation Laws. *Proc. Roy. Soc. London*, **A423**, pp 401–418, 1989.
- E. F. Toro. On Glimm-Related Schemes for Conservation Laws. Technical Report MMU-9602, Department of Mathematics and Physics, Manchester Metropolitan University, UK, 1996.
- E. F. Toro. *Riemann Solvers and Numerical Methods for Fluid Dynamics*. Springer-Verlag, 1997.
- E. F. Toro. Primitive, Conservative and Adaptive Schemes for Hyperbolic Conservation Laws. In *Numerical Methods for Wave Propagation*. Toro, E. F. and Clarke, J. F. (Editors), pages 323–385. Kluwer Academic Publishers, 1998.
- E. F. Toro. *Riemann Solvers and Numerical Methods for Fluid Dynamics, Second Edition*. Springer-Verlag, 1999.
- E. F. Toro and S. J. Billett. Centred TVD Schemes for Hyperbolic Conservation Laws. *IMA J. Numerical Analysis*, **20**, pp 47–79, 2000.
- J. J. W. van der Vegt, van der Ven H., and O. J. Boelens. Discontinuous Galerkin Methods for Partial Differential Equations. In *Godunov Methods: Theory and Applications (Edited Review)*, E. F. Toro (Editor). Kluwer Academic/Plenum Publishers, 2001.
- B. van Leer. On the Relation Between the Upwind-Differencing Schemes of Godunov, Engquist-Osher and Roe. *SIAM J. Sci. Stat. Comput.* **5**, pp 1–20, 1985.

MODEL HYPERBOLIC SYSTEMS WITH SOURCE TERMS: EXACT AND NUMERICAL SOLUTIONS

E. F. TORO

*Department of Computing and Mathematics,
Manchester Metropolitan University,
Chester Street, Manchester, M1 5GD, U.K.
Emails: E.F.Toro@doc.mmu.ac.uk*

AND

M. E. VAZQUEZ-CENDON

*Department of Applied Mathematics,
University of Santiago de Compostela, Spain.
Email: elena@zmat.usc.es*

Abstract. We construct model hyperbolic systems with source terms, give their exact solutions and propose a new numerical approach to solve inhomogeneous hyperbolic problems. Preliminary numerical results, as compared with exact solutions, show that the proposed numerical method is superior to existing methods for dealing with source terms.

1. Introduction

Significant progress has been made, in the last two decades or so, in the design, analysis and application of numerical methods for solving hyperbolic conservation laws. Godunov methods are prominent in these developments and have by now, in many ways, reached a stage of maturity. For background information on Godunov methods the reader should consult the references (Godunov, 1959), (LeVeque, 1992), (Godlewski and Raviart, 1996), (Kröner, 1997), (Harten et. al., 1987), (Roe, 1998), (Toro and Clarke, 1998) and (Toro, 1999) and the many references therein. However, when source or forcing terms are added to the conservation laws the situation is far from clear. This is in many ways paradoxical, as source terms are *algebraic* functions of the unknown and one would expect the

numerical challenges to come from the differential terms rather than from the algebraic terms. On the other hand most of the physics/chemistry may come from the source terms. In fact, if the initial conditions are uniform, flow or wave propagation will be provoked exclusively by the source terms. For many problems simple-minded methods perform deceptively well. Particular difficulties are posed by *stiff* source terms associated with chemical kinetics, by geometric-type source terms and by highly time-dependent source terms. This paper is motivated by time-dependent source terms and by geometric source terms, such as those arising in shallow water models. In Sect. 2 we present hyperbolic models with source terms along with exact solutions. A shallow water example is presented in Sect. 3 and in Sect. 4 we propose a new numerical method to treat source terms. Some numerical examples are shown in Sect. 5 and conclusions are drawn in Sect. 6.

2. Model Hyperbolic System with Source Terms

First we consider the scalar inhomogeneous hyperbolic PDE

$$q(x, t)_t + \lambda q(x, t)_x = s(x, t, q(x, t)) , \quad (1)$$

where the unknown of the problem is $q(x, t)$, λ is a constant wave propagation speed and s is a source term, assumed to be an algebraic function of the unknown $q(x, t)$, allowing for explicit dependence on x and t . We seek solutions of (1) which are of the form

$$q(x, t) = c(x - \lambda t)b(x)g(t) , \quad (2)$$

where $c(s)$ represents a travelling wave and $b(x)$, $g(t)$ are two arbitrary functions of distance x and time t respectively. The initial condition is

$$q(x, 0) = c(x)b(x)g(0) \quad (3)$$

and we seek the corresponding form of the source term $s(x, t, q)$ in (1). Taking partial derivatives of $q(x, t)$ in (2) with respect to t and x and substituting the corresponding expressions in (1) gives

$$q(x, t)_t + \lambda q(x, t)_x = c(x - \lambda t)b(x)g'(t) + \lambda b'(x)c(x - \lambda t)g(t) . \quad (4)$$

We consider algebraic functions $b(x)$ and $g(t)$ that are non-vanishing in the domain of interest. Then the source term takes the form

$$s(x, t, q) = \alpha(x, t)q(x, t) , \text{ with } \alpha(x, t) = \frac{g'(t)}{g(t)} + \lambda \frac{b'(x)}{b(x)} . \quad (5)$$

Expression (2) is the exact solution of (1) with source term (5) and initial condition (3). Note that the source term contains a part that resembles *area*

variation in one-dimensional gas dynamics, or *breadth variation* in one-dimensional shallow water models. It is in fact this analogy what partly motivates this work, as *geometric-type source terms* are recognised to be difficult to deal with numerically. We now extend the previous ideas to construct model hyperbolic systems with source terms

$$\mathbf{U}_t + \mathbf{A}\mathbf{U}_x = \mathbf{S}(x, t, \mathbf{U}), \quad (6)$$

where \mathbf{U} is the vector of unknowns and \mathbf{A} is a constant coefficient matrix. Here we consider the 2×2 case, with $\mathbf{U} = [u_1, u_2]^T$. Denote the real eigenvalues of \mathbf{A} by λ_1 and λ_2 and the the matrix of corresponding right eigenvectors by \mathbf{K} . Then we write

$$\mathbf{A} = \mathbf{K}\Lambda\mathbf{K}^{-1}, \text{ with } \Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}. \quad (7)$$

By transforming to characteristic variables $\mathbf{Q} = \mathbf{K}^{-1}\mathbf{U}$ in the usual way, one obtains the following set of decoupled equations

$$\mathbf{Q}_t + \Lambda\mathbf{Q}_x = \Lambda^{(s)}(x, t)\mathbf{Q}, \quad (8)$$

where $\Lambda^{(s)}$ denotes a diagonal matrix

$$\Lambda^{(s)}(x, t) = \begin{bmatrix} \alpha_1^{(s)}(x, t) & 0 \\ 0 & \alpha_2^{(s)}(x, t) \end{bmatrix}. \quad (9)$$

The form of the source terms in (9) is chosen so as to reproduce the scalar case (5), namely

$$\alpha_1^{(s)}(x, t) = \frac{g'_1(t)}{g_1(t)} + \lambda_1 \frac{b'_1(x)}{b_1(x)} \text{ and } \alpha_2^{(s)}(x, t) = \frac{g'_2(t)}{g_2(t)} + \lambda_2 \frac{b'_2(x)}{b_2(x)}. \quad (10)$$

The functions $b_1(x), b_2(x)$ and $g_1(t), g_2(t)$ are prescribed algebraic functions with the same restrictions as for b and g in the scalar case. Then, the source term can be factored via a matrix $\mathbf{M}^{(s)}(x, t)$, namely

$$\mathbf{S}(x, t, \mathbf{U}) = \mathbf{M}^{(s)}(x, t)\mathbf{U}, \text{ with } \mathbf{M}^{(s)}(x, t) = \mathbf{K}\Lambda^{(s)}(x, t)\mathbf{K}^{-1}. \quad (11)$$

The initial conditions are

$$\mathbf{U}(x, 0) = \mathbf{K} \begin{bmatrix} b_1(x)g_1(t) & 0 \\ 0 & b_2(x)g_2(t) \end{bmatrix} \begin{bmatrix} c_1(x) \\ c_2(x) \end{bmatrix} \quad (12)$$

where c_1 and c_2 are given functions, analogous to the c function in the scalar case. The vector $[c_1(x), c_2(x)]^T$ represents the initial condition for the

homogeneous system associated with (6) written in terms of characteristic variables and which contains only the advective part \mathbf{Q}_0^c . Hence

$$\mathbf{Q}^c(x, 0) = \mathbf{Q}_0^c(x) = \begin{bmatrix} c_1(x) \\ c_2(x) \end{bmatrix} = \mathbf{K}^{-1}\mathbf{U}^c(x, 0) \quad (13)$$

and

$$\mathbf{Q}^c(x, t) = \begin{bmatrix} c_1(x - \lambda_1 t) \\ c_2(x - \lambda_2 t) \end{bmatrix} = \mathbf{K}^{-1}\mathbf{U}^c(x, t), \quad (14)$$

where $\mathbf{U}^c(x, t)$ represents the solution of the homogeneous system in terms of the original variables. The exact solution for the system case is

$$\mathbf{U}(x, t) = \mathbf{K} \begin{bmatrix} b_1(x)g_1(t) & 0 \\ 0 & b_2(x)g_2(t) \end{bmatrix} \begin{bmatrix} c_1(x - \lambda_1 t) \\ c_2(x - \lambda_2 t) \end{bmatrix}. \quad (15)$$

Apart from the restrictions imposed on the form of the source terms in the characteristic variable form of the equations, the previous developments are fairly general and include $n \times n$ systems. Specific examples result from choosing the vector \mathbf{U} of unknowns and the coefficient matrix \mathbf{A} . In the next section we study models associated with free-surface shallow fluids.

3. Example: Linearised Shallow Water

As an example we consider the time-dependent one dimensional linearised shallow water equations

$$\mathbf{U}_t + \mathbf{A}\mathbf{U}_x = \mathbf{S}(x, t, \mathbf{U}), \text{ with } \mathbf{U} = \begin{bmatrix} h \\ u \end{bmatrix} \text{ and } \mathbf{A} = \begin{bmatrix} \bar{u} & \bar{h} \\ g & \bar{u} \end{bmatrix}. \quad (16)$$

The unknowns of the problem are $h(x, t)$ for water depth and $u(x, t)$ for particle velocity. Here \bar{h} is a constant depth, \bar{u} is a constant velocity and g is the acceleration due to gravity. The eigenvalues are $\lambda_1 = \bar{u} - \bar{a}$ and $\lambda_2 = \bar{u} + \bar{a}$, where $\bar{a} = \sqrt{gh}$ is the *celerity*. The matrix of right eigenvectors and its inverse are respectively

$$\mathbf{K} = \begin{bmatrix} \bar{h} & \bar{h} \\ -\bar{a} & \bar{a} \end{bmatrix} \text{ and } \mathbf{K}^{-1} = \frac{1}{2\bar{a}\bar{h}} \begin{bmatrix} \bar{a} & -\bar{h} \\ \bar{a} & \bar{h} \end{bmatrix}. \quad (17)$$

The components of the source term vector $\mathbf{S}(x, t, \mathbf{U}) = [s_1, s_2]^T$, in terms of the original variables, are

$$s_1(x, t, \mathbf{U}) = \frac{1}{2}[\alpha_1^{(s)}(x, t) + \alpha_2^{(s)}(x, t)]h - \frac{1}{2}[\alpha_1^{(s)}(x, t) - \alpha_2^{(s)}(x, t)]\frac{\bar{h}}{\bar{a}}u \quad (18)$$

$$s_2(x, t, \mathbf{U}) = -\frac{1}{2}[\alpha_1^{(s)}(x, t) - \alpha_2^{(s)}(x, t)]\frac{\bar{a}}{\bar{h}}h + \frac{1}{2}[\alpha_1^{(s)}(x, t) + \alpha_2^{(s)}(x, t)]u \quad (19)$$

and the exact solution is

$$\mathbf{U}(x, t) = \begin{bmatrix} \bar{h} & \bar{h} \\ -\bar{a} & \bar{a} \end{bmatrix} \begin{bmatrix} b_1(x)g_1(t) & 0 \\ 0 & b_2(x)g_2(t) \end{bmatrix} \begin{bmatrix} c_1(x - \lambda_1 t) \\ c_2(x - \lambda_2 t) \end{bmatrix}. \quad (20)$$

In Sect. 5 we give examples for the functions $b_i(x)$, $g_i(t)$, $c_i(s)$, for $i = 1, 2$, that resemble source terms due to bottom elevation and breadth variation.

4. A Numerical Method

Here we present a numerical method for inhomogeneous hyperbolic systems that reduces identically to the MUSCL-Hancock scheme (Van Leer, 1984) when the source term vanishes. Our extension of the MUSCL-Hancock scheme is related to the methods proposed by LeVeque (LeVeque, 1998) and that by Greenberg and LeRoux (Greenberg and LeRoux , 1996). We present the ideas in terms of the scalar case (1). The scheme has the following steps:

- Reconstruction-recovery:

$$q_i^{(h)}(x) = q_i^n + \frac{(x - x_i)}{\Delta x} \Delta_i^{(h)}, \quad q_i^{(s)}(x) = q_i^n + \frac{(x - x_i)}{\Delta x} \Delta_i^{(s)}, \quad (21)$$

where $x \in [0, \Delta x]$, $\Delta_i^{(h)}$ is a slope (difference) given by the data $\{q_i^n\}$ in the usual way and $\Delta_i^{(s)}$ is a slope (difference) given by the *steady* version of (1), namely

$$\lambda q(x, t)_x = s(x, t, q(x, t)). \quad (22)$$

The boundary extrapolated values for the homogenous part and the source term part are respectively

$$\left. \begin{array}{l} q_i^{(Lh)} = q_i^n - \frac{1}{2} \Delta_i^{(h)}; \quad q_i^{(Rh)} = q_i^n + \frac{1}{2} \Delta_i^{(h)} \\ q_i^{(Ls)} = q_i^n - \frac{1}{2} \Delta_i^{(s)}; \quad q_i^{(Rs)} = q_i^n + \frac{1}{2} \Delta_i^{(s)} \end{array} \right\} \quad (23)$$

- Evolution. The boundary extrapolated values $q_i^{(Lh)}$, $q_i^{(Rh)}$ are evolved by a time $\frac{1}{2}\Delta t$ according to

$$\bar{q}_i^{(hX)} = q_i^{(hX)} + \frac{1}{2} \frac{\Delta t}{\Delta x} [f(q_i^{(hL)}) - f(q_i^{(hR)})] + \frac{1}{2} \Delta t s(q_i^{(sU)}), \quad (24)$$

with $X = L, R$ and superscript sU denotes the *upwind* extrapolated value, namely

$$q_i^{(sU)} = \begin{cases} q_i^{(sL)} & \text{if } \lambda > 0 \\ q_i^{(sR)} & \text{if } \lambda < 0 \end{cases} \quad (25)$$

- *Numerical Flux.* The intercell numerical flux $f_{i+\frac{1}{2}}$ may be obtained by solving the conventional Riemann problem

$$\left. \begin{aligned} q_t + f(q)_x &= 0, \\ q(x, 0) &= \begin{cases} \bar{q}_i^{(Rh)}, & x < 0, \\ \bar{q}_{i+1}^{(Lh)}, & x > 0, \end{cases} \end{aligned} \right\} \quad (26)$$

so that a Godunov flux is $f_{i+\frac{1}{2}}^{god} = f(q_{i+\frac{1}{2}}(0))$, where $q_{i+\frac{1}{2}}(x/t)$ is the solution of the Riemann problem (26).

- *Updating.* The final step to update the solution from time level n to time level $n+1$ is

$$q_i^{n+1} = q_i^n + \frac{\Delta t}{\Delta x} [f_{i-\frac{1}{2}} - f_{i+\frac{1}{2}}] + \Delta t s(q_i^{(sD)}), \quad (27)$$

where superscript sD denotes the *downwind* extrapolated value, namely

$$q_i^{(sD)} = \begin{cases} q_i^{(sR)} & \text{if } \lambda > 0 \\ q_i^{(sL)} & \text{if } \lambda < 0 \end{cases} \quad (28)$$

Note the reversed bias expressed by (25) and (28). In the predictor step the source term is evaluated on the *upwind* boundary extrapolated state and in the corrector step the source term is evaluated on the *downwind* boundary extrapolated state.

For the linear case with $\lambda > 0$ and source term in (22) $s = \alpha q$, with α a constant, the final scheme in full reads

$$\left. \begin{aligned} q_i^{n+1} &= q_i^n + c \left[(q_{i-1}^n - q_i^n) + \frac{1}{2}(1-c)(\Delta_{i-1}^{(h)} - \Delta_i^{(h)}) \right] \\ &\quad + \Delta t \alpha \left(1 - \frac{1}{2}d \right) \left[\frac{(1-\frac{1}{2}d)}{(1-\frac{1}{2}d)} q_i^n + \frac{1}{2}c(q_{i-1}^n - q_i^n) \right] \end{aligned} \right\} \quad (29)$$

where $c = \frac{\lambda \Delta t}{\Delta x}$ is the Courant number and $d = \frac{\Delta x \alpha}{\lambda}$. The first-order part of the scheme reproduces Godunov's method for the advective part and *upwinding* for the source term, in the usual way, namely

$$q_i^{n+1} = q_i^n + c(q_{i-1}^n - q_i^n) + \Delta t \alpha \left[q_i^n + \frac{1}{2}c(q_{i-1}^n - q_i^n) \right] \quad (30)$$

See (Roe, 1987) and (Vázquez-Cendón, 1999).

5. Numerical Results

Here we present some numerical results on the shallow-water type example described in Sect. 3. All solutions are compared with the exact solutions

presented in Sect. 2. For the computations we choose

$$\left. \begin{aligned} g_1(t) &= e^{\alpha_1 t}, \quad g_2(t) = e^{\alpha_2 t} \\ b_1(x) &= 0.3(\cos(x))^2 + 1, \quad b_2(x) = 0.2x^3 + 1 \\ c_1(x) &= \begin{cases} h_L & \text{if } x < x_D \\ h_R & \text{if } x \geq x_D \end{cases} \\ c_2(x) &= 0 \end{aligned} \right\} \quad (31)$$

The initial conditions are as those for a Riemann problem, with a discontinuity in free-surface h . The source terms include a variation in x along the channel and an exponential variation in time t , which is the most challenging part of this example. The constants are $h_L = 2$, $h_R = 1$, $x_D = 5$,

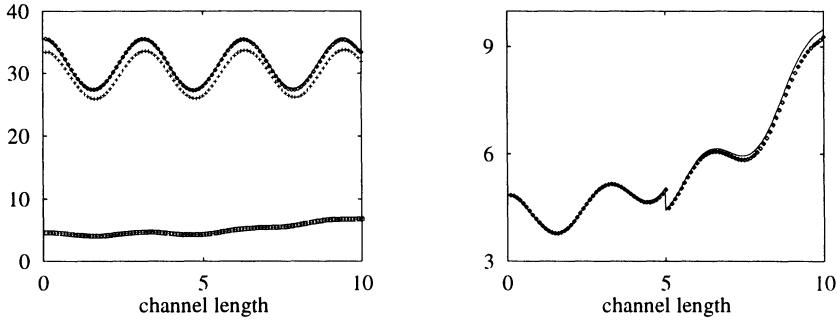


Figure 1. Exact (full line) and numerical (symbols) solutions for free-surface $h(x, t)$. Left plots correspond to sub-critical case at time $t = 2$. Right plots correspond to critical case at time $t = 1$.

$\alpha_1 = 2$, $\alpha_2 = -3$. The length of the domain is 10, the number of cells used is $M = 100$ and the CFL coefficient is $C_{cfl} = 1$. The entries in the coefficient matrix \mathbf{A} are chosen to reproduce two cases of interest, namely (a) *sub-critical case* and (b) *critical case*. For the first case we take $\bar{h} = 3$, $\bar{u} = -3.42$, with corresponding eigenvalues $\lambda_1 = -8.84$ and $\lambda_2 = 1.99$. For the second case we take $\bar{h} = 3$, $\bar{u} = -5.42$, with corresponding eigenvalues $\lambda_1 = -10.84$ and $\lambda_2 = 0$. Fig. 1 shows a comparison between exact solutions (line) and numerical solutions (symbols) for the sub-critical case (left) and the critical case (right). In the sub-critical case we show three numerical results; the top curve (line) is the exact solution; the lowest curve corresponds to a *centred scheme*; the next curve upwards corresponds to the *upwind* method of Ref. (Bermúdez and Vázquez-Cendón, 1994); the next curve up corresponds to the *new scheme* of this paper, the result of which is indistinguishable from the exact solution. For the critical case (right) we show the new scheme (symbol) and the exact solution. The performance of the other schemes was rather poor and results are omitted. It is worth

remarking that these test problems are very demanding on the ability of the numerical methods to capture correctly the time evolution of the solution.

6. Conclusions

Model hyperbolic systems with source terms have been constructed and exact solutions have been provided. In addition, a new numerical method for solving inhomogeneous systems has been presented. Numerical results, as compared with exact solutions, show that the method is superior to existing methods for dealing with source terms. Further work on extensions of the numerical method to non-linear systems is currently in progress.

Acknowledgements

The work of the second author was supported by DGYCIT
MAR97-1055-C02-01.

References

- Bermúdez A and Vázquez-Cendón M E (1994). Upwind Methods for Hyperbolic Conservation Laws with Source Terms. *Computers and Fluids* **23**, pp 8-20.
- Godlewski E and Raviart P A (1996). Numerical Approximation of Hyperbolic Systems of Conservation Laws. Springer.
- Greenberg J M and LeRoux A (1996). A Well-balanced Scheme for the Numerical Processing of Source Terms in Hyperbolic Equations. *SIAM Journal on Numerical Analysis* **33**, pp 1-19.
- Harten A, Engquist B, Osher S and Chakravarthy S R (1987). Uniformly High-order Accurate Essentially Non-oscillatory Schemes, III. *J. Comput. Phys.* **71**, pp 2-47.
- Kröner D (1997). Numerical Schemes for Conservation Laws. Wiley Teubner.
- LeVeque R J (1998). Balancing Source Terms and Flux Gradients in High-resolution Godunov Methods: the Quasi-steady Wave-propagation Algorithm. *J. Comput. Phys.* **146**, pp 346-365.
- LeVeque R J (1992). Numerical Methods for Conservation Laws. Birkhäuser Verlag.
- Godunov S K (1959). A Finite Difference Method for the Computation of Discontinuous Solutions of the Equations of Fluid Dynamics. *Mat. Sb.* **47**, pp 357-393.
- Roe P L (1987). Upwind Differencing Schemes for Hyperbolic Conservation Laws with Source Terms. Proc. First International Conference on Hyperbolic Problems. Carasso, Raviart and Serre (Editors). Springer, pp 41-51.
- Roe P L (1998). The Harten Memorial Lecture—New Applications of Upwinding. Numerical Methods for Wave Propagation, pp 1-31. Toro E F and Clarke J F (Editors). Kluwer Academic Publishers.
- Toro E F (1999). Riemann Solvers and Numerical Methods for Fluid Dynamics. Second Edition, Springer-Verlag.
- Toro E F and Clarke J F (Editors) (1998). Numerical Methods for Wave Propagation. Kluwer Academic Publishers.
- van Leer B (1984). On the Relation Between the Upwind-Differencing Schemes of Godunov, Engquist-Osher and Roe. *SIAM J. Sci. Stat. Comput.* **5**, pp 1-20.
- Vázquez-Cendón M E (1999). Improved Treatment of Source Terms in Upwind Schemes for the Shallow Water Equations with Irregular Geometry. *J. Comput. Phys.* **148**, pp 497-526.

A GODUNOV-TYPE METHOD FOR CAPTURING WATER WAVES

E. H. VAN BRUMMELEN

CWI

P.O. Box 94079

1090 GB Amsterdam

The Netherlands

Email: Harald.van.Brummelen@cwi.nl

AND

B. KOREN

CWI

P.O. Box 94079

1090 GB Amsterdam

The Netherlands

Email: Barry.Koren@cwi.nl

Abstract.

In spite of the absence of shock waves in most hydrodynamic applications, sufficient reason remains to employ Godunov-type schemes in this field. In the instance of two-phase flow, the shock capturing ability of these schemes may serve to maintain robustness and accuracy at the interface. Moreover, approximate Riemann solvers have greatly relieved the initial drawback of computational expensiveness of Godunov-type schemes. In the present work we develop an Osher-type approximate Riemann solver for application in hydrodynamics. Actual computations are left to future research.

1. Introduction

The advantages of Godunov-type schemes (Godunov, 1959) in hydrodynamic flow computations are not as widely appreciated as in gas dynamics applications. Admittedly, the absence of supersonic speeds and hence

shock waves in incompressible flow (the prevailing fluid model in hydrodynamics) reduces the necessity of advanced shock capturing schemes. Nevertheless, many reasons remain to apply Godunov-type schemes in hydrodynamics: Firstly, these schemes have favorable robustness properties due to the inherent upwind treatment of the flow. Secondly, they feature a consistent treatment of boundary conditions. Thirdly, (higher-order accurate) Godunov-type schemes display low dissipative errors, which is imperative for an accurate resolution of boundary layers in viscous flow. Finally, the implementation of these schemes in conjunction with higher-order limited interpolation methods, to maintain accuracy and prevent oscillations in regions where large gradients occur (see, e.g., (Sweby, 1984; Spekreijse, 1987)), is relatively straightforward.

In addition, Godunov-type schemes can be particularly useful in hydrodynamics in case of two-phase flows, e.g., flows suffering cavitation and free surface flows. In these situations, an interface exists between the primary phase (water) and the secondary phase (air, damp, etc.) and fluid properties may vary discontinuously across the interface. In our opinion, the ability of Godunov-type schemes to capture discontinuities is then very useful to maintain robustness and accuracy at the interface. Examples of such interface capturing can be found in, for instance, (Mulder *et al.*, 1992; Chang *et al.*, 1996; Kelecy and Pletcher, 1997).

A disadvantage of the method originally proposed by Godunov is that it requires the solution of an associated Riemann problem with each flux evaluation. In practice, many such evaluations are performed during an actual computation. Consequently, the method is notorious for its high computational costs. To relieve this problem, several approaches have been suggested to reduce the computational costs of the flux evaluations involved, by approximating the Riemann solution. Examples of such approximate Riemann solvers are the flux difference splitting schemes (such as Roe's (Roe, 1981) and Osher's (Osher and Solomon, 1982)).

In the present work we develop an Osher-type flux-difference splitting scheme for the approximate solution of the Riemann problem and we investigate its application in hydrodynamics. Details are presented for the Euler equations for four types of fluids that are commonly used to model the behavior of water, viz., a genuinely compressible fluid, an artificially compressible fluid, a genuinely incompressible fluid, and a two-phase fluid. As a preliminary, we examine the Riemann problem. Next, we give an outline of Osher's approximate Riemann solver. Analysis shows that Osher's scheme suffers loss of accuracy in the presence of centered shock waves and therefore a modified scheme is proposed. Finally, we present the specifics for the aforementioned hydrodynamic applications. Actual computations are deferred to future research.

2. Riemann Problem

In this section we investigate the Riemann Problem:

Definition 1 Let $\mathbf{q} \in \mathbb{R}^n = (q_1, \dots, q_n)^T$, $(x, t) \in \mathbb{R} \times \mathbb{R}^+$ and $\mathbf{f} \in C^1(\mathbb{R}^n, \mathbb{R}^n)$. Consider the Cauchy problem

$$\frac{\partial \mathbf{q}}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{q})}{\partial x} = 0, \quad \forall x \in \mathbb{R}, t \in \mathbb{R}^+, \quad (1a)$$

subject to the initial condition

$$\mathbf{q}(x, 0) = \begin{cases} \mathbf{q}_L, & \text{if } x < 0, \\ \mathbf{q}_R, & \text{if } x > 0, \end{cases} \quad (1b)$$

with \mathbf{q}_L and \mathbf{q}_R constant. The initial value problem (1a) and the initial condition (1b) define the Riemann problem.

First, an introductory analysis is presented. Subsequently, we obtain the general solution to (1).

2.1. PRELIMINARY ANALYSIS

Let $\mathbf{A}(\mathbf{q})$ denote the Jacobian of $\mathbf{f}(\mathbf{q})$, $\mathbf{A}(\mathbf{q}) = \partial_{\mathbf{q}} \mathbf{f}(\mathbf{q})$, and let $\lambda_k(\mathbf{q})$, $k = 1, 2, \dots, n$, $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$, be its eigenvalues and $\mathbf{r}_k(\mathbf{q})$ the corresponding eigenvectors. Equation (1a) constitutes a hyperbolic system if the eigenvalues $\lambda_k(\mathbf{q})$ are real and nonzero. Then, the matrix $\mathbf{A}(\mathbf{q})$ can be decomposed with respect to a basis of its eigenvectors:

$$\mathbf{A}(\mathbf{q}) = \mathbf{R}(\mathbf{q}) \cdot \Lambda(\mathbf{q}) \cdot \mathbf{R}(\mathbf{q})^{-1}, \quad (2)$$

where $\Lambda(\mathbf{q}) = \text{diag}(\lambda_1(\mathbf{q}), \dots, \lambda_n(\mathbf{q}))$ and $\mathbf{R}(\mathbf{q}) = (\mathbf{r}_1(\mathbf{q}), \dots, \mathbf{r}_n(\mathbf{q}))$ contains the eigenvectors. From (Lax, 1957) we adopt the following classification of the eigenpairs $(\lambda_k(\mathbf{q}), \mathbf{r}_k(\mathbf{q}))$:

Definition 2 Consider the matrix $\mathbf{A}(\mathbf{q}) \in \mathbb{R}^{n \times n}$. Let $\lambda_k(\mathbf{q})$, $k = 1, 2, \dots, n$, be its eigenvalues and $\mathbf{r}_k(\mathbf{q})$ the corresponding eigenvectors. An eigenvalue $\lambda_k(\mathbf{q})$ and an eigenvector $\mathbf{r}_k(\mathbf{q})$ are called genuinely nonlinear on a subdomain $\Omega \subseteq \mathbb{R}^n$ if

$$\partial_{\mathbf{q}} \lambda_k(\mathbf{q}) \cdot \mathbf{r}_k(\mathbf{q}) \neq 0, \quad \forall \mathbf{q} \in \Omega. \quad (3)$$

An eigenvalue $\lambda_k(\mathbf{q})$ and an eigenvector $\mathbf{r}_k(\mathbf{q})$ are said to be linearly degenerate on Ω if

$$\partial_{\mathbf{q}} \lambda_k(\mathbf{q}) \cdot \mathbf{r}_k(\mathbf{q}) = 0, \quad \forall \mathbf{q} \in \Omega. \quad (4)$$

The eigenvalues that are genuinely nonlinear for all $\mathbf{q} \in \mathbb{R}^n$ are related to rarefaction waves and shocks in the solution of the Riemann problem. The eigenvalues that are linearly degenerate on \mathbb{R}^n correspond to contact discontinuities in the solution. More complex contact phenomena can occur for eigenvalues that are neither genuinely nonlinear nor linearly degenerate on \mathbb{R}^n , see, e.g., (LeVeque, 1990, pages 48–50).

With each of the eigenpairs $(\lambda_k(\mathbf{q}), \mathbf{r}_k(\mathbf{q}))$ we associate two paths in state space. Firstly, the k -shock path:

Definition 3 Consider hyperbolic system (1a). The k -shock path through \mathbf{q}_L is the set

$$\mathcal{S}_k(\mathbf{q}_L) = \{\mathbf{q} \in \mathbb{R}^n \mid s(\mathbf{q}; \mathbf{q}_L)(\mathbf{q} - \mathbf{q}_L) = \mathbf{f}(\mathbf{q}) - \mathbf{f}(\mathbf{q}_L)\}, \quad (5)$$

where $s(\mathbf{q}; \mathbf{q}_L)$ is referred to as the shock speed.

Secondly, we distinguish the k -path:

Definition 4 Consider the hyperbolic system (1a). The k -path through \mathbf{q}_L is the set

$$\mathcal{R}_k(\mathbf{q}_L) = \{\mathbf{q} \in \mathbb{R}^n \mid \mathbf{q} = \mathbf{h}(\xi), \xi \in \mathbb{R}\}, \quad (6)$$

with $\mathbf{h}(\xi)$ the solution of the ordinary differential equation

$$\begin{aligned} \frac{\partial \mathbf{h}(\xi)}{\partial \xi} &= \mathbf{r}_k(\mathbf{h}(\xi)), & \xi \in \mathbb{R}, \\ \mathbf{h}(\xi_L) &= \mathbf{q}_L, \end{aligned} \quad (7)$$

for some $\xi_L \in \mathbb{R}$.

Furthermore, to each k -path corresponds a set of functions which are invariant on \mathcal{R}_k :

Definition 5 Consider the hyperbolic system (1a). Let $\mathbf{r}_k(\mathbf{q})$ denote the k^{th} eigenvector of the Jacobian $\mathbf{A}(\mathbf{q}) = \partial \mathbf{f}(\mathbf{q}) / \partial \mathbf{q}$. A k -Riemann invariant is any function $\psi_k \in C^1(\mathbb{R}^n, \mathbb{R})$ satisfying

$$\partial_{\mathbf{q}} \psi_k(\mathbf{q}) \cdot \mathbf{r}_k(\mathbf{q}) = 0, \quad \forall \mathbf{q} \in \mathbb{R}^n. \quad (8)$$

There are at most $n - 1$ such k -Riemann invariants with linearly independent gradients in \mathbb{R}^n . Observe that for a linearly degenerate eigenpair $(\lambda_k(\mathbf{q}), \mathbf{r}_k(\mathbf{q}))$ the eigenvalue $\lambda_k(\mathbf{q})$ is a k -Riemann invariant.

2.2. SOLUTION

The general solution to (1) consists of regions in the (x, t) -domain where the solution is constant, separated by simple waves, contact discontinuities

and shock waves. Before constructing the general solution, we first obtain the (weak) solution to (1) in the case that it contains only one of the aforementioned contact phenomena.

We establish that the (weak) solution to the Riemann problem can generally be written in similarity form (see, e.g., (Smoller, 1983)):

Theorem 1 *Suppose a unique solution $\mathbf{q}(x, t)$ to the Riemann Problem (1) exists. Then $\mathbf{q}(x, t)$ can be written in similarity form $\mathbf{q}(x, t) = \mathbf{h}(x/t)$.*

Proof: Assume $\mathbf{q}(x, t)$ solves (1). Then for all $\alpha \in \mathbb{R}$, $\mathbf{q}(\alpha x, \alpha t)$ is also a solution:

$$\frac{\partial \mathbf{q}(\alpha x, \alpha t)}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{q}(\alpha x, \alpha t))}{\partial x} = \alpha [D_2 \mathbf{q}(\alpha x, \alpha t) + \mathbf{A}(\mathbf{q}(\alpha x, \alpha t)) \cdot D_1 \mathbf{q}(\alpha x, \alpha t)] = 0, \quad (9)$$

where D_l denotes differentiation with respect to the l^{th} function-argument. Because the solution is unique by assumption, $\mathbf{q}(x, t) = \mathbf{q}(\alpha x, \alpha t)$. Hence, $\mathbf{q}(x, t) = \mathbf{h}(x/t)$. \square

A (classical) simple wave solution of (1) exists if $\lambda_k(\mathbf{q})$ is a genuinely nonlinear eigenvalue, $\lambda_k(\mathbf{q}_L) < \lambda_k(\mathbf{q}_R)$ and \mathbf{q}_R is on the k -path through \mathbf{q}_L . Note that this implies that the k -Riemann invariants are equal for \mathbf{q}_L and \mathbf{q}_R , i.e., $\psi_k^m(\mathbf{q}_L) = \psi_k^m(\mathbf{q}_R)$, for $m \neq k$, $m = 1, \dots, n$. Assuming that the genuinely nonlinear eigenvector in (7) is normalized such that

$$\partial_{\mathbf{q}} \lambda_k(\mathbf{q}) \cdot \mathbf{r}_k(\mathbf{q}) = 1, \quad \forall \mathbf{q} \in \mathbb{R}^n, \quad (10)$$

we find that $\mathbf{q}(x, t) = \mathbf{h}(x/t)$ according to (7) is the similarity solution in the simple wave region $\lambda_k(\mathbf{q}_L) < x/t < \lambda_k(\mathbf{q}_R)$ (see, e.g., (Smoller, 1983; Lax, 1973)):

Theorem 2 *Suppose $\mathbf{h} \in C^1(\mathbb{R}, \mathbb{R}^n)$ solves (7), with $\mathbf{r}_k(\mathbf{q})$ normalized according to (10), and $\mathbf{q}_R \in \mathcal{R}_k(\mathbf{q}_L)$. Then $\mathbf{q}(x, t) = \mathbf{h}(x/t)$ is the similarity solution of (1) in the simple-wave region $\lambda_k(\mathbf{q}_L) < x/t < \lambda_k(\mathbf{q}_R)$.*

Proof: We will only show that $\mathbf{q}(x, t) = \mathbf{h}(x/t)$ solves (1a). Inserting $\mathbf{q}(x, t) = \mathbf{h}(x/t)$ in (1a), one obtains

$$\frac{\partial \mathbf{h}(x/t)}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{h}(x/t))}{\partial x} = \frac{1}{t} \left(\mathbf{A}(\mathbf{h}(x/t)) - \mathbf{I} \frac{x}{t} \right) \cdot D\mathbf{h}(x/t), \quad (11)$$

where \mathbf{I} stands for the $\mathbb{R}^{n \times n}$ identity matrix and D denotes differentiation with respect to the function-argument. The right-hand side term of (11) vanishes if $x/t = \lambda_k(\mathbf{h}(x/t))$ and $D\mathbf{h}(x/t) = \mathbf{r}_k(\mathbf{h}(x/t))$. The latter trivially follows from (7), the former from (10). Hence, $\mathbf{h}(x/t)$ solves (1a). \square

Outside the wave region the solution remains unchanged. The Riemann

solution $\mathbf{q}(x, t)$ in the case of a k -rarefaction wave is now simply composed of the solutions on the separate regions:

$$\mathbf{q}(x, t) = \begin{cases} \mathbf{q}_L, & \text{if } x/t < \lambda(\mathbf{q}_L), \\ \mathbf{h}(x/t), & \text{if } \lambda(\mathbf{q}_L) < x/t < \lambda(\mathbf{q}_R), \\ \mathbf{q}_R, & \text{if } x/t > \lambda(\mathbf{q}_R). \end{cases} \quad (12)$$

Next, we derive the (weak) Riemann solution in the instance that it contains a single contact discontinuity. The states \mathbf{q}_L and \mathbf{q}_R are connected by a k -contact discontinuity if $(\lambda_k(\mathbf{q}), \mathbf{r}_k(\mathbf{q}))$ is a linearly degenerate eigenpair and \mathbf{q}_R is on the k -path through \mathbf{q}_L . Then, by (4), $\lambda_k(\mathbf{q}_R) = \lambda_k(\mathbf{q}_L)$. The solution to the Riemann problem is now obtained immediately from (12):

$$\mathbf{q}(x, t) = \begin{cases} \mathbf{q}_L, & \text{if } x/t < \lambda(\mathbf{q}_L) = \lambda(\mathbf{q}_R), \\ \mathbf{q}_R, & \text{if } x/t > \lambda(\mathbf{q}_L) = \lambda(\mathbf{q}_R). \end{cases} \quad (13)$$

However, because (13) is discontinuous at $x/t = \lambda(\mathbf{q}_L) = \lambda(\mathbf{q}_R)$, it must be verified that (13) satisfies the weak form of (1a):

$$\oint_{\mathcal{C}} (\mathbf{q} n_t + \mathbf{f}(\mathbf{q}) n_x) d\mathcal{C} = 0. \quad (14)$$

Here \mathcal{C} is any closed curve in (x, t) and $\mathbf{n} = (n_t, n_x)$ denotes the outward pointing unit normal on \mathcal{C} . It can easily be shown that (14) does indeed hold for (13), so that (13) is a valid weak solution.

Finally, we consider the solution to (1) when it comprises a single shock. A shock occurs if $\lambda_k(\mathbf{q})$ is a genuinely nonlinear eigenvalue, $\lambda_k(\mathbf{q}_L) > \lambda_k(\mathbf{q}_R)$ and \mathbf{q}_R is on the k -shock path through \mathbf{q}_L . A solution of the form (12) is then necessarily multiple-valued and must therefore be discarded. Instead, the weak solution reads

$$\mathbf{q}(x, t) = \begin{cases} \mathbf{q}_L, & \text{if } x/t < s(\mathbf{q}_L; \mathbf{q}_R), \\ \mathbf{q}_R, & \text{if } x/t > s(\mathbf{q}_L; \mathbf{q}_R), \end{cases} \quad (15)$$

where $s(\mathbf{q}_L; \mathbf{q}_R)$ denotes the shock speed, determined by the Rankine-Hugoniot relation

$$s(\mathbf{q}_L; \mathbf{q}_R)(\mathbf{q}_L - \mathbf{q}_R) = \mathbf{f}(\mathbf{q}_L) - \mathbf{f}(\mathbf{q}_R). \quad (16)$$

Expression (16) is in fact equivalent to (14). Hence, (15) is a valid weak solution of (1).

The general solution to the Riemann problem consists of $n+1$ (possibly empty) regions Ω_l where the solution is constant, separated by simple waves, contact discontinuities and shock waves. Define $\mathbf{q}_0 = \mathbf{q}_L$, $\mathbf{q}_1 = \mathbf{q}_R$ and let $\mathbf{q}_{l/n}, l = 0, \dots, n$, be the solution in Ω_l . Assuming that $\mathbf{q}_{(l-1)/n}$ is connected

to $\mathbf{q}_{l/n}$ by a simple wave, we denote by $\mathbf{h}_l(x/t)$ the similarity solution in the wave region. Conversely, if $\mathbf{q}_{(l-1)/n}$ is connected to $\mathbf{q}_{l/n}$ by a shock wave, we designate s_l the appropriate shock speed. Then, in succinct form:

$$\mathbf{q}(x, t) = \begin{cases} \mathbf{q}_0, & \text{if } x/t < \sigma_0^+, \\ \mathbf{q}_{l/n}, & \text{if } \sigma_l^- < x/t < \sigma_l^+, \quad l = 1, \dots, n-1, \\ \mathbf{h}_l(x/t), & \text{if } \sigma_{l-1}^+ < x/t < \sigma_l^-, \quad l = 1, \dots, n-1, \\ \mathbf{q}_1, & \text{if } x/t > \sigma_n^-, \end{cases} \quad (17a)$$

where σ_l^\pm denotes the contact speed

$$\sigma_l^\pm = \begin{cases} \lambda_{l+(1\pm 1)/2}(\mathbf{q}_{l/n}) & \text{if } \pm \lambda_{l+(1\pm 1)/2}(\mathbf{q}_{l/n}) < \pm \lambda_{l+(1\pm 1)/2}(\mathbf{q}_{(l\pm 1)/n}), \\ s_{l+(1\pm 1)/2} & \text{otherwise.} \end{cases} \quad (17b)$$

The general solution (17) is schematically depicted in Figure 1. The figure illustrates the contiguity of regions connected by shock waves and contact discontinuities, for instance, $\Omega_{(l-1)/n}$ and $\Omega_{l/n}$, and the separation of regions connected by rarefaction waves, e.g., $\Omega_{l/n}$ and $\Omega_{(l+1)/n}$.

As a side-note, we mention that for general $\mathbf{f}(\mathbf{q})$ and sufficiently large $\|\mathbf{q}_L - \mathbf{q}_R\|$, a solution to (1) can be non-existent (see, e.g., (Smoller, 1983)).

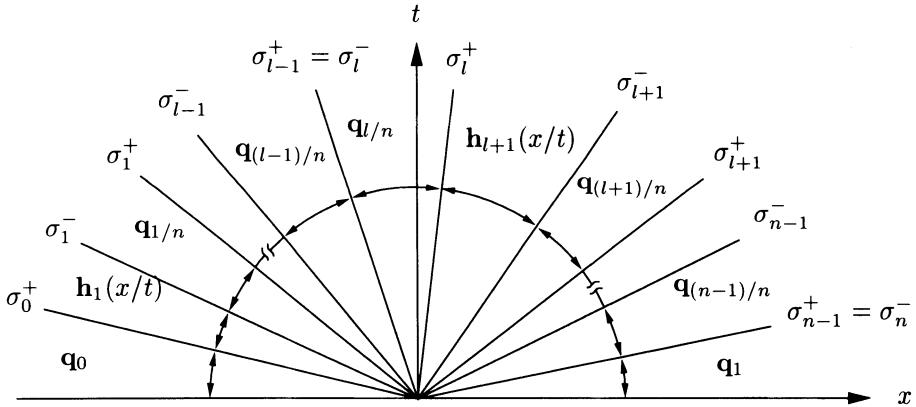


Figure 1. General solution to the Riemann problem

3. Approximate Riemann Solution

In the previous section we established that the solution to the Riemann problem can generally be written in similarity form, $\mathbf{h}(x/t)$. Denoting by

$\mathbf{h}(x/t; \mathbf{q}_L, \mathbf{q}_R)$ the similarity solution for given \mathbf{q}_L and \mathbf{q}_R , $\mathbf{f}(\mathbf{h}(0; \mathbf{q}_L, \mathbf{q}_R))$ expresses the corresponding centered flux, next indicated by $\mathbf{f}(\mathbf{q}_L, \mathbf{q}_R)$. This flux is of particular importance in computational applications: following Godunov's approach, it can be interpreted as the flux between two adjacent cells in the discretized domain. Unfortunately, solving the Riemann problem exactly is computationally expensive and it is therefore necessary to revert to approximate solution techniques.

In this section, we investigate Osher's approximate Riemann solver and a modified Osher-type scheme. We will first present a general outline of the Osher scheme. Subsequently, the approximate Riemann solution employed in Osher's scheme is examined and the computed flux approximation is compared to the exact solution. Finally, we shall propose the modified scheme, based on the preceding analysis.

3.1. OSHER'S SCHEME

In the scheme developed by Osher (Osher and Solomon, 1982; Osher and Chakravarthy, 1983), the centered flux, $\mathbf{f}(\mathbf{q}_L, \mathbf{q}_R) = \mathbf{f}(\mathbf{h}(0; \mathbf{q}_L, \mathbf{q}_R))$, is approximated by:

$$\tilde{\mathbf{f}}(\mathbf{q}_L, \mathbf{q}_R) = \frac{1}{2}\mathbf{f}(\mathbf{q}_L) + \frac{1}{2}\mathbf{f}(\mathbf{q}_R) - \frac{1}{2} \int_{\mathbf{q}_L}^{\mathbf{q}_R} |\mathbf{A}(\mathbf{w})| \cdot d\mathbf{w}, \quad (18)$$

with the absolute value of the Jacobian matrix defined by $|\mathbf{A}(\mathbf{q})| \equiv \mathbf{R}(\mathbf{q}) \cdot |\Lambda(\mathbf{q})| \cdot \mathbf{R}(\mathbf{q})^{-1}$. Here, $|\Lambda(\mathbf{q})| = \text{diag}(|\lambda_1(\mathbf{q})|, \dots, |\lambda_n(\mathbf{q})|)$. Clearly, the integral term is the upwind contribution to the centered flux approximation.

The integral in (18) is evaluated along a path $\Gamma = \{\mathbf{q}(s) : 0 \leq s \leq 1\} \subset \mathbb{R}^n$ in state space, satisfying $\mathbf{q}(0) = \tilde{\mathbf{q}}_0$ and $\mathbf{q}(1) = \tilde{\mathbf{q}}_1$, with $\tilde{\mathbf{q}}_0 = \mathbf{q}_L$ and $\tilde{\mathbf{q}}_1 = \mathbf{q}_R$ or vice versa. This path is composed of sub-paths Γ_l , $l = 1, 2, \dots, n$, where each of the sub-paths connects two adjacent states $\tilde{\mathbf{q}}_{(l-1)/n}$ and $\tilde{\mathbf{q}}_{l/n}$. Moreover, Γ_l is tangential to an eigenvector $\mathbf{r}_{k(l)}$, where $k : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$ is a bijective mapping. It should be appreciated here that Γ_l is thus a section of the $k(l)$ -path through $\tilde{\mathbf{q}}_{(l-1)/n}$, connecting $\tilde{\mathbf{q}}_{(l-1)/n}$ and $\tilde{\mathbf{q}}_{l/n}$. Usual choices for the ordering of the sub-paths are the O-variant $k(l) = n - l$ and the P-variant $k(l) = l$.

Rewriting the integral in (18) as a summation of contributions of the integral over each of the sub-paths,

$$\begin{aligned} \int_{\mathbf{q}_L}^{\mathbf{q}_R} |\mathbf{A}(\mathbf{w})| \cdot d\mathbf{w} &= \sum_{l=1}^n \int_{\Gamma_l} |\mathbf{A}(\mathbf{w}(\xi))| \cdot \mathbf{r}_{k(l)}(\mathbf{w}(\xi)) d\xi = \\ &\quad \sum_{l=1}^n \int_{\Gamma_l} \text{sign}(\lambda_{k(l)}(\mathbf{w})) \mathbf{A}(\mathbf{w}) \cdot d\mathbf{w}. \end{aligned} \quad (19)$$

Obviously, if $\lambda_{k(l)}$ does not change sign along Γ_l , then the sub-integral can be evaluated to $[\mathbf{f}(\tilde{\mathbf{q}}_{l/n}) - \mathbf{f}(\tilde{\mathbf{q}}_{(l-1)/n})] \operatorname{sign}(\lambda_{k(l)})$. Then, by (4), if $\lambda_{k(l)} = \lambda_{k(l+1)} = \dots = \lambda_{k(l+\mu)}$ is a linearly degenerate eigenvalue, the sum in (19) concatenates and we simply obtain

$$\sum_{i=0}^{\mu} \int_{\Gamma_{l+i}} |\mathbf{A}(\mathbf{w})| \cdot d\mathbf{w} = \operatorname{sign}(\lambda_{k(l)}(\mathbf{q}_{l/n})) [\mathbf{f}(\mathbf{q}_{l/n}) - \mathbf{f}(\mathbf{q}_{(l+\mu)/n})]. \quad (20)$$

Hence, the intermediate stages $\tilde{\mathbf{q}}_{(l+i)/n}$, $i = 1, 2, \dots, \mu - 1$ are of no consequence and can be eliminated from the composed path Γ .

As a result of the choice of the sub-paths Γ_l , the intermediate $\tilde{\mathbf{q}}_{l/n}$, $l = 1, 2, \dots, n-1$ can be conveniently determined by means of the Riemann invariants: Because the sub-path $\Gamma_l \subset \mathcal{R}_{k(l)}(\tilde{\mathbf{q}}_{(l-1)/n})$,

$$\psi_{k(l)}^m(\tilde{\mathbf{q}}_{(l-1)/n}) = \psi_{k(l)}^m(\tilde{\mathbf{q}}_{l/n}), \quad l, m = 1, 2, \dots, n, \quad m \neq k(l), \quad (21)$$

see section 2.1. If it is assumed that the k -Riemann invariants in (21) have linearly independent gradients, then by the implicit function theorem, (21) constitutes a solvable system of equations from which the $\tilde{\mathbf{q}}_{l/n}$, $l = 1, 2, \dots, n$ can be extracted. In many practical cases the intermediate stages can then be solved explicitly from (21). Once the intermediate states $\tilde{\mathbf{q}}_{l/n}$ have been obtained, the flux approximation $\tilde{\mathbf{f}}(\mathbf{q}_L, \mathbf{q}_R)$ can be computed using (18), (19).

3.2. ACCURACY

The flux computed by means of the Osher scheme, $\tilde{\mathbf{f}}(\mathbf{q}_L, \mathbf{q}_R)$, relies on an approximate solution of the Riemann problem. Because the approximation can again be written in similarity form, it is useful to introduce the notation $\tilde{\mathbf{f}}(\mathbf{q}_L, \mathbf{q}_R) = \mathbf{f}(\tilde{\mathbf{h}}(0; \mathbf{q}_L, \mathbf{q}_R))$, where $\tilde{\mathbf{h}}(x/t; \mathbf{q}_L, \mathbf{q}_R)$ stands for the approximate similarity solution. In this section we investigate the accuracy of the approximate similarity solution and of the corresponding centered flux approximation.

To evaluate the accuracy of the approximate solution, we examine the inherent representation of simple waves, contact discontinuities and shock waves. In section 3.1 it was emphasized that the sub-paths, Γ_l , in Osher's scheme are actually sections of $k(l)$ -paths. Referring to section 2.2, it follows that the intermediate states $\tilde{\mathbf{q}}_{l/n}$, $l = 0, \dots, n$, in the approximate solution are connected by simple-waves only. Clearly, this representation is correct for simple waves and contact discontinuities. However, shock waves in the actual solution are then replaced by so-called *overturned simple waves*, see (van Leer, 1984). We will now show that this representation is accurate for weak shocks. From (Smoller, 1983) we adopt:

Lemma 1 Suppose \mathbf{q}_L and \mathbf{q}_R are connected by a weak k -shock with shock strength ϵ , i.e., $\mathbf{q}_R \in \mathcal{S}_k(\mathbf{q}_L)$ and $\lambda_k(\mathbf{q}_L) = \lambda_k(\mathbf{q}_R) + \epsilon$, with ϵ a small positive number. Then the change in a k -Riemann invariant across the k -shock is of order $\mathcal{O}(\epsilon^3)$.

Proof: Proof is omitted here, but can be found in (Smoller, 1983, pages 326–333). \square

Then, we obtain:

Theorem 3 Suppose $\mathbf{q}_R \in \mathcal{S}_k(\mathbf{q}_L)$ and $\lambda_k(\mathbf{q}_L) = \lambda_k(\mathbf{q}_R) + \epsilon$. Then a $\tilde{\mathbf{q}}_R \in \mathcal{R}_k(\mathbf{q}_L)$ exists such that $\lambda_k(\tilde{\mathbf{q}}_R) = \lambda_k(\mathbf{q}_R)$ and $|\tilde{\mathbf{q}}_R - \mathbf{q}_R|$ is of order $\mathcal{O}(\epsilon^3)$.

Proof: By definition 5, $\psi_k^m(\mathbf{q}_L) = \psi_k^m(\tilde{\mathbf{q}}_R)$, $k = 1, 2, \dots, n$, $k \neq m$. Then, by lemma 1,

$$\psi_k^m(\tilde{\mathbf{q}}_R) = \psi_k^m(\mathbf{q}_R) + \mathcal{O}(\epsilon^3). \quad (22)$$

System (22) can be augmented with $\lambda_k(\tilde{\mathbf{q}}_R) = \lambda_k(\mathbf{q}_R)$ to obtain n equations for $\tilde{\mathbf{q}}_R$. Because $\text{rank}(\partial_{\mathbf{q}}\psi_1^1, \dots, \partial_{\mathbf{q}}\psi_n^n) = n - 1$ and $\partial_{\mathbf{q}}\lambda_k \in (\partial_{\mathbf{q}}\psi_1^1, \dots, \partial_{\mathbf{q}}\psi_n^n)^\perp$, it follows that $\det(\partial_{\mathbf{q}}\psi_1^1, \dots, \partial_{\mathbf{q}}\psi_n^n, \partial_{\mathbf{q}}\lambda_k) \neq 0$. The result now simply follows by a Taylor expansion around \mathbf{q}_R of the terms in $\tilde{\mathbf{q}}_R$ of the augmented system. \square

From Theorem 3 it may be inferred that the intermediate states obtained by a rarefaction-waves-only approximation are $\mathcal{O}(\epsilon_{\max}^3)$ accurate, with

$$\epsilon_{\max} = \max_{l=1\dots n} (\lambda_l(\mathbf{q}_{(l-1)/n}) - \lambda_l(\mathbf{q}_{l/n}), 0) \quad (23)$$

the strength of the strongest shock.

Although the computed intermediate states are accurate even in the presence of (weak) shocks, the flux approximation $\tilde{\mathbf{f}}(\mathbf{q}_L, \mathbf{q}_R)$ is not necessarily so. By (19), if $\tilde{\mathbf{q}}_R \in \mathcal{R}_k(\mathbf{q}_L)$ and $\lambda_k(\mathbf{q}_L) > 0 > \lambda_k(\tilde{\mathbf{q}}_R)$,

$$\tilde{\mathbf{f}}(\mathbf{q}_L, \tilde{\mathbf{q}}_R) = \mathbf{f}(\mathbf{q}_L) + \mathbf{f}(\tilde{\mathbf{q}}_R) - \mathbf{f}(\mathbf{q}^*), \quad (24)$$

with $\mathbf{q}^* \in \mathcal{R}_k(\mathbf{q}_L)$ such that $\lambda_k(\mathbf{q}^*) = 0$. In contrast, the actual flux corresponding to the k -shock reads $\mathbf{f}(\mathbf{q}_L)$ if $s(\mathbf{q}_R; \mathbf{q}_L) > 0$ and $\mathbf{f}(\mathbf{q}_R)$ if $s(\mathbf{q}_R; \mathbf{q}_L) < 0$. Consequently, the error in the approximate flux in the case of a centered shock with strength ϵ can be of $\mathcal{O}(\epsilon)$. An instance of such failure of Osher's flux is discussed in (van Leer, 1984).

3.3. MODIFIED OSHER SCHEME

In view of the above, a modification of the scheme is advocated. The rarefaction-waves-only approximation of the similarity solution is maintained. However, the centered flux approximation is obtained differently, to avoid loss of accuracy due to centered shock waves.

We propose to extract the intermediate states in the approximate solution to the Riemann problem from

$$\psi_l^m(\tilde{\mathbf{q}}_{(l-1)/n}) = \psi_l^m(\tilde{\mathbf{q}}_{l/n}), \quad l, m = 1, 2, \dots, n, \quad m \neq l, \quad (25)$$

with $\tilde{\mathbf{q}}_0 = \mathbf{q}_L$ and $\tilde{\mathbf{q}}_1 = \mathbf{q}_R$. This is in fact equivalent to (21) with a presumed P-variant ordering of the sub-paths. Next, approximate contact speeds $\tilde{\sigma}_l^\pm$ are obtained:

$$\tilde{\sigma}_l^\pm = \begin{cases} \lambda_{l+(1\pm1)/2}(\tilde{\mathbf{q}}_{l/n}) & \text{if } \pm \lambda_{l+(1\pm1)/2}(\tilde{\mathbf{q}}_{l/n}) < \pm \lambda_{l+(1\pm1)/2}(\tilde{\mathbf{q}}_{(l\pm1)/n}), \\ \tilde{s}_{l+(1\pm1)/2} & \text{otherwise,} \end{cases} \quad (26a)$$

with

$$\tilde{s}_{l+(1\pm1)/2} = \frac{1}{2}\lambda_{l+(1\pm1)/2}(\tilde{\mathbf{q}}_{l/n}) + \frac{1}{2}\lambda_{l+(1\pm1)/2}(\tilde{\mathbf{q}}_{(l\pm1)/n}). \quad (26b)$$

Estimate (26b) of the shock speed is justified by the following theorem, taken from (Smoller, 1983):

Theorem 4 Suppose $\mathbf{q}_R \in \mathcal{S}_k(\mathbf{q}_L)$ and $\lambda_k(\mathbf{q}_L) = \lambda_k(\mathbf{q}_R) + \epsilon$, $\epsilon > 0$. Then the speed of the k -shock connecting \mathbf{q}_L and \mathbf{q}_R satisfies $s(\mathbf{q}_L; \mathbf{q}_R) = \frac{1}{2}\lambda_k(\mathbf{q}_L) + \frac{1}{2}\lambda_k(\mathbf{q}_R) + \mathcal{O}(\epsilon^2)$.

Proof: Proof can be found in (Smoller, 1983, pages 326–333). \square

Once the intermediate states and contact speeds have been established, the approximate Riemann solution can be constructed in a manner similar to (17a). However, considering that our purpose is to compute an approximation to the centered flux, we only need to obtain the central part of the approximate solution:

$$\tilde{\mathbf{h}}(0; \mathbf{q}_L, \mathbf{q}_R) = \begin{cases} \tilde{\mathbf{q}}_0, & \text{if } \tilde{\sigma}_0^+ > 0, \\ \tilde{\mathbf{q}}_{l/n}, & \text{if } \tilde{\sigma}_l^- < 0 < \tilde{\sigma}_l^+, \quad l \in \{1, \dots, n-1\}, \\ \tilde{\mathbf{q}}^*, & \text{if } \tilde{\sigma}_{l-1}^+ < 0 < \tilde{\sigma}_l^-, \quad l \in \{1, \dots, n-1\}, \\ \tilde{\mathbf{q}}_1, & \text{if } \sigma_n^- < 0, \end{cases} \quad (27)$$

with $\tilde{\mathbf{q}}^* \in \mathcal{R}_l(\tilde{\mathbf{q}}_{(l-1)/n})$ such that $\lambda_l(\tilde{\mathbf{q}}^*) = 0$ in case of a centered rarefaction wave. The centered flux approximation is now simply $\tilde{\mathbf{f}}(\mathbf{q}_L, \mathbf{q}_R) = \mathbf{f}(\tilde{\mathbf{h}}(0; \mathbf{q}_L, \mathbf{q}_R))$.

4. Applications in Hydrodynamics

In the previous section we presented a flux-difference splitting scheme that gives an accurate approximation of the centered flux in the Riemann problem, even in the presence of (weak) centered shock waves. A prerequisite

for the flux evaluation is the derivation of the intermediate states $\tilde{\mathbf{q}}_{l/n}$, $l = 1, \dots, n$. Once these states have been obtained, the flux calculation proceeds via straightforward operations.

In this section we derive the intermediate states for the one-dimensional Euler equations for three types of fluids that are commonly used to model the behavior of water. These fluids are, successively, a genuinely compressible fluid, an artificially compressible fluid and an incompressible fluid. Furthermore, we obtain the intermediate states for the Euler equations in the case of an immiscible, compressible two-phase flow.

4.1. COMPRESSIBLE FLUID

Suppose that u , v and w denote the x , y and z components of a fluid velocity $\mathbf{u} \in \mathbb{R}^3$ in a Cartesian coordinate system, respectively, and that $\rho \in \mathbb{R}^+$ denotes the density of the fluid. Consider the hyperbolic system (1a) with $\mathbf{q} = (\rho u, \rho v, \rho w, \rho)^T$ and $\mathbf{f}(\mathbf{q})$ given by

$$\mathbf{f}(\mathbf{q}) = (q_1^2/q_4 + p(q_4), q_1 q_2/q_4, q_1 q_3/q_4, q_1)^T. \quad (28)$$

Then equations (1a) are the Euler equations for a compressible fluid in one dimension. In this section it is assumed that the pressure is related to the density via an equation of state of the form $p = p(\rho)$, with $p \in C^1(\mathbb{R}^+, \mathbb{R}^+)$ an increasing function. An example is Tait's equation of state, which is often used to model the behavior of water:

$$p(\rho) = \alpha\rho^\gamma + \beta, \quad (29)$$

where $\alpha, \gamma \in]0, \infty[$ and $\beta \in \mathbb{R}$ are given constants. Our objective now is to obtain the approximate intermediate states for the Euler equations (1a), (28).

In order to compute the intermediate states from (25), k -Riemann invariants for the system under consideration have to be derived first. The Jacobian of the flux vector (28) reads

$$\mathbf{A}(\mathbf{q}) = \begin{pmatrix} 2q_1/q_4 & 0 & 0 & -q_1^2/q_4^2 + c^2(q_4) \\ q_2/q_4 & q_1/q_4 & 0 & -q_1 q_2/q_4 \\ q_3/q_4 & 0 & q_1/q_4 & -q_3 q_1/q_4 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad (30)$$

where $c(\rho) = \sqrt{\partial p / \partial \rho}$ denotes the speed of sound. Computation of the eigenvalues of $\mathbf{A}(\mathbf{q})$ and the corresponding eigenvectors then yields

$$\lambda_1 = q_1/q_4 - c(q_4), \quad \lambda_{2,3} = q_1/q_4, \quad \lambda_4 = q_1/q_4 + c(q_4), \quad (31)$$

and

$$\begin{aligned}\mathbf{r}_1 &= (q_1/q_4 - c(q_4), q_2/q_4, q_3/q_4, 1)^T, \\ \mathbf{r}_2 &= (0, 1, 0, 0)^T, \\ \mathbf{r}_3 &= (0, 0, 1, 0)^T, \\ \mathbf{r}_4 &= (q_1/q_4 + c(q_4), q_2/q_4, q_3/q_4, 1)^T.\end{aligned}\tag{32}$$

Notice that the eigenvalue λ_k and the eigenvector \mathbf{r}_k are genuinely nonlinear for $k = 1, 4$ and linearly degenerate for $k = 2, 3$. Riemann invariants are then obtained by solving partial differential equations (8), with the eigenvectors according to (32). The details are omitted here, but it can easily be verified that (33) constitutes a complete set of k -Riemann invariants:

$$\begin{aligned}\psi_1^2 &= q_1/q_4 + \phi(q_4), & \psi_1^3 &= q_2/q_4, & \psi_1^4 &= q_3/q_4, \\ \psi_2^1 &= q_1, & \psi_2^3 &= q_3, & \psi_2^4 &= q_4, \\ \psi_3^1 &= q_1, & \psi_3^2 &= q_2, & \psi_3^4 &= q_4, \\ \psi_4^1 &= q_1/q_4 - \phi(q_4), & \psi_4^2 &= q_2/q_4, & \psi_4^3 &= q_3/q_4,\end{aligned}\tag{33a}$$

where $\phi(\rho)$ is defined by

$$\phi(\rho) = \int_0^\rho \frac{c(\eta)}{\eta} d\eta.\tag{33b}$$

The intermediate states can now be extracted from (25), (33). In view of the linear degeneracy of the eigenvalues λ_2 and λ_3 and the arguments presented in section 3.1, we ignore $\tilde{\mathbf{q}}_{1/2}$. We then find that \mathbf{q}_0 is connected to \mathbf{q}_1 via two approximate intermediate states $\tilde{\mathbf{q}}_{1/3}$ and $\tilde{\mathbf{q}}_{2/3}$:

$$\tilde{\mathbf{q}}_{1/3} = \phi^{-1} \left(\frac{1}{2} \psi_1^2(\mathbf{q}_0) - \frac{1}{2} \psi_4^1(\mathbf{q}_1) \right) \begin{pmatrix} \frac{1}{2} \psi_1^2(\mathbf{q}_0) + \frac{1}{2} \psi_4^1(\mathbf{q}_1) \\ \psi_1^3(\mathbf{q}_0) \\ \psi_1^4(\mathbf{q}_0) \\ 1 \end{pmatrix} \tag{34}$$

and

$$\tilde{\mathbf{q}}_{2/3} = \phi^{-1} \left(\frac{1}{2} \psi_1^2(\mathbf{q}_0) - \frac{1}{2} \psi_4^1(\mathbf{q}_1) \right) \begin{pmatrix} \frac{1}{2} \psi_1^2(\mathbf{q}_0) + \frac{1}{2} \psi_4^1(\mathbf{q}_1) \\ \psi_4^2(\mathbf{q}_1) \\ \psi_4^3(\mathbf{q}_1) \\ 1 \end{pmatrix}, \tag{35}$$

where $\phi^{-1}(\psi)$ denotes the inverse of $\phi(\psi)$.

For a fluid that is described by Tait's equation of state, the intermediate states can be determined by substituting (29) in equations (33) to (35). The intermediate velocity components $\tilde{v}_{1/3}$, $\tilde{v}_{2/3}$, $\tilde{w}_{1/3}$ and $\tilde{w}_{2/3}$ are immediately obtained from (33):

$$\begin{aligned}\tilde{v}_{1/3} &= v_0, & \tilde{v}_{2/3} &= v_1, \\ \tilde{w}_{1/3} &= w_0, & \tilde{w}_{2/3} &= w_1.\end{aligned}\quad (36)$$

From (33) it is also clear that $\tilde{u}_{1/3} = \tilde{u}_{2/3} \equiv \tilde{u}_{1/2}$ and $\tilde{\rho}_{1/3} = \tilde{\rho}_{2/3} \equiv \tilde{\rho}_{1/2}$. To determine $\tilde{u}_{1/2}$ and $\tilde{\rho}_{1/2}$, it is necessary to distinguish between the cases $\gamma = 1$ and $\gamma \neq 1$. For $\gamma = 1$ one obtains

$$\begin{aligned}\tilde{u}_{1/2} &= \frac{1}{2}(u_0 + u_1) + \frac{\sqrt{\alpha}}{2} \ln(\rho_0/\rho_1), \\ \tilde{\rho}_{1/2} &= \sqrt{\rho_0 \rho_1} \exp\left(\frac{u_0 - u_1}{2\sqrt{\alpha}}\right).\end{aligned}\quad (37)$$

In case $\gamma \neq 1$, it is convenient to express the density in terms of the speed of sound:

$$\begin{aligned}\tilde{u}_{1/2} &= \frac{1}{2}(u_0 + u_1) + \frac{1}{\gamma - 1}[c(\rho_0) - c(\rho_1)], \\ c(\tilde{\rho}_{1/2}) &= \frac{\gamma - 1}{4}(u_0 - u_1) + \frac{1}{2}[c(\rho_0) + c(\rho_1)].\end{aligned}\quad (38)$$

4.2. ARTIFICIALLY COMPRESSIBLE FLUID

Assume that u , v and w again denote the x , y and z components of a fluid velocity $\mathbf{u} \in \mathbb{R}^3$ in a Cartesian coordinate system, respectively, and that $p \in \mathbb{R}^+$ denotes the fluid pressure. Consider hyperbolic system (1a) with $\mathbf{q} = (u, v, w, p)^T$. Let $\mathbf{f}(\mathbf{q})$ be

$$\mathbf{f}(\mathbf{q}) = (q_1^2 + q_4, q_1 q_2, q_1 q_3, c^2 q_1)^T, \quad (39)$$

with c constant. Equations (1a), (39) are the Euler equations for an artificially compressible fluid in one dimension. Notice that the $\partial p / \partial t$ term that occurs in (1a) in this case, implies compressibility of the fluid.

To determine the intermediate states, we first derive Riemann invariants for (1a), (39). For the Jacobian of $\mathbf{f}(\mathbf{q})$ we simply obtain

$$\mathbf{A}(\mathbf{q}) = \begin{pmatrix} 2q_1 & 0 & 0 & 1 \\ q_2 & q_1 & 0 & 0 \\ q_3 & 0 & q_1 & 0 \\ c^2 & 0 & 0 & 0 \end{pmatrix}. \quad (40)$$

The eigenvalues and corresponding eigenvectors of $\mathbf{A}(\mathbf{q})$ follow by straightforward computation:

$$\lambda_1 = q_1 - \sqrt{q_1^2 + c^2}, \quad \lambda_{2,3} = q_1, \quad \lambda_4 = q_1 + \sqrt{q_1^2 + c^2}, \quad (41)$$

and

$$\begin{aligned} \mathbf{r}_1 &= \left(1, -q_2/\sqrt{q_1^2 + c^2}, -q_3/\sqrt{q_1^2 + c^2}, -q_1 - \sqrt{q_1^2 + c^2} \right)^T, \\ \mathbf{r}_2 &= (0, 1, 0, 0)^T, \\ \mathbf{r}_3 &= (0, 0, 1, 0)^T, \\ \mathbf{r}_4 &= \left(1, q_2/\sqrt{q_1^2 + c^2}, q_3/\sqrt{q_1^2 + c^2}, -q_1 + \sqrt{q_1^2 + c^2} \right)^T. \end{aligned} \quad (42)$$

The eigenpairs $(\lambda_1, \mathbf{r}_1)$ and $(\lambda_4, \mathbf{r}_4)$ are genuinely non-linear, whereas the eigenpairs $(\lambda_2, \mathbf{r}_2)$ and $(\lambda_3, \mathbf{r}_3)$ are linearly degenerate. Riemann invariants are now obtained by solving (8), (42):

$$\begin{aligned} \psi_1^2 &= q_2 \lambda_4, & \psi_1^3 &= q_3 \lambda_4, & \psi_1^4 &= \lambda_4 \exp([2q_4 + q_1 \lambda_4]/c^2), \\ \psi_2^1 &= q_1, & \psi_2^3 &= q_3, & \psi_2^4 &= q_4, \\ \psi_3^1 &= q_1, & \psi_3^2 &= q_2, & \psi_3^4 &= q_4, \\ \psi_4^1 &= q_2 \lambda_1, & \psi_4^2 &= q_3 \lambda_1, & \psi_4^3 &= \lambda_1 \exp([2q_4 + q_1 \lambda_1]/c^2). \end{aligned} \quad (43)$$

The foregoing invariants have linearly independent gradients. Hence, the intermediate states can be obtained from (25), (43).

Considering the linear degeneracy of λ_2, λ_3 , we only need to obtain $\tilde{\mathbf{q}}_{1/3}$ and $\tilde{\mathbf{q}}_{2/3}$. Unfortunately, in this instance we have not succeeded in deriving a closed form expression for these intermediate states. However, from (43) it immediately follows that $\tilde{u}_{1/3} = \tilde{u}_{2/3} \equiv \tilde{u}_{1/2}$ and $\tilde{\rho}_{1/3} = \tilde{\rho}_{2/3} \equiv \tilde{\rho}_{1/2}$. Then, using the expressions for ψ_1^4 and ψ_4^3 , one finds that $\tilde{u}_{1/2}$ is determined by the implicit relation:

$$\left(\frac{\tilde{u}_{1/2} + \sqrt{\tilde{u}_{1/2}^2 + c^2}}{\tilde{u}_{1/2} - \sqrt{\tilde{u}_{1/2}^2 + c^2}} \right) \exp \left(\frac{2\tilde{u}_{1/2}\sqrt{\tilde{u}_{1/2}^2 + c^2}}{c^2} \right) = \frac{\psi_1^4(\mathbf{q}_0)}{\psi_4^3(\mathbf{q}_1)}. \quad (44)$$

Once $\tilde{u}_{1/2}$ has been solved from (44), $\tilde{\mathbf{q}}_{1/3}$ and $\tilde{\mathbf{q}}_{2/3}$ are simply obtained from (43).

4.3. INCOMPRESSIBLE FLUID

We consider the Euler equations for an incompressible flow. Assume that $\mathbf{u} \in \mathbb{R}^3$ denotes the fluid velocity and that the fluid pressure divided by the

(constant) fluid density is designated $p \in \mathbb{R}^+$. Next, let $\mathbf{u}\mathbf{u} \in C^1(\mathbb{R}^3, \mathbb{R}^{3 \times 3})$ be the convective momentum flux tensor. The Euler equations for an incompressible fluid read

$$\frac{\partial \mathbf{u}}{\partial t} + \operatorname{div} \mathbf{u}\mathbf{u} + \nabla p = 0, \quad (45a)$$

$$\operatorname{div} \mathbf{u} = 0. \quad (45b)$$

Due to the absence of a time derivative in (45b), equations (45) do not constitute a hyperbolic system. However, equation (45a) can trivially be recast into an inhomogeneous hyperbolic system governing \mathbf{u} :

$$\frac{\partial \mathbf{u}}{\partial t} + \operatorname{div} \mathbf{u}\mathbf{u} = -\nabla p. \quad (46)$$

Equation (45b) is then interpreted as a constraint on \mathbf{u} and p serves as a Lagrangian multiplier to enforce the constraint. Solving the Euler equations for an incompressible fluid now requires the resolution of the hyperbolic system (46) subject to the constraint (45b). Here, we shall only concern ourselves with the hyperbolic part of the operator. Furthermore, in the following section we will only consider the homogeneous system in one dimension, i.e., we shall neglect the forcing term $-\nabla p$ and (assuming a Cartesian coordinate system is employed) the flux gradients in the y and z direction. We then retrieve an expression of the form (1a), with $\mathbf{q} = (u, v, w)^T$, where u, v, w again denote the x, y, z components of the fluid velocity $\mathbf{u} \in \mathbb{R}^3$ in a Cartesian coordinate system, respectively, and $\mathbf{f}(\mathbf{q})$ given by

$$\mathbf{f}(\mathbf{q}) = (q_1^2, q_1 q_2, q_1 q_3)^T. \quad (47)$$

We acknowledge that the first equation of (1a), (47) is decoupled from the remaining system and can therefore be treated separately. However, for completeness we refrain from doing so.

To obtain the approximate intermediate states for (1a), (47), we first determine Riemann invariants for this system. The Jacobian of $\mathbf{f}(\mathbf{q})$ reads

$$\mathbf{A}(\mathbf{q}) = \begin{pmatrix} 2q_1 & 0 & 0 \\ q_2 & q_1 & 0 \\ q_3 & 0 & q_1 \end{pmatrix}, \quad (48)$$

with the eigenvalues

$$\lambda_1 = q_1^2, \quad \lambda_{2,3} = q_1, \quad (49)$$

and the corresponding eigenvectors

$$\begin{aligned}\mathbf{r}_1 &= (q_1, q_2, q_3)^T, \\ \mathbf{r}_2 &= (0, 1, 0)^T, \\ \mathbf{r}_3 &= (0, 0, 1)^T.\end{aligned}\tag{50}$$

The first eigenpair is neither linearly degenerate nor genuinely nonlinear: the gradient of $\lambda_1(\mathbf{q})$ in the direction of $\mathbf{r}_1(\mathbf{q})$ vanishes for $q_1 = 0$, but is nonzero otherwise. Nevertheless, for our purposes it is sufficient to treat $(\lambda_1, \mathbf{r}_1)$ as a genuinely nonlinear eigenpair, because the eigenvalue vanishes only if $q_1 = 0$ and, therefore, the eigenvalue can change sign only once along $\mathcal{R}_1(\mathbf{q}_L)$. The second and third eigenpair are linearly degenerate. Riemann invariants are obtained by solving (8), (50):

$$\begin{aligned}\psi_1^2 &= q_1/q_2, & \psi_1^3 &= q_1/q_3, \\ \psi_2^1 &= q_1, & \psi_2^3 &= q_3, \\ \psi_3^1 &= q_1, & \psi_3^2 &= q_2.\end{aligned}\tag{51}$$

These invariants have linearly independent gradients in \mathbb{R}^3 .

Because the second and third eigenpair are linearly degenerate, \mathbf{q}_0 and \mathbf{q}_1 are connected via a single intermediate state $\tilde{\mathbf{q}}_{1/2}$. This intermediate state is immediately obtained from (25), (51):

$$\tilde{\mathbf{q}}_{1/2} = \begin{cases} \mathbf{0}, & \text{if } u_0 = 0, \\ \frac{u_1}{u_0} \mathbf{q}_0, & \text{otherwise.} \end{cases}\tag{52}$$

4.4. TWO-PHASE FLOW

In this section we derive the intermediate states for the Euler equations for an immiscible, compressible two-phase flow. The phases are supposed to be separated by a moving interface, which is described by the time-dependent set $\mathcal{I}(t) = \{\mathbf{x} \in \mathbb{R}^3 \mid \theta(\mathbf{x}, t) = 0\}$. Furthermore, we assume $\theta(\mathbf{x}, t)$ to be negative in one phase and positive in the other. As a result of the immiscibility of the phases, the following kinematic condition applies:

$$\frac{\partial \theta}{\partial t} + \mathbf{u} \cdot \vec{\nabla} \theta = 0,\tag{53}$$

where $\mathbf{u} \in \mathbb{R}^3$ again denotes the fluid velocity. Employing the continuity equation for compressible fluids, we can restate kinematic condition (53) in conservation form:

$$\frac{\partial \rho \theta}{\partial t} + \vec{\nabla} \cdot \rho \theta \mathbf{u} = \rho \left(\frac{\partial \theta}{\partial t} + \mathbf{u} \cdot \vec{\nabla} \theta \right) + \theta \left(\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot \rho \mathbf{u} \right).\tag{54}$$

The first term in parentheses vanishes due to (53), the second due to continuity. Hence, $\rho\theta$ is a conserved quantity. Suppose that throughout the entire fluid volume the pressure is related to the density via an equation of state of the form $p = p(\theta, \rho)$. Then, again using u, v, w to designate the velocity components relative to a Cartesian coordinate system and ignoring spatial derivatives in y and z direction, we retrieve (1a), with $\mathbf{q} = (\rho u, \rho v, \rho w, \rho\theta, \rho)^T$ and

$$\mathbf{f}(\mathbf{q}) = (q_1^2/q_5 + p(q_4/q_5, q_5), q_1 q_2/q_5, q_1 q_3/q_5, q_1 q_4/q_5, q_1)^T. \quad (55)$$

Equations (1a), (55) constitute the one-dimensional Euler equations for an immiscible, compressible two-phase flow.

Our first objective now is to derive Riemann invariants for (1a), (55). We define $c_1 = c_1(\theta, \rho) = \sqrt{\partial p / \partial \theta}$ and $c_2 = c_2(\theta, \rho) = \sqrt{\partial p / \partial \rho}$. Then, the Jacobian of (55) reads:

$$\mathbf{A}(\mathbf{q}) = \begin{pmatrix} 2q_1/q_5 & 0 & 0 & c_1^2/q_5 & -q_1^2/q_5^2 - c_1^2 q_4/q_5^2 + c_2^2 \\ q_2/q_5 & q_1/q_5 & 0 & 0 & -q_2 q_1/q_5^2 \\ q_3/q_5 & 0 & q_1/q_5 & 0 & -q_3 q_1/q_5^2 \\ q_4/q_5 & 0 & 0 & q_1/q_5 & -q_4 q_1/q_5^2 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (56)$$

The eigenvalues and eigenvectors of $\mathbf{A}(\mathbf{q})$ are

$$\lambda_1 = q_1/q_5 - c_2, \quad \lambda_{2,3,4} = q_1/q_5, \quad \lambda_5 = q_1/q_5 + c_2, \quad (57)$$

and

$$\begin{aligned} \mathbf{r}_1 &= (q_1/q_5 - c_2, q_2/q_5, q_3/q_5, q_4/q_5, 1)^T, \\ \mathbf{r}_2 &= (0, 1, 0, 0, 0)^T, \\ \mathbf{r}_3 &= (0, 0, 1, 0, 0)^T, \\ \mathbf{r}_4 &= (q_1 c_1^2, 0, 0, -c_2^2 q_5^2 + c_1^2 q_4, q_5 c_1^2)^T, \\ \mathbf{r}_5 &= (q_1/q_5 + c_2, q_2/q_5, q_3/q_5, q_4/q_5, 1)^T. \end{aligned} \quad (58)$$

The eigenvalue λ_k and the eigenvector \mathbf{r}_k are genuinely nonlinear for $k = 1, 5$ and linearly degenerate for $k = 2, 3, 4$. Riemann invariants can now be

obtained by solving (8), (58):

$$\begin{aligned} \psi_1^2 &= q_1/q_5 + \phi, & \psi_1^3 &= q_2/q_5, & \psi_1^4 &= q_3/q_5, & \psi_1^5 &= q_4/q_5, \\ \psi_2^1 &= q_1/q_5, & \psi_2^3 &= q_3, & \psi_2^4 &= q_4, & \psi_2^5 &= p, \\ \psi_3^1 &= q_1/q_5, & \psi_3^2 &= q_2, & \psi_3^4 &= q_4, & \psi_3^5 &= p, \\ \psi_4^1 &= q_1/q_5, & \psi_4^2 &= q_2, & \psi_4^3 &= q_3, & \psi_4^5 &= p, \\ \psi_5^1 &= q_1/q_5 - \phi, & \psi_5^2 &= q_2/q_5, & \psi_5^3 &= q_3/q_5, & \psi_5^4 &= q_4/q_5, \end{aligned} \quad (59a)$$

with $p = p(\theta, \rho)$ and $\phi = \phi(\theta, \rho)$ defined by

$$\phi(\theta, \rho) = \int_0^\rho \frac{c_2(\theta, \eta)}{\eta} d\eta. \quad (59b)$$

Observe that θ is a k -Riemann invariant for $k \in \{1, 5\}$. Hence, it may be inferred that the phase transition is a contact discontinuity. Moreover, because both u and p are k -Riemann invariants for $k \in \{2, 3, 4\}$, the pressure and the normal velocity component are continuous across the interface.

The intermediate states can now be obtained from (25), (59). Because the linearly degenerate eigenvalue q_1/q_5 has algebraic multiplicity 3, only two intermediate states have to be distinguished. Trivially,

$$\begin{pmatrix} \tilde{v}_{1/3} \\ \tilde{w}_{1/3} \\ \tilde{\theta}_{1/3} \end{pmatrix} = \begin{pmatrix} v_0 \\ w_0 \\ \theta_0 \end{pmatrix}, \quad \begin{pmatrix} \tilde{v}_{2/3} \\ \tilde{w}_{2/3} \\ \tilde{\theta}_{2/3} \end{pmatrix} = \begin{pmatrix} v_1 \\ w_1 \\ \theta_1 \end{pmatrix}, \quad (60)$$

and $\tilde{u}_{1/3} = \tilde{u}_{2/3} \equiv \tilde{u}_{1/2}$. Then, $\tilde{\rho}_{1/3}$ and $\tilde{\rho}_{2/3}$ are determined by

$$\begin{aligned} \phi(\theta_0, \tilde{\rho}_{1/3}) + \phi(\theta_1, \tilde{\rho}_{2/3}) &= u_0 - u_1 + \phi(\theta_0, \rho_0) + \phi(\theta_1, \rho_1), \\ p(\theta_0, \tilde{\rho}_{1/3}) &= p(\theta_1, \tilde{\rho}_{2/3}). \end{aligned} \quad (61)$$

We refrain from a further reduction of these expressions and suffice by stating that once the intermediate densities have been obtained, $\tilde{u}_{1/2}$ follows by straightforward computation.

5. Conclusions

In spite of the absence of shock waves in most hydrodynamic applications, sufficient reason remains to employ Godunov-type schemes in this field. The shock capturing ability of these schemes renders them notably useful in the case of two-phase flow. In the present work we developed an Osher-type Riemann solver and we investigated several of its applications in the

field of hydrodynamics. First, the Riemann problem was examined. Subsequently, Osher's approximate Riemann solver was discussed. It was shown that this scheme employs a rarefaction-waves-only approximate Riemann solution and that this approximation is accurate even in the presence of (weak) shocks. Then, it was demonstrated that the centered flux approximation obtained by means of Osher's scheme is not necessarily accurate and, therefore, a modified scheme was proposed. Finally, details were presented for several types of fluid-models commonly used in hydrodynamics.

Acknowledgment

This work was performed under a research contract with the Maritime Research Institute Netherlands.

References

- Godunov S K (1959). Finite difference method for numerical computation of discontinuous solutions of the equations of fluid Dynamics, *Matematicheskii Sbornik* **47**, pp 271–306. (in Russian).
- Sweby P K (1984). High resolution schemes using flux limiters for hyperbolic conservation laws, *SIAM Journal on Numerical Analysis* **21**, pp 995–1011.
- Spekreijse S P (1987). Multigrid solution of monotone second-order discretizations of hyperbolic conservation laws, *Mathematics of Computation* **49**, pp 135–155.
- Mulder W A, Osher S and Sethian J A (1992). Computing interface motion in compressible gas dynamics, *Journal of Computational Physics* **100**, pp 209–228.
- Chang Y C, Hou T Y, Merriman B and Osher S (1996). A level set formulation of Eulerian interface capturing methods for incompressible fluid flows, *Journal of Computational Physics* **124**, pp 449–464.
- Kelecy F J and Pletcher R H (1997). The development of a free surface capturing approach for multidimensional free surface flows in closed containers, *Journal of Computational Physics* **138**, pp 939–980.
- Roe P L (1981). Approximate Riemann solvers, parameter vectors, and difference schemes, *Journal of Computational Physics* **43**, pp 357–372.
- Osher S and Solomon F (1982). Upwind difference schemes for hyperbolic conservation laws, *Mathematics of Computation* **38**, pp 339–374.
- Lax P D (1957). Hyperbolic systems of conservation laws II, *Communications on Pure and Applied Mathematics* **10**, pp 537–566.
- LeVeque R J (1990). Numerical Methods for Conservation Laws, Birkhäuser.
- Smoller J (1983). Shock Waves and Reaction-Diffusion Equations, Springer.
- Lax P D (1973). Hyperbolic systems of conservation laws and the mathematical theory of shock waves, *Regional Conference Series in Applied Mathematics*, **11**, SIAM.
- Osher S and Chakravarthy S (1982). Upwind schemes and boundary conditions with applications to Euler equations in general geometries, *Journal of Computational Physics* **50**, pp 447–481.
- van Leer B (1984). On the relation between the upwind-differencing schemes of Godunov, Engquist-Osher and Roe, *SIAM Journal on Scientific and Statistical Computing* **5**, pp 1–20.

A STAGGERED SCHEME FOR HYPERBOLIC CONSERVATION LAWS APPLIED TO COMPUTATION OF FLOW WITH CAVITATION

D.R. VAN DER HEUL, C. VUIK AND P. WESSELING

J.M. Burgers Center and

Faculty of Information Technology and Systems,

Delft University of Technology,

Mekelweg 4, 2628 CD Delft, The Netherlands

Emails: vdheul@its.tudelft.nl

Abstract. We demonstrate the advantages of discretizing on a staggered grid for the computation of solutions to hyperbolic systems of conservation laws arising from instationary flow of an inviscid fluid with an arbitrary equation of state. The method is used to compute unsteady sheet cavitation, with the homogeneous equilibrium model for two phase flow. Due to strong variations in the speed of sound almost incompressible as well as highly compressible regions occur simultaneously. The compressible pressure correction solution method is able to handle both, because accuracy and efficiency are uniform in the Mach number.

Results show good agreement with those obtained with a cavity interface tracking method.

1. Introduction

This paper describes the development of an efficient and accurate method for computing a flow with cavitation using the homogeneous equilibrium model (HEM). In Section 2 we will briefly discuss the importance of cavitation calculation methods for industrial applications. Section 3 discusses the assumptions of the HEM model, and the requirements that it imposes on the solution procedure. The governing equations are given in Section 4. In Section 5.1 we present the solution procedure we initially employed and compared with the Osher scheme for a model problem for a nonconvex hyperbolic system. Based on a thorough stability analysis, a first order

time integration method is formulated with (almost) unconditional stability for the two dimensional isothermal Euler equations for Mach numbers between 0-30 in Section 5.2. Results obtained with the method are presented in Section 6 and compared with a cavity interface tracking method.

2. Cavitating flow

Hydrodynamic cavitation is the phenomenon, that due to the flow dynamics the pressure in a liquid falls below the vapor pressure and the liquid vaporizes. In heavily loaded hydraulic machinery, e.g. rotary pumps, valves, a controlled amount of cavitation is unavoidable. It is found that the erosive effects of an unsteady cavity, and therefore the lifespan of the equipment, is proportional to the size of the cavity. The goal of a cavitation prediction method is therefore twofold:

1. Optimize the design of hydraulic equipment to minimize the amount of cavitation.
2. Accurately predict the size of the vapor cavities, to enable efficient maintenance scheduling.

In both cases the interest is in macroscopic quantities, e.g. the average thickness and length, more than in details of the internal flow of the cavity.

3. Homogeneous equilibrium model, HEM

A rather simple model for two-phase flow which has gained popularity to model cavitating flow (Dellanoy and Kueny, 1990; Hoeijmakers et al., 1998; Merkle et al., 1998), is the Homogeneous Equilibrium Model. Instead of treating the two phases liquid and vapor separately, the fluidum is treated as a homogeneous liquid/vapor mixture. Based on the assumption of thermodynamic equilibrium and the neglect of velocity slip between the two phases, single phase flow equations can be derived for the mixture, completed with a mixture equation of state. This equation of state makes the density of the mixture equal to the density of the liquid phase if the pressure is above the vapor pressure and equal to the density of the vapor phase, when the pressure falls below the vapor pressure, with a smooth but steep transition in between. The EOS is schematically shown in nondimensional form in Figure 1(a).

Both the pure liquid and pure vapor phase are almost incompressible. In industrial applications water will always contain a small quantity of undissolved gas, that will substantially lower the speed of sound of this mixture, with respect to the speed of sound of pure water. Figure 1(b) shows the nondimensional speed of sound as a function of nondimensional pressure of the EOS of Figure 1(a). The ratio between the speed of sound

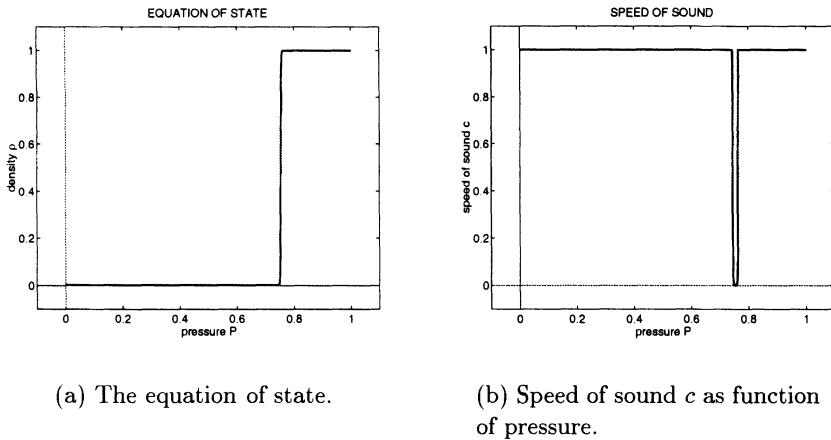


Figure 1. Properties of the liquid/vapor mixture.

in the mixture region and that of the liquid phase is of the order of 10^{-3} . In (Stutz and Reboud, 1997) it is shown experimentally that the velocity in the mixture regions is of the same order of magnitude as in the pure liquid regions. The latter results in a Mach number of the order of 10^{-3} in the liquid phase, and in the order of 10-30 in the mixture region.

Although the model is simple, efficient handling of the low Mach number flow of the liquid phase together with the very high compressibility of the mixture state represents a challenging numerical task. Previously use was made either of artificial compressibility techniques (Hoeijmakers et al., 1998), or SIMPLE type methods (Dellanoy and Kueny, 1990), but we will consider a more efficient approach based on compressible pressure correction.

4. Governing equations

We employ the isothermal Euler equations, completed with the earlier mentioned equation of state. Although the medium in the transition between liquid and vapor state is highly compressible, the liquid state is almost incompressible, characterized by a Mach number of order 10^{-3} . Therefore use is made of the following Mach-uniform pressure based formulation of the compressible Euler equations (Bijl and Wesseling, 1998):

$$\frac{d\rho}{dp} \frac{\partial p}{\partial t} + (\rho U^\alpha)_{,\alpha} = 0, \quad (1)$$

$$\frac{\partial \rho U^\alpha}{\partial t} + (\rho U^\alpha U^\beta)_{,\beta} = -(g^{\alpha\beta} p)_{,\beta}. \quad (2)$$

completed with a barotropic equation of state $\rho = \rho(p)$. We use general coordinates and tensor notation, with $U^\alpha = \mathbf{a}^{(\alpha)} \cdot \mathbf{u}$ denoting the contravariant velocity components, $g^{\alpha\beta}$ the contravariant metric tensor and $\mathbf{a}^{(\alpha)}$ the contravariant base vector with respect to the mapping $Q : \mathbf{x} = \mathbf{x}(\boldsymbol{\xi})$, where \mathbf{x} are Cartesian coordinates, while $\boldsymbol{\xi}$ are general boundary-fitted coordinates. For details on spatial discretisation we refer to (Bijl and Wesseling, 1998).

5. Solution procedure

We will discuss two different solution methods: the one derived from (Bijl and Wesseling, 1998) and a new approach that has more favorable stability properties for the very high Mach number regime. The first method (Bijl and Wesseling, 1998) is compared with the Osher scheme (Osher and Solomon, 1982).

5.1. COMPRESSIBLE PRESSURE CORRECTION

The solution procedure we employ is compressible pressure correction. For sake of brevity we will only discuss the one dimensional case.

First a prediction of the momentum m^* is made using the pressure at the previous time level:

$$\frac{m_{j+\frac{1}{2}}^* - m_{j+\frac{1}{2}}^n}{\delta t} + \frac{1}{\delta x} (u^n m^*)|_{j-\frac{1}{2}}^{j+\frac{1}{2}} = - \frac{1}{\delta x} p^n|_j^{j+1}. \quad (3)$$

The following pressure correction is postulated:

$$m_{j+\frac{1}{2}}^{n+1} = m_{j+\frac{1}{2}}^* - \frac{\delta t}{\delta x} (p^{n+1} - p^n)|_j^{j+1}, \quad (4)$$

and substituted in the mass conservation equation. This leads to the following nonlinear pressure correction equation:

$$\begin{aligned} & \frac{\rho(p_j^n + \delta p_j) - \rho(p_j^n)}{\delta t} + \\ & \frac{1}{\delta x} \left(\frac{\rho(\delta p_{k-\frac{1}{2}} + p_{k-\frac{1}{2}}^n)}{\rho(\delta p_k + p_k^n)} \left(m_k^* - (\delta p_{k+\frac{1}{2}} - \delta p_{k-\frac{1}{2}}) \right) \Big|_{k=j-\frac{1}{2}}^{k=j+\frac{1}{2}} \right) = 0. \end{aligned} \quad (5)$$

Equation (5) is Newton linearized and iteratively solved by a mix of nonlinear Gauss-Seidel and preconditioned GMRES (VanderHeul et al., 1998).

In (VanderHeul et al., 1998) a comparison was made between the staggered discretisation under consideration and the Osher scheme. The latter

was chosen for two reasons. First of all the scheme can be readily applied to general hyperbolic systems and is not restricted to the Euler equations for a perfect gas. Secondly the Osher scheme is proven to converge to a numerical solution that satisfies both the entropy and the Rankine-Hugoniot conditions, provided the shocks are sufficiently weak.

Both schemes were applied to two different Riemann problems for a fluid with a nonconvex equation of state, very similar to the one of the homogeneous equilibrium model. To circumvent the difficulties of low Mach number flow, the 'liquid' and 'vapor' states where chosen to be compressible. Even for these compressible test cases it was found that the staggered scheme is simpler and more efficient than the approximate Riemann solver, but gives solutions of comparable accuracy. To make the Osher scheme able to handle the low Mach number flow of our application some form of time accurate preconditioning with dual timestepping should be applied, rendering that scheme even less efficient compared to the staggered scheme.

Numerical experiments showed that this time integration method, although implicit, has to respect a restriction on the timestep, that becomes more severe when the Mach number is increased (VanderHeul et al., 2000).

5.2. FIRST ORDER TIME INTEGRATION WITH UNCONDITIONAL STABILITY FOR MACH=0-30.

After a thorough analysis of the stability of the time integration method (VanderHeul et al., 2000), we have derived a solution procedure with near optimal stability properties. The method consists of the following 3 steps:

1. solve additional density predictor equation

$$\frac{\rho^* - \rho^n}{\delta t} + (u^n \rho^*)_x = 0, \quad (6)$$

2. solve momentum predictor equation

$$\frac{m^* - m^n}{\delta t} + \left(\left(u^n + \frac{m^n}{\rho^*} \right) m^* - u^n m^n \right)_x = -p_x^n, \quad (7)$$

3. solve pressure correction equation

$$\begin{aligned} & \frac{\rho(p^{n+1}) - \rho^n}{\delta t} + \frac{1}{2} (s^{n+1} (m^* - (p^{n+1} - p^n)_x))_x + \\ & \frac{1}{2} (u^n \rho(p^{n+1}))_x = 0, \quad s^{n+1} = \frac{\rho^{n+1} - \frac{1}{2} \delta x \frac{\partial \rho^{n+1}}{\partial x}}{\rho^{n+1}}. \end{aligned} \quad (8)$$

Some remarks concerning this method are:

- Newton linearisation is used for the convective terms in the momentum conservation equation.
- A Crank-Nicolson type scheme is used for the velocity in the convective terms in the pressure correction equation (8). This gives a more favorable spectrum of the eigenvalues of the time integration methods, resulting in increased damping in the high frequency modes, enhancing almost unconditional stability (VanderHeul et al., 2000).
- The Newton linearized formulation of the convective terms enables extension of the method to second order temporal accuracy, with the θ -method (VanderHeul et al., 2000).
- The Newton linearized formulation of the convective terms enables extension of the method to second order temporal accuracy, with the θ -method (VanderHeul et al., 2000).
- Implicit higher order spatial discretisation on a compact stencil can be achieved by introducing intermediate defect correction steps (Khosla and Rubin, 1997).

One drawback of this improved pressure correction scheme, is the stronger damping of high frequency modes for moderate Mach numbers leading to a loss of accuracy near moving discontinuities, compared to the schemes described in Section 5.1

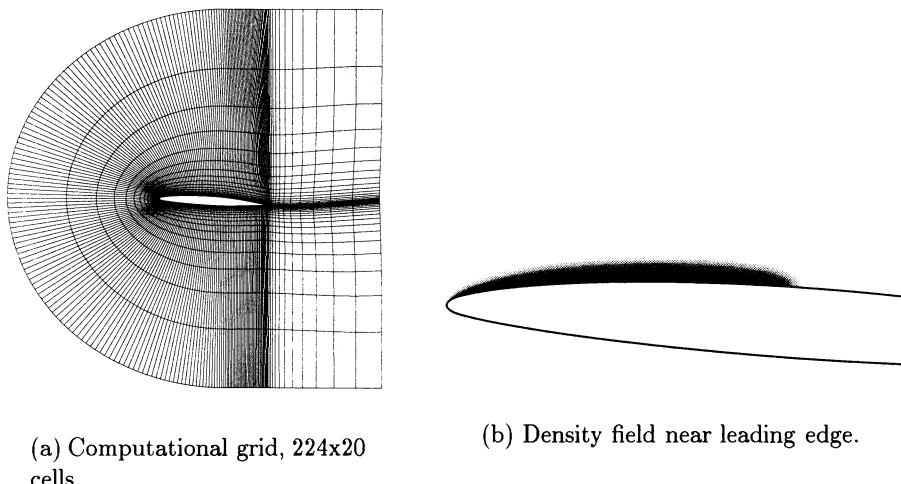


Figure 2. Cavitating flow on NACA66 hydrofoil, $\alpha = 4^\circ$, $\sigma = 0.9$.

6. Results

The cavitating flow over a NACA66 airfoil (Figure 2(a)) has been computed for three different values of the cavitation number σ , defined as:

$$\sigma = \frac{p_\infty - p_{\text{vapor}}}{\frac{1}{2}\rho_\infty V_\infty^2} \quad (9)$$

$c_{\text{vapor state}}/c_{\text{liquid state}}$	1
$c_{\text{liquid state}}/c_{\text{minimal}}$	670
$\rho_{\text{liquid state}}/\rho_{\text{vapor state}}$	100

TABLE 1. Properties of the mixture equation of state.

σ	current method	Deshpande (Deshpande et al., 1992)
1.0	0.32	0.20
0.9	0.40	0.32
0.87	0.65	0.64

TABLE 2. Maximum cavity length as fraction of chord for NACA66 for $\sigma=1.0, 0.9$ and 0.87 .

The parameters in the EOS we have chosen are listed in Table 1. Although the liquid/vapor density ratio of water is of the order of 1000, using a ratio of 100 in the computation has negligible effect on the flow, because so little momentum is carried by the vapor and mixture phase. However, diminishing the liquid/vapor density ratio leads to a reduction of the number of iterations needed to solve the nonlinear pressure equation. Figure 2(b) shows a plot of the density in the vicinity of the airfoil leading edge. The position of the cavitation bubble can be clearly identified.

We compare the maximum attained cavity lengths obtained with our method, with those of the cavity interface tracking method of Deshpande (Deshpande et al., 1992) in Table 2. In the latter method it is assumed that the liquid/vapor interface of the cavity is a streamline. On the iteratively determined liquid/vapor interface a constant pressure boundary condition $p = p_{\text{vapor}}$ is prescribed.

For these moderate values of the cavitation number the cavitation sheet will remain steady. The discrepancy is partly described to the fact that

a first order method is used, that highly smears the shock between the supersonic flow in the cavitation bubble and the subsonic flow in the liquid phase at the trailing edge of the liquid/vapor interface.

7. Conclusions and future extensions

The computation of cavitating flow with the homogeneous equilibrium model presents a demanding application, due to the fact that almost incompressible ($Ma \approx 0.001$) and highly compressible flow ($Ma \approx 20-30$) occur simultaneously. A first order scheme has been derived with unconditional stability for all Mach numbers between 0-30, with Mach uniform efficiency and accuracy. The method has been applied to calculate unsteady sheet cavitation on a hydrofoil. Results show good agreement with a cavity interface tracking method. Extension of the method to second order spatial and temporal accuracy is underway.

Acknowledgments

The work of the first author was supported by the Netherlands Organization for Scientific Research (NWO).

References

- Bijl H and Wesseling P (1998). A Unified Method for Computing Incompressible and Compressible Flows in Boundary Fitted Coordinates. *J. Comp. Phys.* **141**, pp 153-173.
- Dellanoy Y and Kueny J L (1990). Two Phase Flow Approach in Unsteady Cavitation Modelling. *Cavitation and Multiphase Flow, FED-98*, pp 153-158. Furuya O (Editor). New York, ASME.
- Deshpande M , Feng J and Merkle C L (1992). Nonlinear Euler Analysis of 2-d Cavity Flow. *Cavitation and Multiphase Flow Forum* **135**, pp 213-219.
- Hoeijmakers H W M , Janssens M E and Kwan W (1998). Numerical Simulation of Sheet Cavitation. Proc. Third International Symposium on Cavitation, April 7-10, 1998, Grenoble, Michel and Kato (Editors). Volume 2, pp 257-262.
- Merkle C L , Feng J and Buelow P E O (1998). Computational Modeling of the Dynamics of Sheet Cavitation. Proc. Third International Symposium on Cavitation, April 7-10, 1998, Grenoble, Michel and Kato (Editors). Volume 2, pp 307-311.
- Osher S and Solomon F (1982). Upwind Difference Schemes for Hyperbolic Systems of Conservation Laws. *Math. Comp.* **38**, pp 339-374.
- Stutz B and Reboud J L (1997). Two-phase Flow Structure of Sheet Cavitation. *Phys. Fluids* **9**, pp 3678-3686.
- Khosla P K and Rubin S G (1997). A Diagonally Dominant Second-order Accurate Implicit Scheme. *Computers and Fluids* **2**, pp 207-209.
- VanderHeul D R , Vuik C and Wesseling P (1998). A Staggered Scheme for Hyperbolic Conservation Laws. In *Computational Fluid Dynamics '98* **1**, pp 730-735, Wiley.
- VanderHeul D R , Vuik C and Wesseling P (2000). Stability Analysis of Time Integration Schemes for Segregated Solution Methods for Compressible Flow. *Report of the Department of Applied Mathematical Analysis 00-06*, Delft University of Technology, <ftp://ta.twi.tudelft.nl/pub/TWA/reports/00-06.ps>

AN EXPERT SYSTEM TO CONTROL THE CFL NUMBER OF IMPLICIT UPWIND METHODS

DENIS VANDERSTRAETEN

*Katholieke Universiteit Leuven
Department Computer Science
Celestijnenlaan, 200A
B-3001 Heverlee (Belgium)
e-mail: denis@cs.kuleuven.ac.be*

Abstract. Direct solution of complex steady state flows is usually too difficult numerically in the absence of a good initial guess. “Pseudo-transient-continuation (Kelley et al., 1998)” is a physically motivated technique that follows the physical transient solution from a uniform flow. Here, a delicate choice of the time steps (or CFL numbers) is crucial to reduce the number of iterations. However, existing strategies such as SER or an exponential law are not robust; they produce an excessive number of iterations if some initial parameter is not carefully chosen.

In this paper, we investigate this problem and propose an *expert-system* for the automatic determination of the CFL numbers. This expert-system makes use of several numerical and geometrical metrics. Numerical results obtained on a wide variety of test cases show a systematic reduction of the total number of iterations compared to SER or EXP. Furthermore, the system seems more robust in the sense that (a) iteration counts depend only slightly on the parameters, and (b) breakdowns are avoided even if the parameters are overestimated.

1. Introduction

In the absence of a good initial approximation, Newton’s method alone will generally not suffice to obtain the steady state flow modelled by a set of partial differential equations. The usual alternative approach is to rewrite

the equations as a time dependent problem

$$\frac{\partial U}{\partial t} = -R(U), \quad (1)$$

and to march to the steady state by following the physical transient (see (Kell ley et al., 1998) and the references therein).

In this paper, we consider only implicit methods where the time is discretized by a backward Euler technique. Hence, every non-linear iteration k requires the (approximate) solution of a linear system of the form

$$\left(\frac{1}{\gamma^{(k)}} D^{(k)} + J^{(k)} \right) \Delta U^{(k)} = -R(U^{(k)}), \quad (2)$$

where the matrix $J^{(k)}$ is the Jacobian evaluated at $U^{(k)}$ and $\Delta U^{(k)} = U^{(k+1)} - U^{(k)}$ is the correction to the approximate solution. Since the transient states are not of interest, a local time stepping technique may be used, stabilized by a so-called CFL number (denoted $\gamma^{(k)}$) while the diagonal matrix $D^{(k)}$ contains the local time steps.

In this paper, we focus on the problem of finding a *robust* strategy for the determination of $\gamma^{(k)}$ ensuring a *quasi-minimum* number of iterations to reach the steady-state solution. This problem is heuristic in nature and therefore, we propose an *expert-system* that attempts to understand the features of the problem. This system keeps track of important metrics of the iteration history (such as the norm of the residual function) to decide whether to increase or decrease $\gamma^{(k)}$ and by which amount. Robustness is obtained by detecting possible breakdowns at early stages and taking smaller subsequent CFL numbers. Furthermore this new expert system does not depend on some initial parameters which denotes, again, an increased robustness.

2. Expert-System

2.1. EXISTING APPROACHES

Implicit iterative methods defined by (2) are known to remain stable even for large values of $\gamma^{(k)}$. Yet, there is no clear indication about the optimal time stepping strategies. Moreover, an inadequate choice of the CFL number may result in a excessive number of iterations or in a breakdown of the iteration process. Some previous attempts have already been made for the automatic determination of the CFL number and we refer the reader to the literature (Chakravarthy, 1988; Drikakis et al., 1994; Lin et al., 1995; Chen et al., 1993).

Two strategies are usually employed:

Switch Evolution Relaxation: The Switch Evolution Relaxation (SER) strategy

$$\gamma^{(k)} = \gamma_{SER} \left(\|R(U^{(k)})\|_2 \right)^{-1} \quad (3)$$

is a common technique that increases the CFL number as one approaches convergence (Mulder et al., 1985). Unfortunately, the initial parameter γ_{SER} is crucial for the successful application of the strategy. Another drawback encountered with SER is the presence of oscillations.

Exponential law: The exponential law (EXP)

$$\gamma^{(k)} = (\gamma_{EXP})^k \quad (4)$$

limits the oscillations (Johan et al., 1991). The value of the initial parameter is also crucial. For some problem, a very small γ_{EXP} may be required to avoid breakdowns during the initial phase which increases the total number of non-linear iterations.

2.2. EXPERT-SYSTEM

At the beginning of the iteration, complex flow features such as shocks are not resolved or not correctly placed. The CFL number should therefore be kept relatively small to avoid unrealistic transient solutions that would later prevent convergence. Worse, breakdowns may occur in this phase when some nodes have negative density, pressure, or energy. At the other end, when one approaches convergence, the approximate solution enters the radius of convergence of the Newton method and large CFL numbers may be safely taken. These considerations define the key ingredients for an effective and robust strategy: (a) breakdowns or possible breakdowns need to be detected to allow a reduction of the CFL number, and (b) the convergence rate must be estimated to increase the values of $\gamma^{(k)}$ if it is too small.

We have chosen a piecewise exponential function to satisfy these goals. In its simplest form, the CFL numbers are determined by only two exponential functions

$$\gamma^{(k)} = \begin{cases} \gamma_0 \gamma^{(k-1)} & \text{if } k < k_{switch}, \\ \gamma_1 \gamma^{(k-1)} & \text{otherwise.} \end{cases} \quad (5)$$

The first function corresponds to the initial phase and γ_0 is close to 1. In the terminal phase, determined by k_{switch} , the second exponential grows much faster, speeding up the convergence process.

In general, equation (5) is not sufficient to avoid breakdowns in the initial phase. It is therefore necessary to adapt $\gamma^{(k)}$ and/or γ_0 .

The terminal phase: There is no clear definition to decide whether an iterate is close to the solution or not. Also, with SER, we have observed that $\|R(U)\|_2$ alone is not sufficient since it may lead to oscillations.

In the expert-system, several metrics are combined into a single $M^{(k)}$ that estimates the “quality” of the iterate:

- the relative norm of the residual $\|R(U^{(k)})\|_2$,
- the relative correction $\|\Delta U^{(k)}/U^{(k)}\|_2$ where the division is to be taken as a component-wise division,
- the speed at which the norm of the correction tends to 0,
- the percentage of the geometry where a correction occurs, measured by the number of non-zero components of $\Delta U^{(k)}$.

Because the problem is not continuous in nature, the function is irregular and a moving average over the last few iterations is computed to give a smoothing effect. When convergence begins, $M^{(k)}$ becomes an increasing function and the iterative process is said to enter in the terminal phase when $M^{(k)}$ reaches a given threshold.

Breakdowns: Selecting one single exponential during the initial phase is usually not sufficient. Indeed, a large γ_0 may result in a breakdown while a small γ_0 produces excessive iterations. The expert-system automatically adapt γ_0 and decreases the CFL number to avoid breakdowns still constantly trying to speed up convergence:

If a breakdown occurs for $U^{(k)} + \Delta U^{(k)}$, then the iterate is not updated while the exponential is modified by

$$\gamma^{(k+1)} = \frac{1}{2}\gamma^{(k)}, \quad \text{and} \quad \gamma_0 = 1 + \frac{1}{2}(\gamma_0 - 1). \quad (6)$$

This simple procedure is repeated in case of successive breakdowns.

The previous technique alone does not guarantee convergence when a breakdown occurs because $U^{(k)}$ may already be too far from the physical transient. Therefore, if

$$\max \left| \frac{U^{(k+1)}}{U^{(k)}} \right| \geq 1.5 \quad \text{or} \quad \min \left| \frac{U^{(k+1)}}{U^{(k)}} \right| \leq 0.5, \quad (7)$$

then only half of the correction $\Delta U^{(k)}/2$ is accepted while the CFL number is decreased by a factor 2.

Finally, to avoid an unreasonable amount of iterations due to a value of γ_0 close to 1, the inverse transformation $\gamma_0 = 1 + 2(\gamma_0 - 1)$ is performed every few iterations. The number of iteration between two transformations depends on the time since the last breakdown.

3. Numerical experiments

The expert-system has been tested on a wide variety of test cases using the Euler, Navier-Stokes and MHD equations. Four unstructured geometries are considered, both in 2 and 3 dimensions: a Nozzle (5712 triangles, 3015 nodes), a Bow-shock (6152 triangles, 3203 nodes), a NACA airfoil (10924 triangles, 5590 nodes), an Ogive (1 039 022 tetrahedra, 177 846 nodes).

In the following, the three strategies are compared in terms of total number of non-linear iterations. Unless otherwise stated, the values of γ_{SER} and γ_{EXP} are determined — by trial and error — to give the lowest number of iterations while γ_0 is set to γ_{EXP} .

Typical result : Figure 1 presents the typical evolution of the CFL number and the convergence history for the solution of the Euler equations around a NACA airfoil ($M_\infty = 0.83$, $\alpha^\circ = 0$). To illustrate, we have set $\gamma_0 = 1.2 \gamma_{EXP}$.

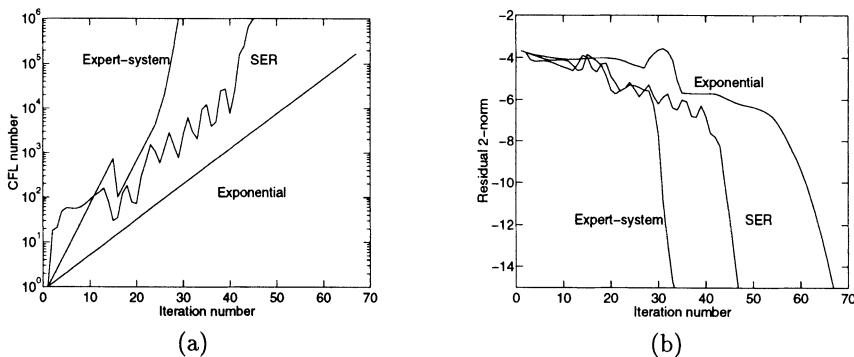


Figure 1. Convergence history of the transient solution of the Euler equations around a NACA airfoil for three CFL strategies: (a) CFL history, (b) residual norm history.

Obviously, the exponential law is not optimal. Indeed, the value of γ_{EXP} is underestimated as $U^{(k)}$ converges to the solution resulting in a slow convergence rate. However, taking a larger γ_{EXP} is not possible since it produces a breakdown in the initial phase.

For this problem, the SER strategy presents an oscillatory behavior slowing down the convergence. On the other hand, close to the solution, large CFL numbers appear and a steep convergence is observed.

By detecting that $U^{(k)}$ is close to the solution, the expert-system produces large CFL numbers earlier. Convergence happens after 38 iterations, which is 19% faster than SER.

Robustness : A proper choice of the initial parameters is crucial to limit the total number of non-linear iterations. Figure 2 shows the total number of

iterations as a function of the initial parameter for the Euler flow around a NACA airfoil. The ‘o’ symbol means that a converged solution was obtained while the ‘x’ symbol gives the iteration at which a breakdown occurred.

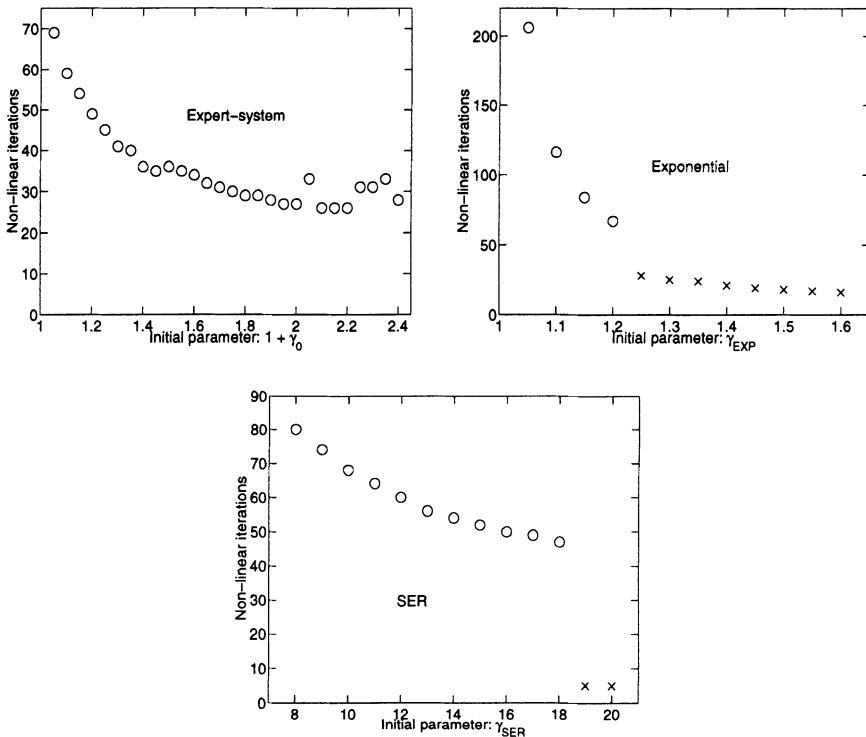


Figure 2. Robustness of the CFL strategies measured as the number of total non-linear iterations as a function of the initial parameter

For both SER and EXP, the number of iteration is a decreasing function up to a limit where convergence does not take place any more. For these two strategies, the decrease in iterations is steep. Hence, if a good estimate of the initial parameter is not available, the number of iteration can be severely increased.

The expert-system does not suffer for these weaknesses. Not only convergence was always observed for all our test cases, but also, the number of iterations remains almost constant for a large range of values of γ_0 .

Validation : Since the expert-system is heuristic in nature, its quality can only be validated empirically by comparison with other strategies on many problems. Table 1 presents iteration counts obtained on seven test cases for the three strategies. For each problem, the “optimal” initial parameter was chosen for SER and the exponential law.

Geometry	Equation	SER		Exponential		Expert-system	
		γ_{SER}	#iter	γ_{EXP}	#iter	γ_0	#iter
Nozzle	Euler	35	17	3.7	16	3.7	18
	MHD	14	17	4.3	14	4.3	14
Naca	Euler	18	47	1.20	67	1.20	51
	N-S	39	33	1.6	31	1.6	42
Bow	Euler	2.3	143	1.03	255	1.02	134
	MHD	3.1	158	1.03	196	1.03	134
Ogive	N-S	1.8	45	1.2	48	1.2	36

TABLE 1. Comparison of the three strategies for 7 test-cases.

On average, the expert-system provides a reduction in the total number of iterations. The reduction is more important when many iterations are needed (bow-shock).

If few iterations are sufficient, such as for the Nozzle problem, the expert-system gives results similar to the other strategies. Indeed, it needs at least 10 iterations to normalize the metrics and, therefore, can not improve the convergence rate. For the Navier-Stokes equations defined on the NACA geometry. Also, for this problem, a possible breakdown is (erroneously) detected. Smaller CFL numbers are thus taken and convergence requires 25% more iterations than SER or EXP.

4. Discussion and conclusion

We have presented a new strategy for the automatic adaptation of the CFL number during the iterative solution of hyperbolic PDE's by implicit upwind methods. This strategy contains many heuristics — hence the name *expert-system* — that attempt to avoid breakdowns of the iteration while providing an acceptable convergence rate. Compared to existing strategies such as the exponential law or the Switch Evolution Relaxation, the expert-system proves to be more efficient: (a) robustness is increased in the sense that the number of non-linear iterations is not much affected by the choice of parameters, and (b) convergence is usually faster with the system than with other strategies.

Based on our experience, the gain obtained by using the expert-system is larger when the problem is more complex, requiring many iterations or when the tolerance criterion on the residual vector is small. On the other side, when a maximum value of the CFL numbers is needed for stability reasons, or when convergence can not be obtained to a good accuracy, the benefits of the expert-system will remain limited.

Acknowledgements

The author is grateful to Arpad Csik, Kurt Sermeus, Herman Deconinck, and Stefaan Poedts (Center for Plasma-Astrophysics and/or von Karman Institute) for providing workable versions of the RA and THOR codes for the numerical solution of the Euler, Navier-Stokes and MHD equations.

This work is funded by the *Fonds voor Wetenschappelijk Onderzoek* (FWO project G0344.98: “Multidimensional upwinding and parallel implicit solvers for MHD”) and by the UIAP P4/02 Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Minister’s Office for Science Technology and Culture. The scientific responsibility rests with its author.

References

- Essers J A, Delanaye M, and Rogiest P (1995). An Upwind Biased Finite Volume Technique for Solving Compressible Navier-Stokes Equations on Irregular Meshes. *AIAA Journal*, **33**, No. 5, pp. 833–842.
- Johan Z, Hughes T J R, and Shakip F (1991). A Globally Convergent Matrix-Free Algorithm for Implicit Time-Marching Schemes Arising in Finite Element Analysis in Fluids. *Comp. Meth. in Appl. Mech. and Eng.*, **87**, pp. 281–304.
- Kelley C T and Keyes D E (1998). Convergence Analysis of Pseudo-Transient Continuation. *SIAM J. Num. Anal.*, **35**, No. 2, pp. 508–523.
- Mulder W and van Leer B (1985). Experiments with Implicit Upwind Methods for the Euler Equations. *J. Comput. Phys.*, **59**, pp. 232–402.
- Chakravarthy S R (1988). High resolution upwind formulations for the Navier-Stokes equations. VKI Lecture Series, Computational Fluid Dynamics, 1988-05.
- Drikakis D and Durst F (1994). Parallelization of inviscid and viscous flow solvers. *Int. J. of Computational Fluid Dynamics*, **3**, 101-121.
- Lin H, Yang D Y and Chieng CC (1995). Variants of Biconjugate Gradient Method for Compressible Navier-Stokes Solver. *AIAA J.*, **33**, No 7, 1177- 1184.
- Chen J P and Whitfield D L (1993). Navier-Stokes Calculations for the unsteady flow-field of turbomachinery. AIAA-93-0676, 31st Aerospace Sciences Meeting, Jan. 11-14, Reno, Nevada.

DISCONTINUOUS GALERKIN METHODS FOR HYPERBOLIC PARTIAL DIFFERENTIAL EQUATIONS

J. J. W. VAN DER VEGT

*Faculty of Mathematical Sciences
University of Twente*

*P.O. Box 217, 7500 AE, Enschede, The Netherlands
e-mail : j.j.w.vandervegt@math.utwente.nl*

AND

H. VAN DER VEN, O.J. BOELENS

*National Aerospace Laboratory NLR
P.O. Box 90502, 1006 BM Amsterdam, The Netherlands
e-mail : venvd@nlr.nl, boelens@nlr.nl*

Abstract. In this paper a survey is given of the important steps in the development of discontinuous Galerkin finite element methods for hyperbolic partial differential equations. Special attention is paid to the application of the discontinuous Galerkin method to the solution of the Euler equations of gas dynamics in time-dependent flows domains and to techniques which reduce the computational complexity of the DG method.

1. Introduction

Finite element methods provide a well developed mathematical framework for the solution of elliptic partial differential equations. For an extensive survey, see for instance (Ciarlet and Lions, 1991). This has motivated the use of finite element methods also for the solution of hyperbolic partial differential equations, but this is not as straightforward as for elliptic pde's. A significant source of problems is caused by the fact that hyperbolic partial differential equations can develop non-smooth solutions, even if the initial data are very smooth. The standard Galerkin finite element discretization for hyperbolic partial differential equations results in oscillatory solutions around discontinuities and complicated stabilization and discontinuity cap-

turing operators are necessary. These methods have resulted in Petrov-Galerkin, streamline upwind Petrov Galerkin (SUPG), and more recently Galerkin Least Squares (GLS) finite element methods. Important contributions to the development of these methods can be found in the series of papers of (Masud and Hughes, 1997; Shakib, Hughes and Johan, 1991; Hansbo and Johnson, 1995; Johnson, Szepessy, and Hansbo, 1990) and the references therein.

Discontinuous Galerkin finite element methods provide an interesting alternative to these methods and combine many of the advantages of finite volume and finite element methods. The concept of discontinuous Galerkin finite element methods has been known for quite some time, and dates back to the early 70s for hyperbolic partial differential equations, but only recently have discontinuous Galerkin methods been applied to real applications. The recent interest in discontinuous Galerkin methods is motivated by the need to develop more accurate numerical discretizations on unstructured adaptive meshes, and to use higher order accurate discretizations for acoustic problems and the direct and large eddy simulation of turbulent flows. The growing interest in discontinuous Galerkin methods was one of the motivations to have a special conference on this topic in Newport, Rhode Island, 1999.

The recent interest in discontinuous Galerkin methods is motivated by some interesting features of this discretization technique. Discontinuous Galerkin finite element methods use a local polynomial representation of the solution and test functions in each element, without requiring continuity across element faces. This results in equations for the polynomial expansion coefficients of the solution which are uncoupled from neighboring elements. An important benefit of this approach is that for higher order accuracy it is not necessary to use complicated reconstruction algorithms to obtain pointwise data from cell averaged data for the flux calculation, as is necessary for finite volume methods. Discontinuous Galerkin methods result in a very local discretization and combine well with grid adaptation using local grid refinement, and parallel computations.

Discontinuous Galerkin methods make it possible to incorporate successful upwind schemes into finite element methods, because the discontinuity in the polynomial representation at the element faces can be interpreted as a Riemann problem, which is a Cauchy initial value problem with two discontinuous initial states. The use of (approximate) Riemann solvers makes it possible to incorporate important physical information into the numerical discretization and results in a robust upwind scheme, which has become very popular for upwind finite volume techniques. An excellent survey of these methods can be found in (Toro, 1997). Another interesting

feature of discontinuous Galerkin finite element discretizations is that they result in an element-wise conservative scheme, whereas SUPG and Galerkin least squares methods only result in globally conservative schemes.

The outline of this paper is as follows. First a survey is presented of the main aspects of discontinuous Galerkin finite element methods for hyperbolic partial differential equations. An attempt will be made to give a critical assessment of discontinuous Galerkin methods, but limited space prevents discussing all aspects in great detail. Recent detailed surveys of discontinuous Galerkin methods can be found in (Cockburn, 1999; Cockburn, Karniadakis and Shu, 2000; Barth, 1998). In the second part of this paper it will be demonstrated how the discontinuous Galerkin finite element method can be used to solve the Euler equations of gas dynamics in time-dependent flow domains using translating-rotating reference frames. This technique is useful to simulate aerodynamic problems, such as aircraft maneuver and propellers. The straightforward use of a discontinuous Galerkin method results, however, in a prohibitively expensive numerical scheme and special attention will be paid to techniques which significantly reduce the computational complexity of discontinuous Galerkin methods and result in a practical algorithm for computational fluid dynamics. The discussion in this paper is mainly limited to discontinuous Galerkin discretizations in space in combination with a Runge-Kutta time integration method. A discussion of the time-discontinuous and space-time discontinuous Galerkin methods is beyond the scope of the present paper.

2. Survey of Discontinuous Galerkin Methods

Discontinuous Galerkin finite element methods have been around for a long time. Their application to elliptic second and fourth order problems already started in the early 1960's, for a brief survey see (Oden, Babuska, and Bammann, 1998). The application to hyperbolic partial differential equations started much later with the work of (Reed and Hill, 1973) for the neutron transport equation for the flux of neutrons $u(\mathbf{x}) \in R^2$:

$$\begin{aligned} \operatorname{div}(\boldsymbol{\mu} u) + \sigma u &= f, & \mathbf{x} \in \Omega \subset R^2, \\ u &= 0, & \text{at } \partial_-\Omega, \end{aligned}$$

with $\boldsymbol{\mu}$ a constant vector, $\partial_-\Omega$ the part of the boundary $\partial\Omega$ with $\boldsymbol{\mu} \cdot \mathbf{n} < 0$, with \mathbf{n} the outward normal vector at $\partial\Omega$, and $\sigma \in R$. This equation is linear and can be solved by a marching scheme if the elements follow the characteristic lines. The first theoretical analyses of the discontinuous Galerkin method were presented by (Lesaint and Raviart, 1974) and (Johnson and Pitkäranta, 1986) for the neutron transport and linear advection equations.

The application of discontinuous Galerkin methods to non-linear scalar hyperbolic conservation laws:

$$\partial_t u + \partial_x f(u) = 0, \quad (x, t) \in R \times (0, T)$$

is more complicated because the marching scheme cannot be applied in general, since the direction of propagation $(1, f'(u))$ is part of the solution. Chavent and Salzano (1982) presented the P^0P^1 -DG method for the one-dimensional water flooding problem with gravity. This method uses a piecewise constant polynomial representation in time and a linear representation in space. This explicit scheme can be easily extended to more spatial dimensions, but has a severe stability restriction and is not suitable for practical calculations.

A significant step forward was made by (Chavent and Cockburn, 1989) which introduced a monotonicity preserving slope limiter, similar to the MUSCL scheme of (van Leer, 1974), and the use of a Godunov flux to account for the discontinuities at element faces. They proved that the scheme satisfies a maximum principle and is total variation bounded in the means (TVBM). Linear stability analysis however shows that the numerical scheme is only stable if the CFL number satisfies the condition $cfl \cong O(h^{\frac{1}{2}})$ as the element length $h \downarrow 0$.

This time step limitation can be alleviated by using the total variation diminishing Runge-Kutta schemes introduced by (Shu and Osher, 1988). This has been the topic of a series of papers by (Cockburn and Shu, 1988; Cockburn and Shu, 1989a; Cockburn, Lin and Shu, 1989b; Cockburn, Hou and Shu, 1990; Cockburn and Shu, 1991) in which they studied the $RK\Lambda\Pi P^k$ method. The $RK\Lambda\Pi P^k$ method is a combination of a TVD Runge-Kutta time integration method with a discontinuous Galerkin finite element discretization in space using polynomials of degree k . The local projection limiter $\Lambda\Pi$ is used to ensure monotonicity of the solution. It was shown by (Cockburn and Shu, 1988; Cockburn and Shu, 1991) that this method is stable and second order accurate in space and time for $k = 1$ when applied to scalar conservation laws and if the CFL number is chosen in the range $cfl \in [0, \frac{1}{3}]$. Cockburn and Shu extended the analysis of the $RK\Lambda\Pi P^k$ method to polynomials of degree $k \geq 1$ in (Cockburn and Shu, 1989a) and showed that this method is TVBM and converges to the entropy solution for the scalar conservation law. The order of accuracy of the $RK\Lambda\Pi P^k$ method is $k + 1$, except at critical points. Special attention is paid to the design of a TVB limiter in order to minimize the effects of the limiter in smooth parts of the flow. The use of a TVB limiter for practical applications is, however, not straightforward since it is difficult to

estimate the coefficients in the limiter, which depend on the derivatives of the solution and the mesh size.

The extension to one-dimensional systems is discussed in (Cockburn, Lin and Shu, 1989b), where the application to the Euler equations of gas dynamics is discussed in detail. Special attention was paid to the construction of the slope limiter and the most successful and robust approach uses the characteristic decomposition, which essentially decouples the equations in a system of scalar conservation laws. The theoretical framework of the discontinuous Galerkin discretization for multi-dimensional scalar conservation laws, in combination with the TVD Runge-Kutta time integration method, was firmly established in (Cockburn, Hou and Shu, 1990). Their main results can be summarized as:

- If the quadrature rules for the flux integrals over the element faces and volumes in the discontinuous Galerkin discretization are exact for polynomials of degree $2k + 1$ and $2k$, respectively, then the difference between the discrete approximation $L_h(u_h, \gamma_h)$ of $-\operatorname{div} \mathbf{f}(u)$ (including the boundary operator γ_h) can be estimated as:

$$\| L_h(u_h, \gamma_h) + \operatorname{div} \mathbf{f}(u) \|_{L^\infty(\Omega)} \leq C h^{k+1} |\mathbf{f}(u)|_{W^{k+2,\infty}(\Omega)},$$

with \mathbf{f} the flux function in the multi-dimensional scalar conservation law $\partial_t u + \partial_x \mathbf{f}(u) = 0$ and $W^{k,p}(\Omega)$ the Sobolev seminorm, see for instance (Ciarlet and Lions, 1991).

- Let the coefficients α_{il} of the Runge-Kutta time discretization be positive and such that $\sum_{l=0}^{i-1} \alpha_{il} = 1$, for $i = 1, \dots, k+1$. Set $w_h = u_h + \Delta t L_h(u_h, \gamma_h)$, and suppose that the following maximum principle is satisfied:

$$\bar{u}_h, \gamma_h \in [a, b] \Rightarrow \bar{w}_h \in [a - Mh^2, b + Mh^2], \quad (1)$$

where M is some nonnegative parameter and a overbar denotes the element mean value, then:

$$\bar{u}_h^n \in [a - (k+1)mMh^2, b + (k+1)mMh^2], \quad \text{form } m = 0, \dots, n.$$

if $\operatorname{cfl} \in [0, \operatorname{cfl}_0 / |\max_{i,l} \{|\frac{\beta_{il}}{\alpha_{il}}\}|]$. The coefficients cfl_0 , a_0 and b_0 are defined as:

$$\begin{aligned} \operatorname{cfl}_0 &= \sup_{n=1, \dots, N; e \in \partial K; K \in \mathcal{T}_h} \Delta t \frac{|e|}{|K|} \| \mathbf{f}' \cdot \mathbf{n} \|_{L^\infty[a_0, b_0]}, \\ a_0 &= \inf_{\mathbf{x} \in \Omega, t \in (0, t^{n+1}), \mathbf{y} \in \partial \Omega} \{u_0(\mathbf{x}), \gamma(t, \mathbf{y})\}, \\ b_0 &= \sup_{\mathbf{x} \in \Omega, t \in (0, t^{n+1}), \mathbf{y} \in \partial \Omega} \{u_0(\mathbf{x}), \gamma(t, \mathbf{y})\}, \end{aligned}$$

with $|e|$ and $|K|$ the length of the edges and the volume of element K with unit outward normal vector \mathbf{n} . This result shows that the combination of a TVD Runge-Kutta time integration method and a slope limiter satisfying a TVB condition results in a stable discretization with a reasonable CFL number restriction. The use of a TVD Runge-Kutta time integration method is essential and results in a very robust discretization. If one uses Runge-Kutta schemes which are not TVD then the numerical scheme will suffer from severe stability restrictions.

- For a general class of triangulations (**B**-triangulations) a $\Lambda\Pi_h$ projection was defined which satisfies the maximum principle (1).

The series of articles of (Cockburn and Shu, 1988; Cockburn and Shu, 1989a; Cockburn, Lin and Shu, 1989b; Cockburn, Hou and Shu, 1990; Cockburn and Shu, 1991) give the discontinuous Galerkin finite element method in combination with a TVD-Runge Kutta time integration method a solid mathematical background. The first applications to two-dimensional problems can be found in (Bey and Oden, 1991), which applied the method to the supersonic flow about a compression corner and a forward facing step. Special attention was paid to the local projection limiter for higher order elements, but the limiter resulted in significant smearing of discontinuities. (Lin and Chin, 1993) used a second order discontinuous Galerkin discretization for the Euler and Navier-Stokes equations and applied the method to the transonic flow in a channel with a circular bump and an oscillating NACA 0012 profile.

The use of a discontinuous Galerkin finite element method to three-dimensional problems, in particular in computational fluid mechanics, is fairly recent. As mentioned in the introduction this is motivated by the very local behavior of the discontinuous Galerkin discretization, which makes it a good candidate for use on unstructured meshes. In a series of articles (van der Vegt, 1995a; van der Vegt and van der Ven, 1995b; van der Vegt and van der Ven, 1998a) demonstrated that the discontinuous Galerkin method can be used for practical aerodynamic calculations in combination with grid adaptation using local grid refinement. Calculations on the ONERA M6 wing and a delta wing, both in transonic flow, showed that significant improvement in the capturing of shocks and vortical structures can be obtained. Also, a face based dynamic data structure, suitable for an adaptive discontinuous Galerkin discretization, was discussed and implemented. A general discontinuous Galerkin method for arbitrary Lagrangian Eulerian hydrodynamics was developed by (Kershaw, Prasad, Shaw and Milovich, 1998). Special attention was paid to the development of a slope limiter and it was demonstrated that DG methods combine very well with object oriented programming.

Three-dimensional applications, both for the Euler and Navier-Stokes equations in combination with local grid refinement, are also demonstrated by (Baumann, 1997; Baumann and Oden, 1999), which studied the Shuttle Orbiter at a Mach number of 7.4. The most important contribution in this work is, however, the theoretical analysis of discontinuous Galerkin methods using broken spaces and the extension to the Navier-Stokes equations, see also (Oden, Babuska, and Baumann, 1998). This theoretical analysis is closely linked to the work of (Bey and Oden, 1996), which presented a very complete a posteriori error estimate for the neutron transport equation.

The application of discontinuous Galerkin finite element methods to three-dimensional problems requires, however, significant improvements in computational efficiency in order to be practical. The most significant computational expense is the use of Gauss quadrature rules for the calculation of the element face and volume integrals, because this requires for each Gauss quadrature point the calculation of the flux. Several approaches to alleviate this problem have been proposed. If only second order accuracy is required (van der Vegt and van der Ven, 1998a) demonstrated that using special element and face averages, in combination with the exact calculation of the geometric contributions, results in a practical second order accurate scheme. This method reduces the number of flux evaluations to one per element face and is analyzed in (van der Vegt and van der Ven, 1998a; van der Vegt and van der Ven, 2000; van der Ven and van der Vegt, 2000).

An alternative approach, which is not limited to second order accuracy, is followed by (Atkins and Shu, 1998). They expand the flux in terms of the basis functions used to represent the solution and integrate this representation analytically. This results in a significant reduction in the number of flux evaluations, but due to the non-linearity of the flux, and also because contributions of ρ^{-1} are required, it is not possible to obtain an exact representation for the flux. Also, the use of more advanced upwind schemes is non-trivial and Atkins and Shu therefore limit themselves to the relatively simple Lax-Friedrichs flux. The flux integrals for the Euler equations of gas dynamics can also be evaluated with the procedure used by (Lowrie, Roe, and van Leer, 1995) which uses the parameter vector $\mathbf{w} = \sqrt{\rho}(1, \mathbf{u}, H)^T$, with ρ the density, \mathbf{u} the fluid velocity, and H the total enthalpy. The flux vector and conservative variables then are quadratic functions of \mathbf{w} . This method, however, requires a significant amount of storage for all the coefficients and is also difficult to use for advanced upwind schemes.

The application of the Runge-Kutta discontinuous Galerkin method to the Euler equations of gas dynamics in two-dimensions using a higher order accurate discretization is investigated by several people. Polynomials of degree $k = 1$ and $k = 2$ are used by (Cockburn and Shu, 1998a), which

show that the discontinuous Galerkin discretization is capable of capturing very fine details in the instability in the contact discontinuity inside a double Mach reflection and also behind a forward facing step. Discretizations up to fourth order accuracy were used by (Bassi and Rebay, 1997b), which also demonstrated that a superparametric representation of the elements at the boundary is necessary to prevent unphysical solutions. The use of higher order accurate discontinuous Galerkin methods for acoustics is investigated by (Hu, Hussaini, and Rasetarinera, 1999), which consider the dispersion and dissipation properties of the DG method using Fourier analysis.

Recently, (Lowrie and Morel, 1999) demonstrated that discontinuous Galerkin discretizations can be significantly more accurate than high resolution finite volume schemes for stiff hyperbolic systems. Discontinuous Galerkin finite element methods also combine very well with parallel computations as is demonstrated by (Biswas, Devine, and Flaherty, 1994; van der Ven and van der Vegt, 1997; van der Ven and van der Vegt, 1998).

Despite the significant progress made in the development of discontinuous Galerkin finite element methods there still are a number of important open issues. The two most important ones are the significant increase in memory use for higher order discretizations and the effects of the limiter. The use of a limiter prevents convergence to a steady state, because the limited slopes do not satisfy the discretized equations and tend to grow again, which results in limit cycle behavior. The limiter is also triggered by small disturbances and tends to be active in large parts of the flow field, causing a reduction in accuracy in smooth parts of the flow field. More efficient stabilization techniques are still a topic of active research. The memory use of DG methods rapidly increases with increasing order of the polynomials. It can be reduced by using more sophisticated basis functions, such as Legendre polynomials in one dimension, but for general elements this still is an open issue.

3. Discontinuous Galerkin Discretization of the Euler Equations in Moving Flow Domains

In the remaining part of this paper the discontinuous Galerkin finite element method will be applied to the Euler equations of gas dynamics in three-dimensional time-dependent flow domains. The Euler equations of gas dynamics are an important example of hyperbolic partial differential equations, and the use of a discontinuous Galerkin finite element method is non-trivial due to the occurrence of non-smooth solutions, such as shocks and contact discontinuities.

3.1. ALE WEAK FORMULATION OF THE EULER EQUATIONS

The calculation of the flow field in moving and deforming flow domains can be done efficiently using an arbitrary Lagrangian Eulerian (ALE) formulation. In this section we will discuss the ALE weak formulation for the Euler equations of gas dynamics using translating-rotating coordinate systems. This approach has important applications in several areas, such as aircraft maneuver and propellers. The ALE formulation, both for finite volume and finite elements methods, is analyzed in detail by (Lesoinne and Farhat, 1996). Special attention is paid to the geometric conservation law (GCL). The GCL was formulated by (Thomas and Lombard, 1979) and states that a uniform flow field should remain uniform under mesh movement and deformation. This imposes restrictions on the way grid velocities are evaluated in finite volume methods. For finite element discretizations the GCL condition can, however, be satisfied relatively easy if one calculates the element integrals with sufficient accuracy.

The Euler equations of gas dynamics at a point (\mathbf{x}, t) fixed in space and time are defined as:

$$\frac{\partial \mathbf{U}(\mathbf{x}, t)}{\partial t} + \frac{\partial \mathbf{F}_j(\mathbf{U}(\mathbf{x}, t))}{\partial x_j} = 0,$$

where the vectors with conserved flow variables $\mathbf{U} : R^3 \times [t_0, T] \rightarrow R^5$, and the fluxes \mathbf{F}_j , ($j = 1, 2, 3$); $\mathbf{F}_j : R^5 \rightarrow R^5$, are defined as:

$$\mathbf{U} = \begin{pmatrix} \rho \\ \rho u_i \\ \rho E \end{pmatrix}; \quad \mathbf{F}_j = \begin{pmatrix} \rho u_j \\ \rho u_i u_j + p \delta_{ij} \\ u_j (\rho E + p) \end{pmatrix},$$

with $i \in \{1, 2, 3\}$, and ρ, p, E denote the density, pressure, and specific total energy, u_i the velocity component in the Cartesian coordinate direction x_i of the velocity vector $\mathbf{u} : \Omega \times [t_0, T] \rightarrow R^3$, and δ_{ij} the Kronecker delta symbol. The Euler equations are complemented with initial and boundary conditions:

$$\begin{aligned} \mathbf{U}(\mathbf{x}, t_0) &= \mathbf{U}_0(\mathbf{x}), & \mathbf{x} \in \Omega(t_0), \\ \mathbf{U}(\mathbf{x}, t) &= \mathcal{B}(\mathbf{U}, \mathbf{U}_w, t), & t \in [t_0, T], \mathbf{x} \in \partial\Omega(t), \end{aligned}$$

with \mathbf{U}_w the boundary operator, and the equation of state for an ideal gas $p = (\gamma - 1)\rho(E - \frac{1}{2}u_i u_i)$, with γ the ratio of specific heats at constant pressure and volume.

Introduce a translating-rotating coordinate system $O_1 X_1 Y_1 Z_1$ relative to the inertial coordinate system $O_0 X_0 Y_0 Z_0$. The relation between between

points in both coordinate systems is:

$$\mathbf{x}(t) = \mathcal{C}(t) \mathbf{x}^{(1)} + \mathbf{r}(t), \quad (2)$$

with \mathcal{C} the rotation matrix between the inertial and rotating reference frame, and \mathbf{r} the position of the origin of the coordinate system $O_1X_1Y_1Z_1$ relative to $O_0X_0Y_0Z_0$. The superscript (1) indicates that a vector has components relative to the coordinate systems $O_1X_1Y_1Z_1$, otherwise the components are always with respect to $O_0X_0Y_0Z_0$. The Jacobian of the transformation between the two coordinate systems is equal to one. The flow domain $\Omega(t)$ becomes independent of time when expressed relative to the translating-rotating reference frame $O_1X_1Y_1Z_1$ and is denoted $\Omega^{(1)}$. The ALE weak formulation of the Euler equations can be obtained from the general ALE weak formulation discussed by (Lesoinne and Farhat, 1996) using the coordinate transformation (2):

Find a $\mathbf{U} \in [L^1(\Omega)]^5$, such that for all $\mathbf{W} \in [C^1(\Omega)]^5$, we have:

$$\begin{aligned} & \frac{d}{dt} \int_{\Omega^{(1)}} \mathbf{W}^T(\mathbf{x}^{(1)}) \mathbf{U}(\mathbf{x}^{(1)}, t) d\Omega^{(1)} \\ & + \int_{\partial\Omega^{(1)}} \mathbf{W}^T(\mathbf{x}^{(1)}) (n_k(\mathbf{x}^{(1)}, t) \mathbf{F}_k(\mathbf{U}) - \mathbf{n}(\mathbf{x}^{(1)}, t) \cdot \mathbf{s}(\mathbf{x}^{(1)}, t) \mathbf{U}(\mathbf{x}^{(1)}, t)) d(\partial\Omega^{(1)}) \\ & - \int_{\Omega^{(1)}} \frac{\partial \mathbf{W}^T(\mathbf{x}^{(1)})}{\partial x_k^{(1)}} (\mathbf{F}_p(\mathbf{U}) \mathcal{C}_{pk}(t) - s_k^{(1)}(\mathbf{x}^{(1)}, t) \mathbf{U}(\mathbf{x}^{(1)}, t)) d\Omega^{(1)} = 0. \end{aligned} \quad (3)$$

The components of the normal vector \mathbf{n} and grid velocity $\mathbf{s} = d\mathbf{x}/dt$ in both coordinates systems are related as: $n_k^{(1)} = \mathcal{C}_{jk} n_j$ and $s_k^{(1)} = \mathcal{C}_{jk} s_j$. The use of a mixed formulation, with vectors having components in both the inertial and the translating-rotating reference frame, has as main benefit that the conservation form is maintained and no source terms, such as the Coriolis and centrifugal forces, appear.

3.2. DISCONTINUOUS GALERKIN FINITE ELEMENT DISCRETIZATION

The flow domain $\Omega^{(1)}$ is discretized using a triangulation \mathcal{T}_h with elements K . As basic elements in \mathcal{T}_h we use hexahedrons and each of the elements $K \in \mathcal{T}_h$ is related to the cubic master element $\hat{K} = [-1, 1]^3$ by means of the isoparametric transformation F_K :

$$F_K : \hat{\mathbf{x}} = (\xi, \eta, \zeta)^T \in \hat{K} \rightarrow \mathbf{x}^{(1)} \in K : \mathbf{x}^{(1)}(\xi, \eta, \zeta) = \sum_{i=1}^{m_K} \mathbf{x}_i^{(1)} \chi_i(\hat{\mathbf{x}}), \quad (4)$$

with $m_K = 8$ for hexahedrons and χ_i the standard tri-linear basis functions. The discontinuous Galerkin discretization is now obtained using a sequence a function spaces:

- Define $P^k(\hat{K})$ as the space of polynomial functions of degree $\leq k$ on the master element \hat{K} : $P^k(\hat{K}) = \text{span}\{\hat{\phi}_j, j = 0, \dots, M\}$. In this paper M is restricted to 3, so the four basis functions $\hat{\phi}_j$ are: $\hat{\phi}_j = 1, \xi, \eta, \zeta, (j = 0, \dots, 3)$.
- Define $P^k(K)$ as the space of functions associated to functions in $P^k(\hat{K})$ through the mapping F_K : $P^k(K) = \text{span}\{\phi_j = \hat{\phi}_j \circ F_K^{-1}, j = 0, \dots, 3\}$.
- Define $\mathbf{V}_h^1(K) = \{\mathbf{P}(K) = (p_1, \dots, p_5)^T \mid p_i \in P^1(K)\}$.

It is important to realize that the polynomial expansions in each element are purely local without any connection to neighboring elements. This is the main difference of the discontinuous Galerkin finite element method with standard node based Galerkin methods. The approximate flow field \mathbf{U}_h can be defined using the basis functions ϕ_j , but for the definition of the slope limiter and the multigrid convergence acceleration it is beneficial to split \mathbf{U}_h into an element mean $\bar{\mathbf{U}}_h$ and a fluctuation $\tilde{\mathbf{U}}_h$. This can be accomplished by splitting $\mathbf{V}_h^1(K)$ into two spaces: $\bar{\mathbf{V}}_h^0(K)$ and $\tilde{\mathbf{V}}_h^1(K)$, such that:

$$\mathbf{V}_h^1(K) = \bar{\mathbf{V}}_h^0(K) \oplus \tilde{\mathbf{V}}_h^1(K),$$

with :

$$\bar{\mathbf{V}}_h^0(K) = \{\mathbf{P}(K) = (p_1, \dots, p_5)^T \mid p_i \in P^0(K)\},$$

$$\tilde{\mathbf{V}}_h^1(K) = \{\mathbf{P}(K) = (p_1, \dots, p_5)^T \mid \int_K p_i dK = 0, p_i \in P^1(K)\}.$$

The element mean flow field $\bar{\mathbf{U}}_h \in \bar{\mathbf{V}}_h^0(K) \otimes C^1[0, T]$ can now be defined as:

$$\bar{\mathbf{U}}_h(t) = \frac{1}{|K|} \int_K \mathbf{U}(\mathbf{x}^{(1)}, t) dK,$$

and the flow field fluctuations $\tilde{\mathbf{U}}_h \in \tilde{\mathbf{V}}_h^1(K) \otimes C^1[0, T]$ as:

$$\tilde{\mathbf{U}}_h(\mathbf{x}^{(1)}, t) = \mathbf{U}_h(\mathbf{x}^{(1)}, t) - \bar{\mathbf{U}}_h(t) = \sum_{m=1}^3 \hat{\mathbf{U}}_m(K, t) \psi_m(\mathbf{x}^{(1)}),$$

with the basisfunctions $\psi_m \in P^1(K)$ defined as:

$$\psi_m(\mathbf{x}^{(1)}) = \phi_m(\mathbf{x}^{(1)}) - \frac{1}{|K|} \int_K \phi_m(\mathbf{x}^{(1)}) dK, \quad m = 1, 2, 3.$$

The element mean and fluctuating flow fields are orthogonal with respect to the L_2 inner product, and this relation can be used to simplify the discretized equations. The test functions \mathbf{W}_h are also split into an element mean and a fluctuating part, $\bar{\mathbf{W}}_h \in \bar{\mathbf{V}}_h^0(K)$ and $\tilde{\mathbf{W}}_h \in \tilde{\mathbf{V}}_h^1(K)$.

The discontinuous Galerkin finite element discretization for the Euler equations in a translating-rotating reference frame is obtained if we introduce the polynomial representations for $\bar{\mathbf{U}}_h$ and $\tilde{\mathbf{U}}_h$, and the equivalent expressions for the test functions $\bar{\mathbf{W}}_h$ and $\tilde{\mathbf{W}}_h$, into the weak formulation (3):

$$|K| \frac{d\bar{U}_i(K)}{dt} = - \int_{\partial K} (n_k F_{ik}(\mathbf{U}_h) - n_k s_k U_i) d(\partial K) \equiv -\bar{R}_i(K) \quad (5)$$

$$\begin{aligned} M_{nm} \frac{d\tilde{U}_{mi}(K)}{dt} &= - \int_{\partial K} \phi_n (n_k F_{ik}(\mathbf{U}_h) - n_k s_k U_i) d(\partial K) \\ &\quad + \int_K \frac{\partial \phi_n}{\partial x_k^{(1)}} \left(F_{ip}(\mathbf{U}_h) \mathcal{C}_{pk}(t) - s_k^{(1)} U_i \right) dK + \frac{1}{|K|} M_{n0} \bar{R}_i(K), \end{aligned} \quad (6)$$

with the matrix $M \in R^{3 \times 3}$ defined as:

$$M_{nm}(K) = \int_K \phi_n \phi_m dK.$$

The equation for the element mean flow field $\bar{\mathbf{U}}_h$ (5) is identical to the equations for a finite volume discretization and is only weakly coupled with the equation for the flow field fluctuations (6). The mass matrix of the moving elements in the translating-rotating flow domain needs to be calculated only once since the grid is not deforming. An analytic expression for the mass matrix can be found in (van der Vegt and van der Ven, 1998a). For use in the time integration method and the multigrid convergence acceleration scheme it is beneficial to express (5) and (6) symbolically as:

$$\mathcal{M} \frac{d}{dt} \hat{\mathbf{U}}_m = \mathcal{R}_m(\mathbf{U}_h), \quad m = 0, \dots, 3 \quad (7)$$

with $\mathcal{M} \in R^{4 \times 4}$ defined as:

$$\mathcal{M} = \begin{bmatrix} |K| & 0 \\ 0 & M \end{bmatrix}.$$

4. Flux Calculation

The discontinuous Galerkin discretization results in expressions for the flow field \mathbf{U}_h and test functions \mathbf{W}_h which are discontinuous at element faces. This discontinuity can be interpreted as a Riemann problem from gas dynamics and can be used to give a suitable definition for the flux at the

element boundary ∂K . This is accomplished by replacing the flux function with a numerical flux. Any of the well-known (approximate) Riemann solvers, such as those from Godunov, Roe, Lax-Friedrichs or Osher, can be used in the definition of the numerical flux. For an overview of (approximate) Riemann solvers for gas dynamics, see (Toro, 1997). This procedure introduces upwinding into the discontinuous Galerkin formulation and does not require the design of elaborate stabilization and discontinuity capturing operators.

In this paper the Osher scheme is used because it is a very accurate upwind scheme with good shock capturing capabilities. More details can be found in (Osher and Chakravarthy, 1983). The main difference in calculating the Osher flux for moving elements in comparison with non-moving elements is that the eigenvalues used in determining the path integrals in phase space must be corrected for the grid velocity. More details can be found in (van der Vegt and van der Ven, 2000). The Osher flux for moving element faces can be split into the standard Osher flux for non-moving grids \mathbf{H} and a part directly related to the grid velocity:

$$\mathbf{H}^c(\mathbf{U}_h^{\text{int}(K)}, \mathbf{U}_h^{\text{ext}(K)}) = \mathbf{H}(\mathbf{U}_h^{\text{int}(K)}, \mathbf{U}_h^{\text{ext}(K)}) - \mathbf{n} \cdot \mathbf{s} \mathbf{G}(\mathbf{U}_h^{\text{int}(K)}, \mathbf{U}_h^{\text{ext}(K)}),$$

where the flux vector \mathbf{G} is obtained by replacing the normal flux vector $n_k \mathbf{F}_k(\mathbf{U})$ in the Osher flux with \mathbf{U}_h .

If only second order accuracy is required then the flux integration does not require Gauss quadrature rules and can be done using special face and element flux averages, after which the geometric contributions can be calculated analytically or numerically. This method was proposed and analyzed by (van der Vegt and van der Ven, 1998a; van der Vegt and van der Ven, 2000) and is briefly described below. The main benefit of this method is that only one flux calculation for each element face is necessary and relatively simple and exact relations for the geometric contributions are obtained, which automatically satisfy the GCL.

The integral of the numerical flux \mathbf{H}^c over the element faces can be approximated using the following approximation to the flux integrals at the element boundary ∂K :

$$\begin{aligned} \int_{\partial K} \phi_n \mathbf{H}^c dK &\cong \frac{1}{2} \left(\mathbf{F}_j(\bar{\mathbf{U}}_h^{\text{int}(K)}) + \mathbf{F}_j(\bar{\mathbf{U}}_h^{\text{ext}(K)}) \right) \int_{\partial K} \phi_n n_j d(\partial K) - \\ &\quad \frac{1}{2} \left(\sum_{\alpha} \int_{\Gamma_{\alpha}} |\partial \hat{\mathbf{F}}| d\Gamma \right) \int_{\partial K} \phi_n d(\partial K) - \\ &\quad \mathbf{G}(\bar{\mathbf{U}}_h^{\text{int}(K)}, \bar{\mathbf{U}}_h^{\text{ext}(K)}) \int_{\partial K} \phi_n \mathbf{n} \cdot \mathbf{s} d(\partial K). \end{aligned}$$

The integral along Γ_α uses a path in phase space between $\bar{\mathbf{U}}_h^{\text{int}(K)}$ and $\bar{\mathbf{U}}_h^{\text{ext}(K)}$. For details, see (Osher and Chakravarthy, 1983). The flow states $\bar{\mathbf{U}}_h^{\text{int}(K)}$ and $\bar{\mathbf{U}}_h^{\text{ext}(K)}$ in the element face are defined as:

$$\begin{aligned}\bar{\mathbf{U}}_h^{\text{int}(K)} &= \frac{1}{|K|} \left(\bar{\mathbf{U}}_h(K) + \sum_{m=1}^3 \hat{\mathbf{U}}_{m,K} \int_{\partial K} \psi_{m,K} d(\partial K) \right) \\ \bar{\mathbf{U}}_h^{\text{ext}(K)} &= \frac{1}{|K|} \left(\bar{\mathbf{U}}_h(K') + \sum_{m=1}^3 \hat{\mathbf{U}}_{m,K'} \int_{\partial K} \psi_{m,K'} d(\partial K) \right),\end{aligned}$$

with K' the index of the element connected to element K at the ∂K . The suffices K and K' of $\psi_m(\mathbf{x}^{(1)})$ refer to the limit of $\psi_m(\mathbf{x}^{(1)})$ taken from the interior and exterior of element K at the boundary ∂K , respectively.

Analytic expressions for the element face moments $\int_{\partial K} \phi_n n_j dK$ are given in (van der Vegt and van der Ven, 1998a). The flux contribution at element faces which is related to the body motion is evaluated using the representation for the velocity of a point in the moving reference frame:

$$\mathbf{s}(t) = \frac{d\mathbf{x}}{dt} = \mathbf{v}(t) + \boldsymbol{\omega}(t) \times \mathbf{r}_b = \mathbf{v}^{(1)}(t) + \boldsymbol{\omega}^{(1)}(t) \times \mathbf{r}_b^{(1)},$$

with \mathbf{v} the velocity vector of the origin of the moving reference frame, $\boldsymbol{\omega}$ the angular velocity vector of the moving reference frame, and $\mathbf{r}_b = \mathbf{x} - \mathbf{x}_b$ a vector pointing from the center of rotation \mathbf{x}_b in coordinate system $O_1X_1Y_1Z_1$ to a point \mathbf{x} in this reference system. Introducing the representation for the grid velocity \mathbf{s} into the approximation of the element face flux integrals we obtain:

$$\int_{\partial K} \phi_n \mathbf{n} \cdot \mathbf{s} d(\partial K) = \left(\mathbf{v}_0^{(1)} \cdot \int_{\partial K} \phi_n \mathbf{n}^{(1)} d(\partial K) + \boldsymbol{\omega}^{(1)} \cdot \int_{\partial K} (\mathbf{r}_b^{(1)} \times \mathbf{n}^{(1)}) \phi_n d(\partial K) \right).$$

The volume flux integrals in (6) are calculated analogously:

$$\int_K \frac{\partial \phi_n}{\partial \mathbf{x}^{(1)}} \cdot \mathbf{s}^{(1)} \mathbf{U}_h dK \cong \bar{\mathbf{U}}_h \left(\mathbf{v}_0^{(1)} \cdot \int_K \frac{\partial \phi_n}{\partial \mathbf{x}^{(1)}} dK + \boldsymbol{\omega}^{(1)} \cdot \int_K \mathbf{r}_b^{(1)} \times \frac{\partial \phi_n}{\partial \mathbf{x}^{(1)}} dK \right).$$

The advantage of this splitting is that we can calculate the contribution of the grid velocity to the fluxes exactly and therefore automatically satisfy the geometric conservation law. The geometric integrals can be calculated analytically or can be found in (van der Vegt, 1998b).

5. Slope Limiter

The discontinuous Galerkin finite element method requires a slope limiting algorithm to obtain monotone solutions. A local projection limiter, which

guarantees monotonicity for multi-dimensional scalar equations, was derived by (Cockburn, Hou and Shu, 1990). Several alternative multi-dimensional limiters have also been proposed which attempt to minimize the loss in accuracy caused by the limiting process, but they are difficult to apply to hexahedral elements. In this paper the Barth and Jespersen slope limiter with the modifications proposed by (Venkatakrishnan, 1995), is used, because it results in a robust formulation and is easy to apply to hexahedral elements. The limiting operator $\Pi_h(\mathbf{U}_h) \in [0, 1]$ is applied directly to the flow field fluctuations $\tilde{\mathbf{U}}_h$, because the element mean flow field $\bar{\mathbf{U}}_h$ remains unchanged by the limiter: $\Pi_h(\bar{\mathbf{U}}_h) = \bar{\mathbf{U}}_h$.

6. Implicit Time Integration

Calculations of unsteady flows frequently suffer from a large disparity between the physically relevant time scales and the time step limitations imposed by the stability constraints of explicit time integration methods. These limitations can be alleviated for the discontinuous Galerkin method presented in this paper by using a three-point backward implicit time integration method to integrate the semi-discrete equations (7) in time. The resulting set of non-linear equations is solved by augmenting these equations with a pseudo-time derivative of the flow field expansion coefficients $\partial\tilde{\mathbf{U}}_m/\partial\tau$ and marching the solution to a steady state in pseudo-time:

$$\frac{\partial\hat{\mathbf{U}}_m}{\partial\tau} = \mathcal{L}_m(\hat{\mathbf{U}}_m, \hat{\mathbf{U}}_m^n, \hat{\mathbf{U}}_m^{n-1}), \quad m = 0, \dots, 3, \quad (8)$$

with:

$$\mathcal{L}_m(\hat{\mathbf{U}}_m, \hat{\mathbf{U}}_m^n, \hat{\mathbf{U}}_m^{n-1}) = \mathcal{R}_m(\mathbf{U}_h) - \mathcal{M}\left(\frac{3}{2}\hat{\mathbf{U}}_m - 2\hat{\mathbf{U}}_m^n + \frac{1}{2}\hat{\mathbf{U}}_m^{n-1}\right)/\Delta t,$$

and Δt the global time step. At steady state the new solution then is equal to $\hat{\mathbf{U}}_m^{n+1}$. This technique was first proposed by (Jameson, 1991), and made unconditionally stable by (Melson, Sanetrik and Atkins, 1993) for the Jameson scheme.

The equations in pseudo-time τ are integrated using the third order accurate TVD Runge-Kutta time integration method from (Shu and Osher, 1988):

1. Set $\hat{\mathbf{U}}_m^{(0)} = \hat{\mathbf{U}}_m^n$
2. For $i = 1, 2, 3$ compute the intermediate Runge-Kutta stages:

$$\hat{\mathbf{U}}_m^{(i)} = \Pi_h\left(\sum_{l=0}^{i-1} \left(\alpha_{il}\mathbf{U}_m^{(l)} + \beta_{il}\Delta\tau\mathcal{L}_m(\hat{\mathbf{U}}_m^{(l)}, \hat{\mathbf{U}}_m^n, \hat{\mathbf{U}}_m^{n-1})\right)\right)$$

3. $\hat{\mathbf{U}}_m^{n+1} = \hat{\mathbf{U}}_m^3$

The coefficients in the TVD Runge-Kutta scheme are equal to: $\alpha_{10} = 1, \alpha_{20} = \frac{3}{4}, \alpha_{21} = \frac{1}{4}, \alpha_{30} = \frac{1}{3}, \alpha_{32} = \frac{2}{3}, \beta_{10} = 1, \beta_{21} = \frac{1}{4}, \beta_{32} = \frac{2}{3}$ and zero otherwise. This Runge-Kutta scheme is stable for CFL numbers less than one, but all calculations discussed in this paper have been done using a CFL number of 0.7. The convergence to steady state is accelerated using the FAS multigrid scheme proposed by (Brandt, 1977). The restriction operator uses volume weighted averages and the prolongation operator pure injection. The FAS algorithm is only applied to the element mean flow field. The application of the FAS scheme to the equations for the flow field fluctuations has not been successful, because the corrections in the flow field fluctuation expansion coefficients caused by the limiter disturb the multigrid process. Also, the restriction and prolongation operators are considerably more complicated for the flow field fluctuations for hexahedral elements.

7. Applications

The discontinuous Galerkin finite element discretization is tested by simulating the unsteady flow field about a generic wing, called simple strake wing (SIS). The wing consists of two parts, an outer part and a strake connected to it and represents a generic model for a fighter aircraft. The mesh used for the calculations consists of 189184 grid points with hexahedral elements. The wing is oscillating in pitch with a mean angle of attack $\alpha = 6.157^\circ$ and amplitude $\Delta\alpha = 2.141^\circ$. The oscillating frequency is $\omega = 0.241$ and the free stream Mach number $M_\infty = 0.899$. The simulations were started by first calculating the steady flow field at an angle of attack $\alpha = 8.298^\circ$, see Fig. 1. The flow field has a shock close to the trailing edge and a lambda shock from the junction of the strake with the leading edge of the wing to the wing tip. Both the strake and outer wing generate a vortex system, which merge in the far wake. This vortex system produces significant additional lift. The hysteresis curves of both the lift coefficient C_L and the drag coefficient C_D are plotted in Fig. 2. In these figures also the results of experiments and the effect of different time steps (20 and 40 time steps per period), mesh size (one time coarsened=C and fine=M) in the simulation are plotted.

Figure 2 shows that the hysteresis effects are small for this specific condition. The comparison of the induced drag force with the experiments is good, whereas the lift force in the calculations is slightly overpredicted, as can be expected from inviscid flow simulations. Fig. 3 shows the zeroth harmonic and the real and imaginary part of the first harmonic of the pressure coefficient at the spanwise location $y/b = 0.5$, with b the wing

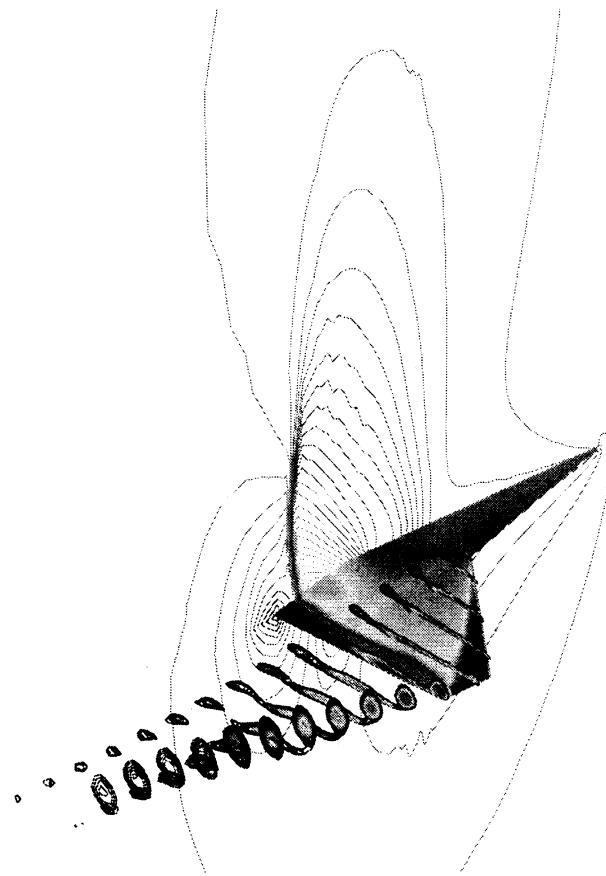


Figure 1. Steady pressure field on Simple Strake wing and total pressure loss in the wake ($\alpha = 8.298^\circ$, $M_\infty = 0.899$).

span, for both the experiments and the numerical results for different time steps Δt and mesh size (one time coarsened=C and fine=M). Apart from the position of the shock the agreement is generally good. These simulations show that discontinuous Galerkin finite element discretization can be used for practical aerodynamic calculations, which until now have been done only in a few cases, see (van der Vegt, 1995a; van der Vegt and van der Ven, 1995b; van der Vegt and van der Ven, 1998a; Baumann, 1997; Baumann and Oden, 1999).

8. Conclusions

Discontinuous Galerkin methods combine many of the nice features of finite volume and finite element methods and can be developed into algorithms

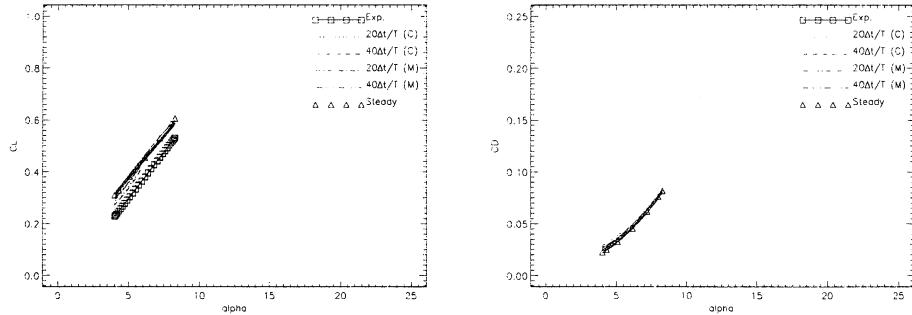


Figure 2. Hysteresis curves for lift force C_L and drag force C_D ($\alpha = 8.298^\circ$, $M_\infty = 0.899$).

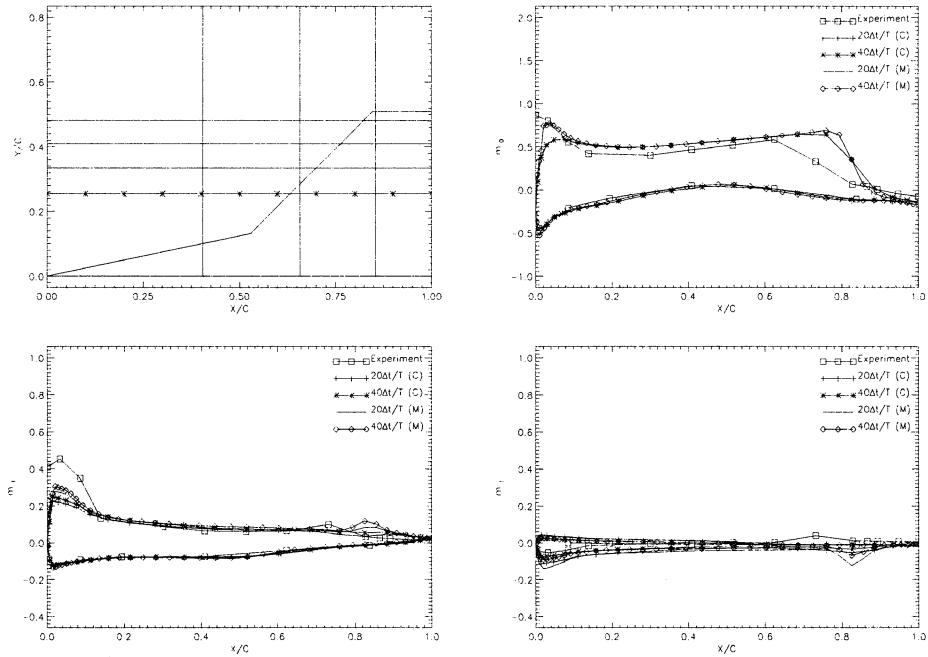


Figure 3. Position of cross-section and zeroth harmonic and real and imaginary part of first harmonic of the pressure coefficient C_p , ($\alpha = 6.157^\circ$, $\Delta\alpha = 2.141^\circ$, $\omega = 0.241$, $M_\infty = 0.899$).

suitable for complicated CFD calculations. In this paper only hyperbolic partial differential equations have been discussed, but presently there is also a significant development of DG methods for elliptic and parabolic partial differential equations. Extensions of the DG method to the solution

of the Navier-Stokes equations can for instance be found in (Cockburn and C.-W. Shu, 1998b; Bassi and Rebay, 1997a; Oden, Babuska, and Baumann, 1998). Discontinuous Galerkin methods have been shown to combine well with local grid refinement and parallel computing, but several important issues remain. The most important ones are the use of a slope limiter, which prevent convergence to steady state, and the significant memory use of DG methods. These issues will have to be addressed in the near future in order to develop DG methods into truly efficient CFD algorithms.

Acknowledgment

The research presented in this paper greatly benefitted from the support and discussions with Dr. B. Oskam from the National Aerospace Laboratory NLR. This work was partially funded by the Netherlands Agency for Aerospace Programmes (NIVR) under contract 7601N.

References

- H.L. Atkins and C.-W. Shu, Quadrature-free implementation of discontinuous Galerkin methods for hyperbolic equations, *AIAA J.* **36**, 775 (1998).
- T.J. Barth, Numerical methods for gasdynamic systems on unstructured meshes, in Proceedings "An introduction to recent developments in theory and numerics for conservation laws", Eds. D. Kröner et al., *Lecture Notes in Computational Science and Engineering*, **5**, 195 (1998).
- F. Bassi and S. Rebay, A high-order accurate discontinuous finite element method for the numerical solution of the compressible Navier-Stokes equations, *J. Comput. Phys.* **131**, 267 (1997).
- F. Bassi and S. Rebay, High-order accurate discontinuous finite element solution of the 2D Euler equations, *J. Comput. Phys.* **138**, 251 (1997).
- C.E. Baumann, An hp-adaptive discontinuous finite element method for computational fluid dynamics, Ph.D. dissertation, The University of Texas at Austin, Aug. 1997.
- C.E. Baumann and J.T. Oden, A discontinuous *hp* finite element method for the Euler and Navier-Stokes equations, to appear in *Int. J. Numer. Meth. Fluids* (1999).
- K.S. Bey and J.T. Oden, A Runge-Kutta discontinuous finite element method for high speed flows, AIAA Paper 91-1575-CP (1991).
- K.S. Bey and J.T. Oden, *hp*-Version discontinuous Galerkin methods for hyperbolic conservation laws, *Comput. Methods Appl. Mech. Engrg.* **133**, 259 (1996).
- R. Biswas, K. D. Devine, and J. Flaherty, Parallel, adaptive finite element methods for conservation laws, *Appl. Numer. Math.*, **14**, 255 (1994).
- A. Brandt, Multi-level adaptive solutions to boundary value problems, *Math. Comp.*, **31**, 333 (1977).
- G. Chavent and G. Salzano, A finite element method for the 1D water flooding problem with gravity, *J. of Comput. Physics* **45**, 307 (1982).
- G. Chavent and B. Cockburn, The local projection P^0 - P^1 -discontinuous-Galerkin finite element method for scalar conservation laws, *MAN Math. Modelling and Numer. Anal.* **23**, 565 (1989).
- P.G. Ciarlet and J.L. Lions, *Handbook of Numerical Analysis*, Volume II, part 1, Finite element methods, North Holland, Amsterdam (1991).
- B. Cockburn and C.-W. Shu, The Runge-Kutta local projection P^1 discontinuous Galerkin finite element method for scalar conservation laws, Proceeding of the First

- National Fluid Dynamics Congress, Cincinnati, Ohio (1988).
- B. Cockburn and C.-W. Shu, TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws II: General framework, *Math. Comput.* **52**, 411 (1989).
- B. Cockburn, S.-Y. Lin and C.-W. Shu, TVD-Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws III: One-dimensional systems, *J. Comput. Phys.* **84**, 90 (1989).
- B. Cockburn, S. Hou and C.-W. Shu, The Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws IV: The multidimensional case, *Math. Comput.* **54**, 545 (1990).
- B. Cockburn and C.-W. Shu, The Runge-Kutta local projection P^1 -discontinuous Galerkin finite element method for scalar conservation laws, *Math. Modelling and Num. Anal.* **25**, 337 (1991).
- B. Cockburn and C.-W. Shu, The Runge-Kutta discontinuous Galerkin method for conservation laws V, *J. Comput. Phys.* **141**, 199 (1998).
- B. Cockburn and C.-W. Shu, The local discontinuous Galerkin method for time-dependent convection-diffusion systems, *SIAM J. Numer. Anal.* **35**, 2440 (1998).
- B. Cockburn, Discontinuous Galerkin methods for convection dominated problems, in *High order methods for computational physics*, edited by H. Deconinck and T.J. Barth, Lecture Notes in Computational Science and Engineering, **9** (Springer Verlag, 1999), p. 69.
- B. Cockburn, G. Karniadakis and C.-W. Shu, An overview of the development of discontinuous Galerkin methods, in *Discontinuous Galerkin Methods, Theory, Computation and Applications*, edited by B. Cockburn, G. Karniadakis and C.-W. Shu, Lecture Notes in Computational Science and Engineering, **11**, 3 (Springer Verlag, 2000).
- P. Hansbo and C. Johnson, *Streamline Diffusion Finite Element Methods for Fluid Flow*, in Von Karman Institute for Fluid Dynamics, Lecture Series 1995-02 (1995).
- F.Q. Hu, M.Y. Hussaini, and P. Rasetarinera, An analysis of the discontinuous Galerkin method for wave propagation problems, *J. of Comput. Physics*, **151**, 921 (1999).
- A. Jameson, *Time Dependent Calculations using Multigrid, with Applications to Unsteady Flows past Airfoils and Wings*, AIAA Paper 91-1596 (1991).
- C. Johnson and J. Pitkaranta, An analysis of the discontinuous Galerkin method for a scalar hyperbolic equation, *Math. Comp.*, **46**, 1 (1986).
- C. Johnson, A. Szepessy, and P. Hansbo, On the convergence of shock-capturing streamline diffusion finite element methods for hyperbolic conservation laws, *Math. Comp.* **54**, 107 (1990).
- D.S. Kershaw, M.K. Prasad, M.J. Shaw and J.L. Milovich, 3D Unstructured mesh ALE hydrodynamics with the upwind discontinuous finite element method, *Comput. Methods Appl. Mech. Engrg.* **158**, 81 (1998).
- P. Lesaint and P.A. Raviart, On a finite element method for solving the neutron transport problem, in *Mathematical Aspects of Finite Elements in Partial Differential Equations*, edited by C. de Boor (Academic Press, 1974), p. 89.
- M. Lesoinne and C. Farhat, Geometric conservation laws for flow problems with moving boundaries and deformable meshes, and their impact on aeroelastic computations, *Comput. Methods Appl. Mech. Engrg.* **134**, 71 (1996).
- D.P. Lockard and H.L. Atkins, Efficient implementation of the quadrature-free discontinuous Galerkin method, AIAA paper 99-3309, in *Proc. 14th AIAA CFD Conference, Norfolk Virginia* (1999).
- S.-Y. Lin and Y.-S. Chin, Discontinuous Galerkin finite element method for Euler and Navier-Stokes equations, *AIAA J.*, **31**, 2016 (1993).
- R.B. Lowrie, P.L. Roe, and B. van Leer, A space-time discontinuous Galerkin method for the time-accurate numerical solution of hyperbolic conservation laws, AIAA paper 95-1658-CP, in *Proc. 12th AIAA CFD Conference, San Diego, California* (1995).
- R.B. Lowrie and J.E. Morel, Discontinuous Galerkin for stiff hyperbolic systems, AIAA paper 99-3307, in *Proc. 14th AIAA CFD Conference, Norfolk Virginia* (1999).

- A. Masud and T.J.R. Hughes, A space-time Galerkin/least-squares finite element formulation of the Navier-Stokes equations for moving domain problems, *Comput. Methods Appl. Mech. Engrg.* **146**, 91 (1997).
- N.D. Melson, M.D. Sanetrik and H.L. Atkins, Time-accurate Navier-Stokes calculations with multigrid acceleration, in *Proc. 6th Copper Mountain Confer. on Multigrid Methods* (1993).
- J.T. Oden, I. Babuska, and C.E. Baumann, A discontinuous hp finite element method for diffusion problems, *J. Comput. Phys.* **146**, 491 (1998).
- S. Osher and S. Chakravarthy, Upwind schemes and boundary conditions with applications to Euler equations in general geometries, *J. Comput. Phys.* **50**, 447 (1983).
- W.H. Reed and T.R. Hill, Triangular mesh methods for the neutron transport equation, technical report LA-UR-73-479, Los Alamos Scientific Laboratory, New Mexico (1973).
- F. Shakib, T.J.R. Hughes and Z. Johan, A new finite element method for computational fluid dynamics: X. The compressible Euler and Navier-Stokes equations, *Comput. Methods Appl. Mech. Engrg.* **89**, 141 (1991).
- C.-W. Shu and S. Osher, Efficient implementation of essentially non-oscillatory shock-capturing schemes, *J. Comput. Phys.* **77**, 439 (1988).
- P.D. Thomas and C.K. Lombard, Geometric conservation law and its application to flow computations on moving grids, *AIAA J.* **17**, 1030 (1979).
- E.F. Toro, *Riemann solvers and numerical methods for fluid dynamics*, Springer Verlag, Berlin, Germany (1997).
- J.J.W. van der Vegt, Anisotropic grid refinement using an unstructured discontinuous Galerkin method for the three-dimensional Euler equations of gas dynamics, in *Proc. 12th AIAA CFD Conference, San Diego, California, 1995*. [AIAA Paper 95-1657-CP]
- J.J.W. van der Vegt and H. van der Ven, Hexahedron Based Grid Adaptation for Future Large Eddy Simulation, in *Proc. Progress and Challenges in CFD Methods and Algorithms, Seville, Spain, 1995*. [AGARD CP-578, p. 22-1]
- J.J.W. van der Vegt and H. van der Ven, Discontinuous Galerkin finite element method with anisotropic local grid refinement for inviscid compressible flows, *J. Comput. Phys.* **141**, 46 (1998).
- J.J.W. van der Vegt, Technical Publication TP-98239, National Aerospace Laboratory NLR, Amsterdam, The Netherlands, <http://www.nlr.nl/public/library/1999-1/iw-index.html> (1998).
- J.J.W. van der Vegt and H. van der Ven, Space-time discontinuous Galerkin finite element method with dynamic grid motion for inviscid compressible flows, submitted to *J. Comput. Phys.* (2000).
- H. van der Ven and J.J.W. van der Vegt, Experiences with Advanced CFD Algorithms on NEC SX-4, in *Proc. Vector and Parallel Processing VECPAR '96*, edited by Palma and Dongarra, Lect. Notes in Computer Science, (Springer Verlag, 1997).
- H. van der Ven and J.J.W. van der Vegt, Partitioning and parallel development of an unstructured, adaptive flow solver on the NEC SX-4, in *Proc. Parallel Computational Fluid Dynamics '97 Conference*, edited by D.R. Emerson et al., (North Holland, 1998).
- H. van der Ven and J.J.W. van der Vegt, Accuracy, resolution, and computational complexity of a discontinuous Galerkin finite element method, in *Discontinuous Galerkin Methods, Theory, Computation and Applications*, edited by B. Cockburn, G. Karniadakis and C.-W. Shu, Lecture Notes in Computational Science and Engineering, **11**, 439 (Springer Verlag, 2000).
- B. van Leer, Towards the ultimate conservative scheme, II Monotonicity and conservation combined in a second order scheme, *J. Comput. Phys.* **14**, 361 (1974).
- V. Venkatakrishnan, Convergence to steady state solutions of the Euler equations on unstructured grids with limiters, *J. Comput. Phys.* **118**, 120 (1995).

SOLVING INCOMPRESSIBLE TWO-PHASE FLOWS WITH A COUPLED TVD INTERFACE TRACKING / LOCAL MESH REFINEMENT METHOD

S. VINCENT, J-P. CALTAGIRONE

Laboratoire MASTER-ENSCPB,

Bordeaux 1 University,

Avenue Pey-Berland, BP 108, 33402 Talence, France

Emails: vincent@lmaster.u-bordeaux.fr

calta@lmaster.u-bordeaux.fr

Abstract. A local multigrid refinement method, which is adaptative in time and space and which refines the grid at the cell scale, has been developed to solve the different scales of interfaces involved in two-phase flows. On each grid level, the Navier-Stokes equations are approximated by Finite Volumes on a MAC grid. Moreover, an advection equation on the phase function is solved by a Lax-Wendroff TVD scheme. Composite boundary conditions are proposed to solve the flow on multigrid calculation domains. Scalar problems and real flows such as Rayleigh-Taylor instabilities or bubble oscillations are presented.

1. Introduction

The classical method for tracking interfaces, such as the volume of fluid method (Hirt and Nichols, 1981), the marker method (Daly, 1967) or the level set method (Sethian, 1996), all consist of two steps: the interface position is first estimated by particles moving with the free boundary or by the solution of a conservation equation; the distribution of the two phases is then reconstructed to ensure mass conservation or a smooth profile. Since the grid spacing is smaller than the finer scales of the structures developing at the interface, these methods are efficient and provide accurate results. However, when the precision of the grid becomes insufficient, numerical artifacts are generated by these classical methods (poor mass conservation, artificial surface tension,...). As soon as we get interested in

non-symmetric three-dimensional simulations, the memory limit of super-computers restricts the computations to two-phase flows with macro-scale interfacial structures.

To optimize the computational time and the memory requirements while increasing the precision of the numerical solution we propose to couple a local mesh refinement technique with an implicit solver for the equations of motion and an original explicit interface capturing method based on TVD schemes. Several examples are presented to illustrate the abilities of the multi-scale solver.

2. Governing equations and motion equation solver on a single grid

The numerical simulation of unsteady and incompressible interfacial flows involving strong interface deformations is classically achieved by the implementation of a fixed Cartesian grid method, where the Navier-Stokes equations written as a 1-fluid model are coupled with an advection equation on a phase function:

$$\begin{aligned} \nabla \mathbf{u} &= 0 \\ \rho \left[\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} \right] &= -\nabla p + \rho \mathbf{g} + \nabla \cdot [\mu (\nabla \mathbf{u} + \nabla^T \mathbf{u})] + \sigma \kappa \mathbf{n}_i \delta_i \quad (1) \\ \frac{\partial C}{\partial t} + \mathbf{u} \cdot \nabla C &= 0 \end{aligned}$$

where \mathbf{u} is the velocity, p is the pressure, C is the phase function, \mathbf{g} is the gravity, ρ is the density, μ is the dynamic viscosity, σ is the surface tension, κ is the mean curvature of the interface, \mathbf{n}_i is a normal to the interface and δ_i is a Dirac function indicating the interface. The system (1) is called the 1-fluid model because it models a two-phase flow by 1 fluid with varying densities and viscosities. The phase function C repairs the whole phase of the flow by assuming $C=1$ in one fluid and $C=0$ in the other fluid. The interface is naturally defined as $C=0.5$. In system (1), the discontinuous character of the physical characteristics is described by the phase function as follows:

$$\begin{aligned} &\text{if } C \geq 0.5, \rho = \rho_1 \text{ and } \mu = \mu_1 \\ &\text{if } C < 0.5, \rho = \rho_0 \text{ and } \mu = \mu_0 \quad (2) \end{aligned}$$

where ρ_0 , ρ_1 , μ_0 and μ_1 are the characteristics of fluids 0 and 1 respectively. The motion equation system is approximated by implicit Finite Volumes on a staggered grid. The coupling between pressure and velocity is dealt with thanks to an augmented Lagrangian procedure (Vincent and Caltagirone, 1999). A penalization term is added to the momentum equation to enforce

the divergence free condition. The implicit and discrete 1-fluid system reads

$$\begin{aligned}
 & \frac{\frac{3}{2}\mathbf{u}_0^{n+1} - 2\mathbf{u}_0^n + \frac{1}{2}\mathbf{u}_0^{n-1}}{\Delta t} + (\mathbf{u}_0^n \cdot \nabla) \mathbf{u}_0^{n+1} - r \nabla (\nabla \cdot \mathbf{u}_0^{n+1}) = -\frac{1}{\rho} \nabla p_0^n \\
 & + \frac{1}{\rho} \nabla \cdot [\mu (\nabla \mathbf{u}_0^{n+1} + \nabla^T \mathbf{u}_0^{n+1})] + \mathbf{g} - \mathbf{B}_u^v (\mathbf{u}_0^{n+1} - \mathbf{u}_\infty) + \sigma \kappa \mathbf{n}_i \delta_i \\
 & p_0^{n+1} - p_0^n = -r \nabla \cdot \mathbf{u}_0^{n+1} \\
 & \frac{\partial \mathbf{u}_0^{n+1}}{\partial \mathbf{n}} = \mathbf{B}_u^s (\mathbf{u}_0^{n+1} - \mathbf{u}_\infty) \text{ on the boundaries.} \\
 & \frac{C_0^{n+1} - C_0^n}{\Delta t} + \mathbf{u}_0^n \cdot \nabla C_0^n = 0
 \end{aligned} \tag{3}$$

where Δt is the time scale, \mathbf{n} is a normal to the boundaries, r is a numerical parameter controlling incompressibility and \mathbf{u}_∞ is a reference velocity. \mathbf{B}_u^s is a diagonal matrix, whose components are Fourier-like control parameters imposing the Boundary conditions (Dirichlet if $\mathbf{B}_u^{s,l} = +\infty$ and Neumann if $\mathbf{B}_u^{s,l} = 0$). The volume control parameters $\mathbf{B}_u^{v,l}$ are used to enforce a reference velocity \mathbf{u}_∞ in the calculation domain. The subscript 0 refers to the variables on the coarse grid G_0 and exponent n to time ($n\Delta t$). In the motion equation system, the temporal derivatives are approximated by a Gear scheme of second order whereas the spatial derivatives are discretised by a quick scheme (non-linear terms) and a centred scheme (diffusive term). The superficial tension term is modelled by a continuum surface force CSF method (Brackbill, Kothe and Zemach, 1992). The linear system generated by the previous discretisations is solved by an iterative Bi-Conjugate Gradient Stabilised algorithm BiCgStab (Van Der Vorst, 1992) preconditioned with a Modified and Incomplete LU method MILU. The numerical treatment of the advection equation on the phase function C is detailed in the following section.

3. Multigrid interface capturing method

An interface capturing method (Vincent and Caltagirone, 1999) has been developed to describe the interface deformations. The principle is to solve the hyperbolic advection equation on phase function C thanks to an explicit Lax-Wendroff TVD scheme, which is able to take into account the discontinuities of the phase function while keeping a monotone solution and suitable precision in time and space. This method is very interesting because it is easy to implement even in three-dimensions, and it can deal with more than two fluids.

To limit the computation nodes far away from the free surface and concentrate the calculation points on the interface, an original one-cell local multigrid method (OCLM) is proposed. Starting on a coarse grid G_0 , we

define a refinement criterion $R_c = \|\nabla C\|$ to detect the points to be refined on each multigrid level G_l . If $R_c(M) \neq 0$ with $M \in G_{l-1}$, the control volume around pressure point M is refined and a fine calculation domain $G_{l,s}$ is built. For all grid level l , the coarse solution $(\mathbf{u}_{l-1}^{n+1}, p_{l-1}^{n+1}, C_{l-1}^{n+1})$ are prolongated on G_l using a classical Q1 interpolation procedure and for all calculation domain s of level l , (3) is solved to get $(\mathbf{u}_l^{n+1}, p_l^{n+1}, C_l^{n+1})$. Each fine grid solution of level l is restricted to G_{l-1} thanks to a direct injection technique for the pressure and the velocity and a Full Weighting Interface Control Volume FWICV (Hackbusch, 1985) procedure for the phase function. The time step is then incremented and another multigrid solving procedure is initiated. Finally, thanks to the OCLM method, we have access to a multi-scale solution, which adapts in time and space to the interface. A hierarchy of embedded sub-domains is obtained, each fine calculation grid $G_{l,s}$ corresponding to a coarse control volume around a detected point of G_{l-1} cut by 3 in each space direction. An odd cutting is necessary to ensure a perfect connection between the fine grids $G_{l,s}$ (Fig. 1).

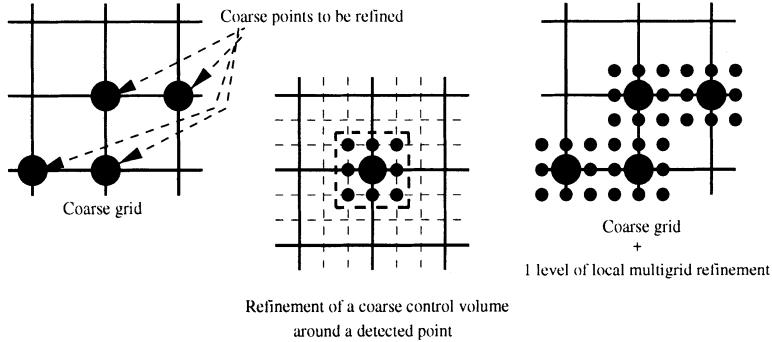


Figure 1. Local mesh refinement technique.

We have extended the 3×3 fine calculation domains to 5×5 grids to deal with the boundary conditions. Defining realistic boundary conditions in such a way that an accurate solution is calculated on the fine grids is the main difficulty in the OCLM method. The boundaries of the fine grids $G_{l,s}$ do not correspond inevitably with those of the physical problem on G_0 . We have imposed the interpolated values of the coarse grid G_{l-1} at the boundaries of $G_{l,s}$ by means of Dirichlet boundary conditions ($\mathbf{B}_u^{s,l} = +\infty$). Nevertheless, due to the reduced size of the fine grids, the physical system is too stressed and the prolongation operator does not preserve the divergence free property. In this way, the Dirichlet boundary conditions force the numerical solver to converge to a non suitable solution. To relax the constraints at the boundaries of the multigrid calculation grids, composite boundary conditions were introduced into the Navier-Stokes equation sys-

tem such as $\mathbf{B}_u^{s,l} \approx \frac{3}{2h_l}$ and $\mathbf{B}_u^{v,l} \approx \frac{2\mathbf{B}_u^{s,l}}{h_{l-1}}$, where h_l is the space scale on G_l . This numerical idea results from theoretical work which was developed to verify the conservation of the symmetric components of the shear stress between two multigrid levels (Vincent and Caltagirone, 2000). Orders of magnitude of the composite boundary condition parameters $\mathbf{B}_u^{s,l}$ and $\mathbf{B}_u^{v,l}$ are included between 10^4 and 10^9 .

The method has been validated on the vortex test of (Rider and Kothe, 1995). A concentration circle is strongly distorted in an analytical shearing velocity field. A pseudo-analytical solution is presented in (Rider and Kothe, 1995). The local mesh refinement and the multigrid solution with 3 grid levels are presented in Fig. 2. Starting on a 70×70 coarse grid, the absolute error between the analytical solution and the multigrid one is less than 0.1%. A second order convergence rate is observed on mass conservation, on the position of the interface and on the velocity field.

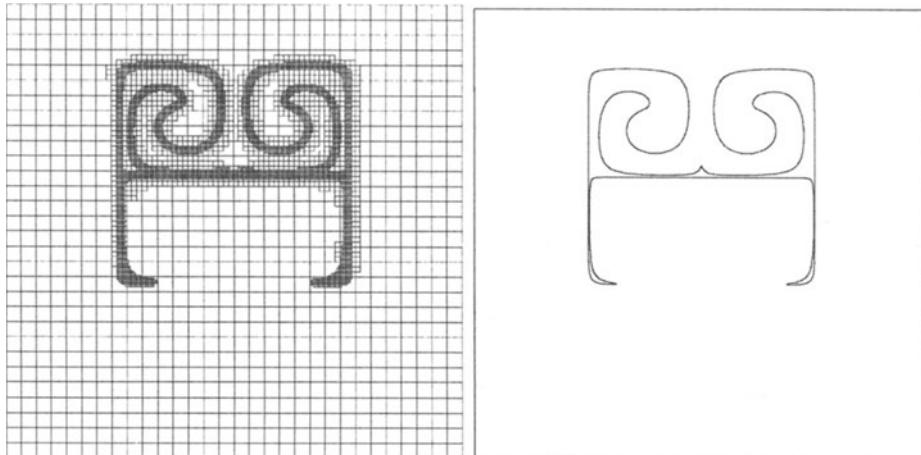


Figure 2. Multigrid simulation of the vortex problem. The local mesh refinement (left) and the multigrid solution of the interface position are presented (right).

4. Numerical simulation of 2D interfacial flows

To highlight the abilities of the OCLM method to solve two-phase flows dealing with stongly varying interfaces, we present two widely studied problems: the non-linear oscillations of a bubble initially perturbed by a Legendre polynomial of mode 4 (Fig. 3) and Rayleigh-Taylor instabilities (Fig. 4).

In the first test, a two-dimensional drop is initially pertubed in a zero gravity field and the free surface oscillates around its circular equilibrium

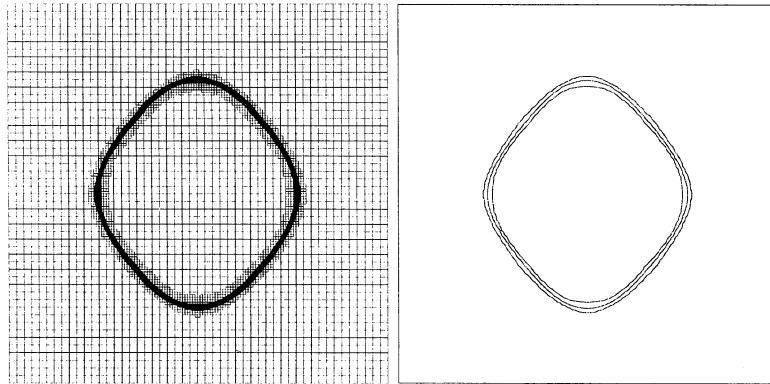


Figure 3. Multigrid simulation of non-linear bubble oscillation. The local mesh refinement (left) and the phase function (isovalue 0.01, 0.5 and 0.99) are presented (right).

shape. The physical characteristics are $\frac{\rho_1}{\rho_0} = \frac{\mu_1}{\mu_0} = 100$ and $\sigma = 0.5 N.m^{-1}$. Three grids levels were computed with a 50×50 coarse mesh and $h_0 = 0.8 mm$.

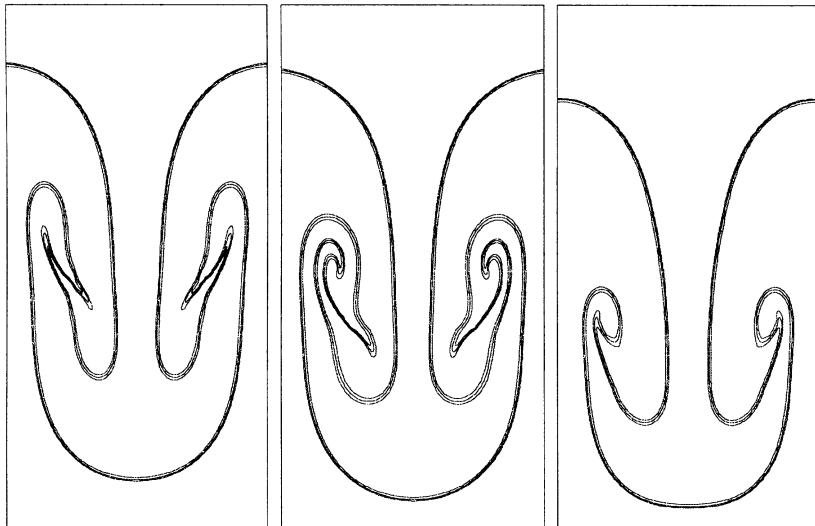


Figure 4. Multigrid simulation of Rayleigh-Taylor instabilities: $Re = 200$, $Re = 500$ and $Re = 700$ (from left to right). Isovalues 0.01, 0.5 and 0.99 of C are plotted.

We also illustrate strong interface deformations by Rayleigh-Taylor instabilities corresponding to Reynolds numbers $Re = \frac{\rho L u}{\mu}$ equal to 200, 500 and 700, Atwood numbers $A = \frac{\rho_1 - \rho_0}{\rho_1 + \rho_0}$ equal to 0.11, 0.33 and 0.5, and

zero surface tension. Starting on a coarse 30×60 grid, several local mesh refinement levels were built to track the dynamics of the interface. On G_0 , the space scale h_0 is chosen equal to 0.8mm .

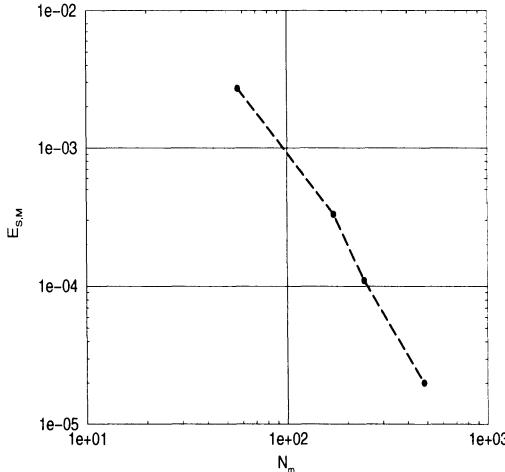


Figure 5. Evolution of the difference $E_{S,M}$ between a single grid and a multigrid solution according to the square root N_N of the whole of the calculation points.

The multigrid solutions are second order accurate with respect to mass conservation. A comparison between the multigrid solution with two levels of local mesh refinement and the solution on an equivalent single grid demonstrates an almost second order convergence rate. The difference $E_{S,M}$ between these solutions is presented in Fig. 5.

Let R_{GI} be a geometrical improvement ratio between the number of calculation points on the multigrid levels G_0 to $G_{l_{max}}$ ($l_{max} = 1, 2$) and the number of computation nodes required to obtain an equivalent single grid solution. For $R_{GI} \rightarrow 0$, the best improvement in terms of calculation points and consequently in computational memory, is obtained, whereas for $R_{GI} \rightarrow 1$, the local mesh refinement is ineffective. Fig. 6 describes the behaviour of R_{GI} for Rayleigh-Taylor instability ($A=0.33$ and $\text{Re}=500$). The gains in calculation points are always as high as 85% in spite of the strong stretching of the interface which induces large refined zones of the grid.

5. Conclusions

Thanks to the coupled augmented Lagrangian / TVD interface capturing method and to the OCLM grid refinement method, several objectives have

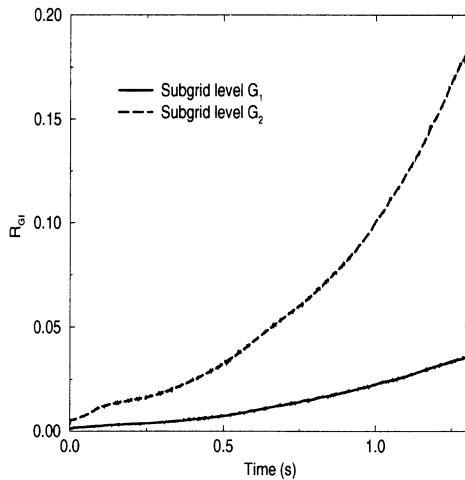


Figure 6. Behaviour of the geometrical improvement ratio R_{GI} for multigrid solutions with $l_{max} = 1$ and 2 of Rayleigh-Taylor instability ($A=0.33$ and $Re=500$).

been reached: strong deformations of the phase field and thin interface structures are accurately captured by the OCLM method, the multigrid solutions converge to a unique solution with a second order rate, the diffusion of the TVD scheme is controlled on 4 or 5 fine cells whatever the number of grid levels, the memory is reduced by at least 80% in all the simulations and the computational time is decreased from 30 to 60%.

References

- Vincent S and Caltagirone JP (1999). Efficient solving method for unsteady incompressible interfacial flow problems. *Int. J. Numer. Meth. Fluids* **30**, pp 795-811.
- Brackbill JU, Kothe BD and Zemach C (1992). A continuum method for modeling surface tension. *J. Comput. Phys.* **100**, pp 335-354.
- Van Der Vorst HA (1992). ABi-CGSTAB: a fast and smoothly converging variant of Bi-CG for the solution of non-symmetric linear systems. *SIAM Journal on Stat. Comput.* **13**, pp 631-644.
- Hirt CW and Nichols BD (1981). Volume of fluid (vof) methods for the dynamics of free boundaries. *J. Comput. Phys.* **39**, pp 201-225.
- Daly BJ (1967). Numerical study of two-fluid Rayleigh-Taylor instability. *Phys. Fluids*, **10**, pp 297-307.
- Sethian JA (1996). Level set methods. Cambridge University Press.
- Hackbusch W (1985). Multi-grid methods and applications. SCM Vol. 4, Springer, Berlin
- Vincent S and Caltagirone JP (2000) One Cell Local Multigrid method for solving unsteady incompressible multi-phase flows. *J. Comput. Phys.* in correction.
- Rider WJ and Kothe DB (1995). Stretching and tearing interface tracking methods. *AIAA paper*, **95**, pp 1717.

A LARGE TIMESTEP GODUNOV-TYPE MODEL OF GLOBAL ATMOSPHERIC CHEMISTRY AND TRANSPORT

B.M.-J.B.D. WALKER AND N. NIKIFORAKIS

*Department of Applied Mathematics
and Theoretical Physics, University of Cambridge,
Silver Street, Cambridge, CB3 9EW, U.K.*

*Emails: bmjbdw2@damtp.cam.ac.uk,
nn10005@damtp.cam.ac.uk*

Abstract. One of the main shortcomings of Eulerian methods in global atmospheric Chemistry and Transport Models (CTMs) is the computational cost due to the stringent stability restriction on the Courant-Friedrichs-Levy (CFL) number to one. Our aim was to implement a scheme which decreases the number of time steps required, while retaining the main advantages of the high-resolution class of methods. This article outlines an implementation of a CFL equal to 2 version of the Weighted Average Flux (WAF) method in a CTM. The scheme, which has increased computational efficiency compared to conventional Riemann-problem-based methods, is briefly presented. It is evaluated both for a planar geometry two-dimensional problem, which has an exact solution, and for a case-study of stratospheric transport on a spherical geometry using meteorological data analyses. Results indicate that the scheme can operate at timesteps bound by a CFL number of 2 without loss of accuracy, and achieves significant savings of cpu-time.

1. Introduction

We aim at implementing efficient numerical solvers in global atmospheric Chemistry and Transport Models (CTMs). The governing equations for the advection and chemical interaction of atmospheric species, which need to be solved by these models, take the form

$$\frac{\partial n_i}{\partial t} + \nabla \cdot (n_i \mathbf{v}) = Q_i - L_i n_i, \quad i = 1, \dots, N_s, \quad (1)$$

where n_i are the individual chemical species number densities (for N_s species). Equations (1) are in effect scalar conservation laws with source terms, the latter accounting for chemical production and loss (Q and L). The velocity field, that is wind values ($\mathbf{v} = \mathbf{v}(x, y, z, t)$), and other thermodynamic quantities, such as pressure or temperature, are either provided by meteorological analyses or by forecasts from General Circulation Models (GCMs). The simulations are then known as *off-line* or *coupled*, respectively.

We propose to achieve a cpu-time saving by reducing the number of timesteps needed to reach a specific time (compared to most contemporary schemes) while solving equation (1). We hence consider a scheme with a CFL number restriction of 2, originally formulated by Toro and Billett in 1997, which we will refer to as WAF2. These authors consider the linear advection equation with constant coefficients, Burgers' equation and the compressible unsteady Euler equations. Part of our work is the extension of the formulation of the WAF2 scheme to variable-coefficient scalar laws like (1); this is summarised in section 2. The second part of this work addresses issues specific to the implementation of the scheme in a global atmospheric chemistry and transport model, which are related to the spherical geometry of the problem. The resulting scheme is then validated using a problem on planar geometry, while the CTM itself is tested with a simulation of the behaviour of the Northern hemisphere polar vortex using meteorological analyses.

2. Outline of the numerical scheme

The original numerical scheme is a large time-stepping extension of the weighted average flux (WAF) method (Toro, 1989), referred to as WAF1, to a maximum CFL restriction of 2 (Toro and Billett, 1997). To describe the underlying idea briefly let us consider the homogeneous, one-dimensional version of equation (1), for a single generic variable u

$$u_t + f(u)_x = 0, \quad (2)$$

where f is the scalar flux function. The WAF family of methods relies on the finite volume formula

$$u_i^{n+1} = u_i^n + \frac{\Delta t}{\Delta x} (f_{i-1/2} - f_{i+1/2}) \quad (3)$$

to update the solution from time level $t = n$ to $t = n+1$ for given space and time intervals Δx and Δt respectively, $f_{i-1/2}$ and $f_{i+1/2}$ being numerical fluxes. The latter are calculated by considering the generalised intercell flux

$$f_{i+1/2} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \frac{1}{x_2 - x_1} \int_{x_1}^{x_2} f(\hat{u}(x, t)) dx dt, \quad (4)$$

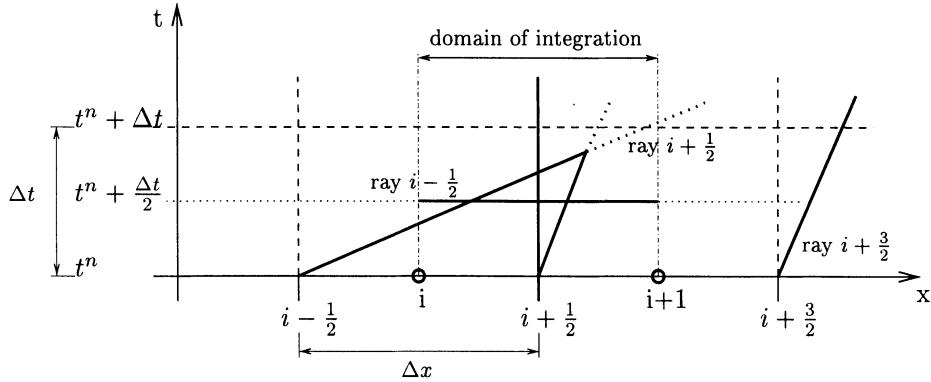


Figure 1. Cells taken into account in the calculation of the numerical flux $i + 1/2$ for the WAF2 scheme.

where \hat{u} is the solution of local Riemann problems defined by a piecewise-constant approximation of the data of any two consecutive cells and the conservation law. The space and time limits in formula (4) can be taken as $x_1 = -\Delta x/2$, $x_2 = \Delta x/2$, $t_1 = 0$ and $t_2 = \Delta t$, where, in local space/time coordinates, $x = 0$ is the intercell boundary and $t = 0$ is the time at timelevel n . The timestep is such that $|c| = (|a|\Delta t)/\Delta x \leq 2$ and the flux is evaluated by a midpoint rule in time, i.e. at $\Delta t/2$ (Figure 1). The implication of propagating information on the grid at a speed faster than that of the advection, is that waves from neighbouring Riemann problems may reach the region where the weighted average takes place, as seen on Figure 1, so they have to be taken into account while averaging. Wave interactions may take place in this case; these interactions are considered to be linear, that is waves pass through each other without any changes either in direction or in intensity. To evaluate the fluxes for the whole range of CFL numbers $|c| \leq 2$, the following WAF-based scheme is used (Toro and Billett, 1997):

$$f_{i+1/2} = \frac{1}{2}(f_i + f_{i+1}) - \frac{1}{2}\psi_{i+1/2,L}\Delta f_{i-1/2} - \frac{1}{2}\phi_{i+1/2}\Delta f_{i+1/2} - \frac{1}{2}\psi_{i+1/2,R}\Delta f_{i+3/2}, \quad (5)$$

where $\Delta f_{i+1/2} = f_{i+1} - f_i$, and

$$\psi_{i+1/2,L} = \begin{cases} |c_{i-1/2}| & \text{if } c_{i-1/2} > 1 \\ 0 & \text{otherwise} \end{cases}, \quad \psi_{i+1/2,R} = \begin{cases} -(|c_{i-1/2}| - 1) & \text{if } c_{i-1/2} < -1 \\ 0 & \text{otherwise} \end{cases}$$

$$\phi_{i+1/2} = \begin{cases} c_{i+1/2} & \text{if } -1 < c_{i-1/2} < 1 \\ \operatorname{sgn}(c_{i+1/2}) & \text{otherwise} \end{cases}.$$

This is a unified Riemann-problem-based extension of the Warming-Beam and Lax-Wendroff schemes; the ϕ and ψ functions operate the switching

between the Lax-Wendroff and Warming-Beam schemes respectively. The scheme is second order accurate in space and time; for a detailed description the reader is referred to the paper by Toro and Billett, 1997. In the same paper, appropriate limiter functions which ensure that the solution remains free of unphysical oscillations in the neighbourhood of steep gradients are constructed.

3. Implementation for planar kinematics

To evaluate the flux function (5) we need to solve the local Riemann problems defined by equation (2), which, in the case of a scalar conservation law with variable coefficients $a = a(x, t)$, is

$$u_t + (a(x, t)u)_x = 0, \quad (6)$$

and by the value of the scalar at any two consecutive cells. Although equation (6) is linear, the dependence of the velocity field on time and space can have a nonlinear effect on the scalar u , similar to Burgers' equation. The solution of the local linear problem in this case is a single wave which, depending on the sign of the coefficients, can be either a propagating discontinuity or an expansion fan. This gives rise to five possible wave patterns, namely left or right shocks and left, right or centred expansion fans. An algorithm for the exact solution of the Riemann problem needs a number of conditional statements to cover all five cases. This naturally incurs additional computational expense. Following Nikiforakis and Toro (2000), we propose to use an approximate Riemann solver which avoids this problem. To this end we solve the local Riemann problems for any two consecutive cells i and $i + 1$ exactly, but for the approximate problem resulting from replacing the original PDE (6) with the linear *constant* coefficient PDE

$$u_t + (a_{i+\frac{1}{2}} u)_x = 0. \quad (7)$$

Here, $a_{i+\frac{1}{2}}$ is the characteristic speed at the interface $i + \frac{1}{2}$ and is assumed locally constant. As for the choice of $a_{i+\frac{1}{2}}$, we took a linear interpolation, in which case we obtain $a_{i+\frac{1}{2}} = \frac{1}{2}[a_i^n + a_{i+1}^n]$. Higher-order interpolation schemes may also be used. The WAF2 scheme can now be applied to linear advection problems with variable coefficients. Extensions of the scheme to two and three dimensions, and forcing terms (eg. chemistry), are implemented by operator splitting (Nikiforakis and Toro, 2000).

To validate the scheme we consider a problem of kinematic frontogenesis which has an exact solution (see Davies-Jones, 1985, and references therein for its formulation). The velocity field is an idealised Rankine vortex which progressively distorts a scalar field. Results for time $t = 4$ are shown in

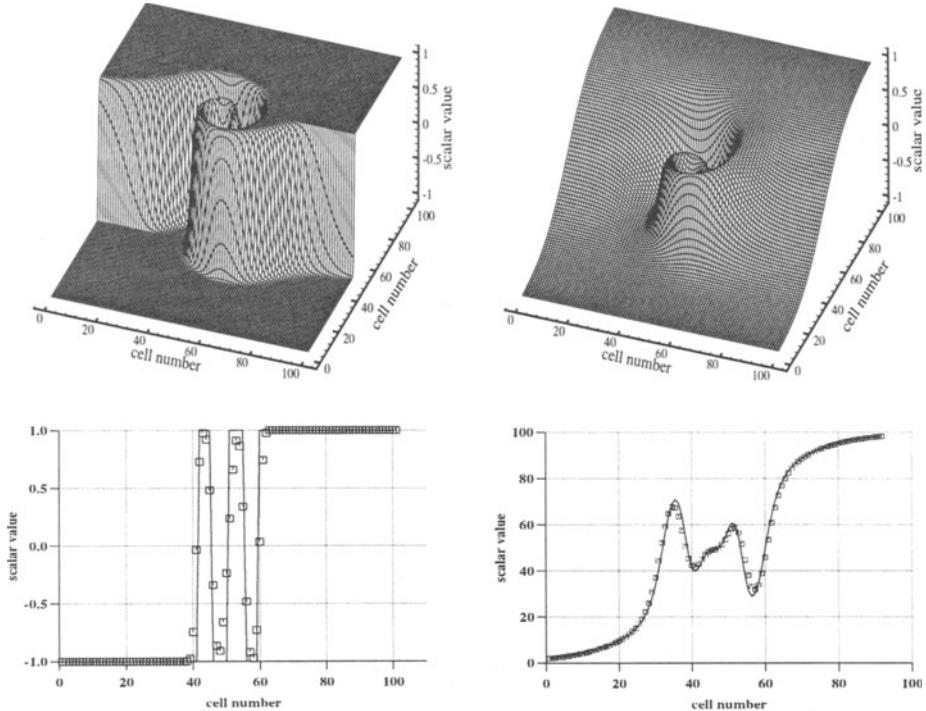


Figure 2. Top: the scalar field of the numerical solution for the cyclogenesis problem with discontinuous and smooth initial conditions, using the WAF2 scheme. Bottom: superimposed exact (lines) and numerical (points) solutions for data slices.

figure 2, for initial data in the form of a (discontinuous) step function and a smooth hyperbolic tangent distribution; the degree of accuracy of the scheme can be assessed by superimposed exact and numerical solutions for data retrieved through sections of the scalar field. The scheme captures smooth profiles well and there is little smearing of the discontinuities; no spurious oscillations appear. The results are very similar to those obtained with WAF1 (Nikiforakis and Toro, 2000), with only half the number of time steps.

4. The scheme for global atmospheric transport

The WAF2 scheme was implemented in a model of chemistry and transport, first presented in the paper by Nikiforakis and Toro (2000) for a WAF1 solver. The schemes have to be discretised on a regular latitude-longitude

grid (on a spherical surface), where the convergence of the meridians to the poles results in cells of variable size. To accommodate the change in cell area along constant longitude sweeps we use the general finite volume update formula

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{A_i} (f_{i+1/2} L_{i+1/2} - f_{i-1/2} L_{i-1/2}), \quad (8)$$

along both longitude and latitude directions, where $L_{i+1/2}$ are lengths associated with the side through which the flux is calculated and A_i is the area of the i th cell, A_i and $L_{i+1/2}$ being calculated on the spherical surface. The timestep is calculated at the beginning of every iteration. It depends on the ration of the magnitude of the wind velocity components along constant latitudes and longitudes and the corresponding cell-widths.

The model is initialised with fields either of chemical tracers or potential vorticity supplied by meteorological analyses. The latter also provide the winds at regular intervals (usually every six hours), which are then interpolated for shorter times, according to the CFL number of the scheme which prescribes the Δt chosen. In this paper we show results obtained on a single stratospheric surface.

As a validation exercise, a period during the 1991-92 Northern hemisphere winter is considered; this was studied as part of the EASOE (European Arctic Stratospheric Ozone Experiment) campaign (November 1991 - April 1992) which was planned in an atmosphere of growing concern about potential stratospheric ozone decline over populated latitudes of the Northern hemisphere, in order to improve our understanding of its behaviour. The evolution of the stratosphere during this period has been thoroughly documented, making it a well-suited operational-case benchmark. The high degree of distortion of the Northern polar vortex, and the filamentary structures related with wave breaking events, make it a demanding test of the capabilities of any method.

The model has been used for a single-layer (450K) integration using ECMWF analyses to advect potential vorticity (PV - a measure of the spin of air, conserved by a parcel of fluid under certain conditions used by atmospheric scientists) in the period going from 16 to 26 January 1992. Results using WAF1 and WAF2 are shown in Figure 3 for 20 and 26 January 1992. The main features of the flow, described in a paper written by Plumb *et al* in 1994, are accurately captured in these simulations. These include the distinctive kidney shape of the polar vortex on the 20th, and the fine filamentary structure which develops after the 22 January and eventually wraps around the main body of the vortex, which is clearly visible over the US on the 26 January. The monotonicity of the methods maintains the steep PV gradients across the edge of the vortex and along the filaments without any spurious oscillations appearing in their vicinity. Also,

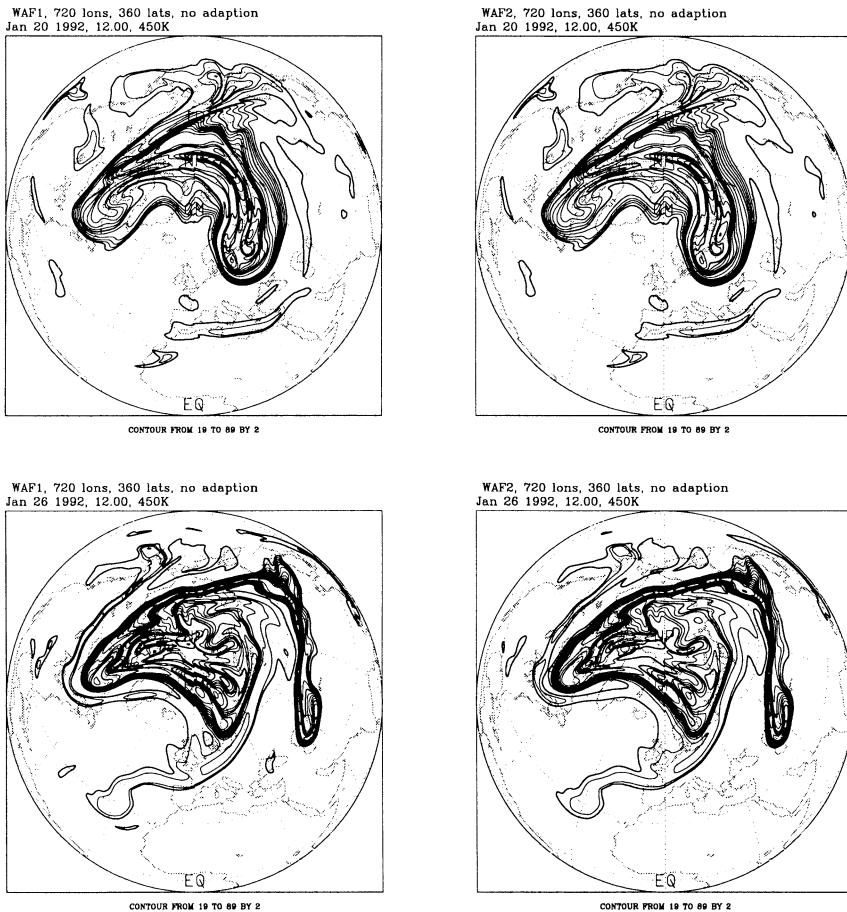


Figure 3. Potential vorticity on the 450K stratospheric surface at 12 o'clock midday, 20 and 26 January 1992, using the WAF1 (left) and WAF2 schemes; the results are nearly identical, but the cpu cost of the WAF2 simulation was significantly lower.

the fine filamentary structures remain intact owing to the low diffusivity of the methods.

The results the two integrations are very similar, but the contour lines cannot be expected to be identical because of the factor of two difference in the number of timesteps, which affects cumulative numerical errors. The cpu times for these runs are compared in table 1. The use of WAF2 for CFL number smaller than one results in increased cpu time, because of the overheads due to the extra cells taken into account while calculating the fluxes. However, for a CFL number twice as large, WAF2 enables a 30% saving on the WAF1 run; further savings are expected once the WAF2 scheme is optimised.

solver	normalised times
WAF1 at CFL no.= 0.9:	1
WAF2 at CFL no.= 0.9:	1.43
WAF2 at CFL no.= 1.9:	0.72

TABLE 1. Normalised cpu times for the single-layer stratospheric integrations, using WAF1 and WAF2.

5. Concluding remarks

We have presented an implementation of a WAF-type scheme with CFL number 2 restriction, in a global atmospheric Chemistry and Transport Model, enabling a substantial cpu-time saving, compared to its CFL number 1 counterpart, without sacrificing accuracy. This scheme can be extended to solve dynamical systems (eg. the primitive equations in general circulation models); it is suitable for use in three-dimensional Adaptive Mesh Refinement (AMR) models.

Acknowledgements

This work was supported by the Natural Environment Research Council via UGAMP (UK Universities Global Atmospheric Modelling Program). The authors would like to thank Dr M.E. Hubbard for his help in the preparation of this paper.

References

- Davies-Jones, R. 1985 Comments on "A kinematic analysis of frontogenesis associated with a non-divergent vortex". *J. Atm. Sci.* **42**, 2073-2075.
- Nikiforakis N. and Toro E.F. 1996. Evaluation of the WAF method for kinematic and dynamic atmospheric modelling problems. *Numerical Methods for Fluid Dynamics V*, Morton and Baines (eds) Oxford University Press.
- Nikiforakis N. and Toro E.F. 2000. Riemann-problem-based methods for global atmospheric off-line equations; submitted to *Quart. J. Roy. Met. Soc..*
- Plumb, R. A., Waugh, D. W., Atkinson, R. J., Newman, P. A., Lait, L. R., Schoeberl, M. R., Browell, E. V., Simmons, A. J. and Lowenstein, M., 1994 Intrusions into the lower stratospheric Arctic vortex during the winter 1991-1992. *J. Geoph. Res.*, **99**, D1,1089-1105.
- Toro E F (1989). A Weighted Average Flux method for hyperbolic conservation laws. *Proc. Roy. Soc. London*, **A423**, 401-418.
- Toro E F (1999). Riemann Solvers and Numerical Methods for Fluid Dynamics. Second Edition, Springer-Verlag.
- Toro E.F. and Billett S.J. 1997. A unified Riemann-problem-based extension of the Warming-Beam and Lax-Wendroff schemes. *IMA J. Num. Ana.* **17**, 61-102.

APPROXIMATE RIEMANN SOLVERS, GODUNOV SCHEMES AND CONTACT DISCONTINUITIES

B. WENDROFF

Retired Fellow, Los Alamos National Laboratory

Los Alamos, NM 87544, U.S.A.,

Adjunct Professor, University of New Mexico

Email: bbw@lanl.gov

Web: <http://math.unm.edu/~bbw>

Abstract. This presentation divides logically into three parts. In Part I we briefly review the one-dimensional (1D) 3-state approximate Riemann solver (HLL) of [A. HARTEN, P. D. LAX AND B. VAN LEER, *On upstream differencing and Godunov-type schemes for hyperbolic conservation laws*, SIAM Rev., 25 (1983), pp. 35-61], and its implementation (HLLE) of [B. EINFELDT, *On Godunov-type methods for gas dynamics*, SIAM J. Numer. Anal., 25 (1988), pp. 294-318]. We propose here a two-dimensional (2D) 9-state approximate Riemann solver and a corresponding 2D Godunov scheme. The result of applying this to some 2D shock interaction problems studied in [C. W. SCHULZ-RINNE, J. P. COLLINS, AND H. M. GLAZ, *Numerical solution of the Riemann problem for two-dimensional gas dynamics*, SIAM J. Sci. Comput., 14 (1993), pp. 1394-1414] are shown.

In Part II we return to one dimension and review the 4-state contact-corrected version HLLC developed by Toro and others [E. F. TORO, *Riemann Solvers and Numerical Methods for Fluid Dynamics*, Springer, 1997]. This version preserves a stationary contact discontinuity exactly. Since, as is well known, the Lax-Friedrichs (LF) scheme can be considered a special case of HLL, but involving a single constant state rather than three, a 2-state contact corrected version LFC that has the same property is easily formulated. This can also be used as the predictor in a two step Lax-Wendroff scheme, leading to LWC. We show how LFC, LWC, and their composition, [R. LISKA AND B. WENDROFF, *Composite schemes for conservation laws*, SIAM J. Numer. Anal., 35 (1998), pp. 2250-2271], behave on some of the test problems in Toro's book. In Part III we explore the possibility of extending the contact correction to two dimensions. The easiest way to do this is by dimensional splitting and the result of this is shown on a

two dimensional shock contact interaction problem, also from the work of Schulze-Rinne et al. A truly two-dimensional 4-state extension of LFC is feasible and the result of this for the same interaction problem is shown. In principle HLLC could be formulated in two dimensions using the same construction as was used for HLLE, but this would involve 16 constant states and the associated Godunov scheme would likely be very inefficient.

1. Introduction

There are three parts to this work. Part I is concerned with approximate Riemann solvers and associated Godunov schemes in one and two dimensions. In section 2 we briefly review the one-dimensional (1D) 3-state approximate Riemann solver and associated Godunov scheme HLL of Harten, Lax, and van Leer (Harten et al, 1983) and its implementation HLLE of Einfeldt (Einfeldt, 1988). In section 3 we show what a truly two-dimensional (2D) Godunov's method would be if one could solve 2D Riemann problems exactly. We then present several nine-state 2D versions of HLLE in section 3.1. We did this for the four-quadrant Riemann problem, but in principle this could be done for other configurations. The associated Godunov scheme is outlined, and then some computational details are described in section 3.2. In section 4 we present the result of two calculations, namely, cases 3 and 4 from (Schulz-Rinne et al, 1993). These are four-shock configurations and were chosen simply to show that our 2D HLLE scheme is consistent with the solution of these problems obtained by more accurate means. In Part II we return to 1D and consider the 4-state contact-corrected version HLLC developed by Toro and others (Toro, 1999), (Batten et al, 1997). This version preserves a stationary contact discontinuity exactly. Since, as is well known, the Lax-Friedrichs (LF) scheme can be considered a special case of HLL, but involving a single constant state rather than three, a 2-state contact corrected version LFC that has the same property is easily formulated. This can also be used as the predictor in a two step Lax-Wendroff scheme, leading to LWC. We show how LFC, LWC, and their composition behave on some of the test problems in Toro's book. However, we show by some examples that exact stationary contact preservation is not always a good thing.

In Part III we explore the possibility of extending the contact correction to two dimensions. The easiest way to do this is by dimensional splitting and we show the result of this on a two dimensional shock contact interaction problem, also from the work of Schulze-Rinne et al. A truly two-dimensional 4-state extension of LFC is feasible and the result of this for the same

interaction problem is shown. Figure 18 shows that there are difficulties with this particular formulation. In principle HLLC could be formulated in two dimensions using the same construction as was used for HLLE, but this would involve 16 constant states and the associated Godunov scheme would likely be very inefficient.

2. Godunov's Method

We begin with a review of Godunov's original method, following (Einfeldt, 1988). Consider the system of conservation laws

$$v_t + f_x(v) = 0,$$

where

$$v = \begin{pmatrix} \rho \\ m \\ E \end{pmatrix}, \quad f(v) = \begin{pmatrix} m \\ \frac{m^2}{\rho} + p \\ \frac{m}{\rho}(E + p) \end{pmatrix}.$$

The gas dynamic state consists of the density ρ , the velocity q , and internal energy density e . Then $m = \rho q$, $E = \rho(e + \frac{1}{2}q^2)$, and the pressure p is given by an equation of state, $p = p(\rho, e)$. We restrict ourselves to the ideal gas equation of state, $p = (\gamma - 1)\rho e$.

The integral form of the conservation law is

$$\begin{aligned} \int_a^b v(x, t_1) dx &= \int_a^b v(x, t_0) dx - \int_{t_0}^{t_1} f(v(b, s)) ds \\ &\quad + \int_{t_0}^{t_1} f(v(a, s)) ds \end{aligned} \quad (1)$$

for any $a < b$ and $0 \leq t_0 < t_1$.

The Riemann problem (centered at x_0, t_0) asks for the solution of (1) with initial data

$$v(x, t_0) = \begin{cases} u_0, & x < x_0 \\ u_1, & x > x_0 \end{cases}. \quad (2)$$

The solution is a function of $(x - x_0)/(t - t_0)$,

$$\omega((x - x_0)/(t - t_0)) \quad (3)$$

The structure of this solution is well-known. A description in terms of elementary waves (shock, contact or rarefaction), and an algorithm for solving the nonlinear equations defining the solution can be found in (Toro, 1999). A rigorous mathematical treatment is in (Smoller, 1994).

The Godunov scheme for general initial data is first to approximate the initial states by piecewise constant states, then use the solution of the local Riemann problems to construct new piecewise constant states by averaging. Thus, suppose we have a grid with cell centers $x_i = i\Delta x$, $i = 0, \pm 1, \pm 2, \dots$, cell endpoints $x_{i+\frac{1}{2}} = (i + \frac{1}{2})\Delta x$ and discrete time steps $t^n = n\Delta t$. Then given

$$v_i^n = \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} v(x, t^n) dx$$

Godunov defines v_i^{n+1} by

$$v_i^{n+1} = \frac{1}{\Delta x} \int_0^{\frac{\Delta x}{2}} \omega_{i-\frac{1}{2}}^n(x/\Delta t) dx + \frac{1}{\Delta x} \int_{-\frac{\Delta x}{2}}^0 \omega_{i+\frac{1}{2}}^n(x/\Delta t) dx, \quad (4)$$

where $\omega_{i-\frac{1}{2}}^n((x - x_{i-\frac{1}{2}})/(t - t^n))$ is the solution of (1 and (2) centered at $(x_{i-\frac{1}{2}}, t^n)$ and with $u_0 = v_{i-1}^n$, $u_1 = v_i^n$. Since an exact Riemann solution is used, it follows from (1) that there is a numerical flux F such that

$$v_i^{n+1} = v_i^n - \lambda(F_{i+\frac{1}{2}}^n - F_{i-\frac{1}{2}}^n), \quad (5)$$

where

$$F_{i+\frac{1}{2}}^n = f(\omega_{i+\frac{1}{2}}^n(0)), \quad (6)$$

and

$$\lambda = \frac{\Delta t}{\Delta x},$$

and therefore the integrals in (4) do not have to be computed.

2.1. HLLE

The approximate solution of (2) proposed by Einfeldt (Einfeldt, 1988), based on the HLL solver of Harten, Lax and van Leer (Harten et al, 1983), and now known as the HLLE solver consists of two speeds, $b^0 < b^1$, the two initial states u_0 and u_1 , and a third intermediate state $u_{\frac{1}{2}}$. Then an approximate ω is (for $x_0 = 0$, $t_0 = 0$)

$$\omega(x/t) = \begin{cases} u_0, & x < b^0 t \\ u_{\frac{1}{2}}, & b^0 t < x < b^1 t \\ u_1, & x > b^1 t \end{cases}, \quad (7)$$

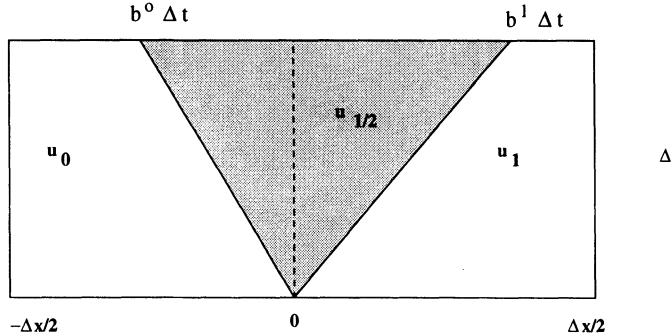


Figure 1. Structure of the HLLE solution

where the left and right speeds $b^0 < b^1$ and the intermediate state $u_{\frac{1}{2}}$ are to be determined. The speeds are meant to be approximations to the smallest and largest signal velocities of the exact Riemann problem. Einfeldt assumes that

$$\Delta t \max(|b^0|, |b^1|) \leq \Delta x/2. \quad (8)$$

The intermediate state is defined to satisfy the integral conservation law in the interval $(-\Delta x/2, \Delta x/2)$, that is,

$$\int_{-\Delta x/2}^{\Delta x/2} \omega(x'/t) dx = \frac{\Delta x}{2}(u_0 + u_1) - \Delta t(f(u_1) - f(u_0)),$$

or

$$u_{\frac{1}{2}} = \frac{b^1 u_1 - b^0 u_0 + f(u_0) - f(u_1)}{b^1 - b^0}. \quad (9)$$

Figure 1 shows the space time structure of the approximate solution. For simplicity, the figure is drawn assuming that $b^0 \leq 0$ and $b^1 \geq 0$.

The Godunov scheme at a generic mesh cell ($i = 0$) advances the state u_0 to its new value u_0^1 according to

$$\Delta x u_0^1 = b_{-\frac{1}{2}}^1 \Delta t u_{-\frac{1}{2}} - b_{\frac{1}{2}}^0 \Delta t u_{\frac{1}{2}} + (\Delta x - (b_{-\frac{1}{2}}^1 - b_{\frac{1}{2}}^0) \Delta t) u_0$$

as in Figure 2.

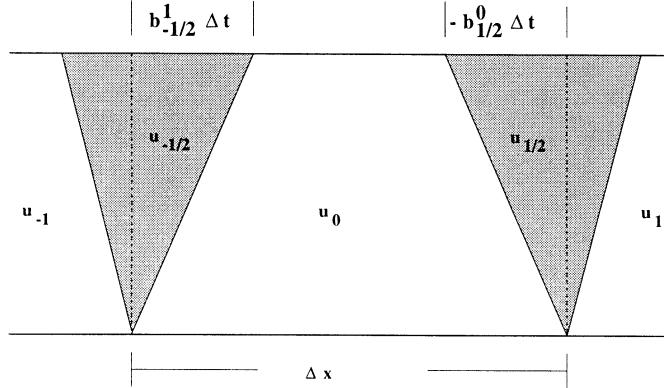


Figure 2. The HLLE-Godunov scheme

However, just as in the exact case, there is a flux form. For the full set of difference equations we have (5) with

$$F_{i+\frac{1}{2}} = \frac{b_{i+\frac{1}{2}}^+ f(v_i) - b_{i+\frac{1}{2}}^- f(v_{i+1})}{b_{i+\frac{1}{2}}^+ - b_{i+\frac{1}{2}}^-} + \frac{b_{i+\frac{1}{2}}^+ b_{i+\frac{1}{2}}^-}{b_{i+\frac{1}{2}}^+ - b_{i+\frac{1}{2}}^-} (v_{i+1} - v_i), \quad (10)$$

where

$$b_{i+\frac{1}{2}}^+ = \max(0, b_{i+\frac{1}{2}}^1), \quad b_{i+\frac{1}{2}}^- = \min(0, b_{i+\frac{1}{2}}^0). \quad (11)$$

The speeds remain to be chosen. The speeds proposed in (Einfeldt, 1988) are defined in Appendix A for the 2D gas dynamic equations in terms of Roe averages. For the 1D system here just suppress the second index, j , and set $q^x = q$, $q^y = 0$.

2.2. STABILITY, ENTROPY, AND POSITIVITY

The derivation of the exact and approximate Godunov schemes assumes that waves generated at the cell endpoints do not interact during one time step, that is, that (8) holds. In fact, as pointed out in (Harten et al, 1983), the exact Godunov scheme is valid as long as waves from one endpoint do not reach an adjacent endpoint in one time step. A similar situation holds for the HLLE scheme, in that there is a stability condition derived in (Einfeldt, 1988) that is less restrictive than (8). To review this condition, suppose that f' is a constant matrix A , with eigenvalues $\beta_1 < \beta_2 < \beta_3$. Then

$$b^0 = b_{i+\frac{1}{2}}^0 = \beta_1, \quad b^1 = b_{i+\frac{1}{2}}^1 = \beta_3,$$

and

$$v_i^{n+1} = (1 - \lambda\sigma)v_i^n + \frac{1}{2}\lambda(\sigma + A)v_{i-1}^n + \frac{1}{2}\lambda(\sigma - A)v_{i+1}^n,$$

where

$$\sigma = \frac{A(b^+ + b^-) - 2b^+b^-}{b^+ - b^-}. \quad (12)$$

By construction, $\sigma \pm A \geq 0$, so all coefficients are positive if $\lambda\sigma \leq 1$, which is the case if

$$\max(b^+, -b^-)\lambda \leq 1. \quad (13)$$

It is shown in (Einfeldt et al, 1991) that both the HLLE scheme with its stability condition and the exact Godunov scheme are positivity-conserving, that is, if the density and internal energy are initially positive then they are for all time. We refer the reader to (Harten et al, 1983) for a definition and discussion of entropy and the entropy inequality. It is shown there that exact Godunov satisfies the entropy inequality, and that a scheme like HLLE does also if the speeds b^0 and b^1 bound the exact wave speeds. According to (Batten et al, 1997) this is not always the case for the choice for the chosen speeds, thus, the behavior of entropy for HLLE is unknown.

3. Godunov and HLLE in Two Dimensions

The system of conservation laws is

$$v_t + f_x(v) + g_y(v) = 0,$$

where

$$v = \begin{pmatrix} \rho \\ m \\ n \\ E \end{pmatrix}, \quad f(v) = \begin{pmatrix} m \\ \frac{m^2}{\rho} + p \\ \frac{mn}{\rho} \\ \frac{m}{\rho}(E + p) \end{pmatrix}, \quad g(v) = \begin{pmatrix} n \\ \frac{mn}{\rho} \\ \frac{n^2}{\rho} + p \\ \frac{n}{\rho}(E + p) \end{pmatrix}, \quad (14)$$

where if q^x, q^y are respectively the x and y components of velocity then $m = \rho q^x$, $n = \rho q^y$, $E = \rho(e + \frac{1}{2}((q^x)^2 + (q^y)^2))$, and p and e are pressure and internal energy, as before.

The integral form of the conservation law, specifically for rectangles, is

$$\begin{aligned} \int_a^b \int_c^d v(x, y, t_1) dx dy &= \int_a^b \int_c^d v(x, y, t_0) dx dy \\ &- \int_{t_0}^{t_1} \int_c^d f(v(b, y, s)) dy ds + \int_{t_0}^{t_1} \int_c^d f(v(a, y, s)) dy ds \\ &- \int_{t_0}^{t_1} \int_a^b g(v(x, d, s)) dx ds + \int_{t_0}^{t_1} \int_a^b g(v(x, c, s)) dx ds \end{aligned} \quad (15)$$

for any $a < b$, $c < d$ and $0 \leq t_0 < t_1$.

There is an infinite variety of two-dimensional Riemann problems, but only one type which is relevant for a rectangular grid, namely, the four quadrant problem (centered at x_0, y_0, t_0), that is, find $v(x, y, t)$ satisfying (15) with initial data

$$v(x, y, t_0) = \begin{cases} u_{00}, & x < x_0, y < y_0 \\ u_{01}, & x < x_0, y > y_0 \\ u_{11}, & x > x_0, y > y_0 \\ u_{10}, & x > x_0, y < y_0 \end{cases}. \quad (16)$$

Unlike the situation in one dimension, where everything is known about the Riemann problem, here almost nothing is known. The complexity of the solution can be seen from the numerical results of (Schulz-Rinne et al, 1993). Nevertheless, we can proceed formally and suppose a solution which is a function of $((x - x_0)/(t - t_0), (y - y_0)/(t - t_0))$,

$$\omega((x - x_0)/(t - t_0), (y - y_0)/(t - t_0)). \quad (17)$$

A literal extension of Godunov's idea for general initial data in two dimensions is first to approximate the initial states by piecewise constant states, then use the solution of the two-dimensional local Riemann problems to construct new piecewise constant states by averaging. Thus, suppose we have a uniform rectangular grid with cell centers $x_i = i\Delta x$, $y_j = j\Delta y$, $i, j = 0, \pm 1, \pm 2, \dots$, cell vertices $x_{i+\frac{1}{2}} = (i + \frac{1}{2})\Delta x$, $y_{j+\frac{1}{2}} = (j + \frac{1}{2})\Delta y$ and discrete time steps $t^n = n\Delta t$.

Call the $\Delta x \times \Delta y$ cells centered on the vertices the *dual cells*, and call the original cells the *primary cells*. The cell arrangement and the Riemann data are shown in Figure 3.

Then given piecewise-constant data in the primary cells, namely,

$$v_{i,j}^n = \frac{1}{\Delta x \Delta y} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} v(x, y, t^n) dy dx$$

we could define $v_{i,j}^{n+1}$ by solving the 2D Riemann problem in each dual cell and projecting these solutions onto the primary cells, obtaining

$$\begin{aligned} v_{i,j}^{n+1} = & \frac{1}{\Delta x \Delta y} \int_0^{\frac{\Delta x}{2}} \int_0^{\frac{\Delta y}{2}} \omega_{i-\frac{1}{2}, j-\frac{1}{2}}^n(x/\Delta t, y/\Delta t) dy dx \\ & + \frac{1}{\Delta x \Delta y} \int_0^{\frac{\Delta x}{2}} \int_{-\frac{\Delta y}{2}}^0 \omega_{i-\frac{1}{2}, j+\frac{1}{2}}^n(x/\Delta t, y/\Delta t) dy dx \end{aligned}$$

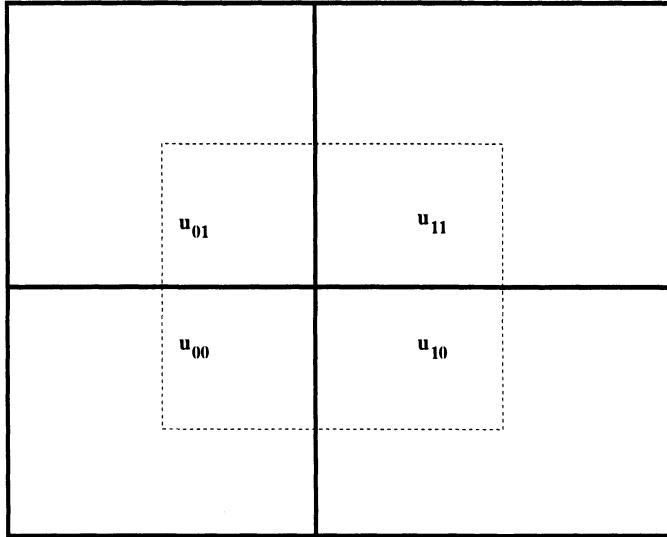


Figure 3. Primary (dark lines) and dual (dotted lines) cell arrangement with Riemann data

$$\begin{aligned}
 & + \frac{1}{\Delta x \Delta y} \int_{-\frac{\Delta x}{2}}^0 \int_{-\frac{\Delta y}{2}}^0 \omega_{i+\frac{1}{2}, j+\frac{1}{2}}^n(x/\Delta t, y/\Delta t) dy dx \\
 & + \frac{1}{\Delta x \Delta y} \int_{-\frac{\Delta x}{2}}^0 \int_0^{\frac{\Delta y}{2}} \omega_{i+\frac{1}{2}, j-\frac{1}{2}}^n(x/\Delta t, y/\Delta t) dy dx
 \end{aligned} \quad (18)$$

where

$$\omega_{i-\frac{1}{2}, j-\frac{1}{2}}^n((x - x_{i-\frac{1}{2}})/(t - t^n), (y - y_{j-\frac{1}{2}})/(t - t^n))$$

is the solution of (15) and (16) centered at $(x_{i-\frac{1}{2}}, y_{j-\frac{1}{2}}, t^n)$ and with $u_{00} = v_{i-1, j-1}^n$, $u_{01} = v_{i-1, j}^n$, $u_{11} = v_{i, j}^n$, $u_{10} = v_{i, j-1}^n$.

This construction only makes sense if waves generated by the 2D Riemann problems in any dual cells do not interact with the waves from its neighbors. So if σ is the maximum signal speed, maximized over all directions and all vertices, then we require $\sigma \Delta t < \min(\Delta x/2, \Delta y/2)$.

3.1. SOME NINE-STATE RIEMANN SOLVERS

Unfortunately, the preceding section does not define a numerical method since there is no possibility of solving (16) exactly. This can be seen in (Schulz-Rinne et al, 1993), where it is shown that even the simplest non-trivial 2D Riemann problems have extremely complex solutions. So let us

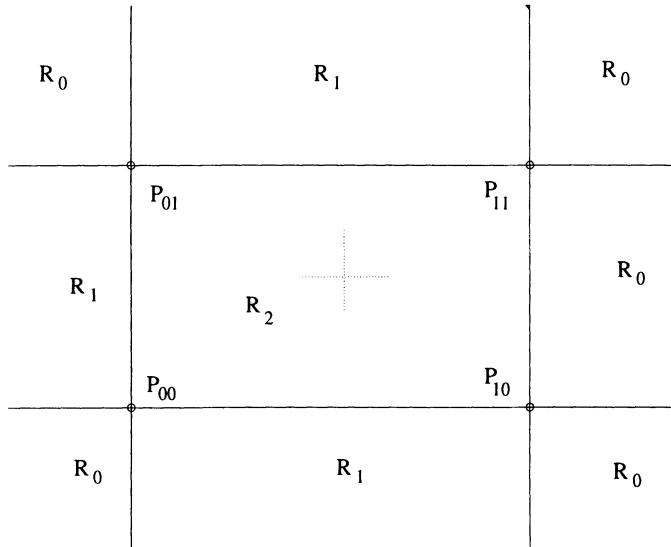


Figure 4. Structure of the 2D solution at later time. Dotted lines indicate initial center of the Riemann problem

consider what an approximate solution might look like in Figure 4. This is a schematic of a dual cell after a small time interval of evolution of the 2D Riemann problem. Because signals propagate with finite speed, the cell can be divided into nine pieces of three different types, R_0 , R_1 , and R_2 . In the corner R_0 regions the data are undisturbed from their initial states. In the edge R_1 regions the states are completely determined by the associated 1D Riemann problems. The central region R_2 contains the 2D interaction.

In analogy with the one-dimensional case, our two-dimensional HLLE solution will define nine specific regions in terms of speeds, and then assign a constant state to each of the nine regions. There are several ways to do this, but in each instance the four corner regions will have their initial values, in the edge strips the constant state will be the intermediate state obtained from a one-dimensional HLLE solver, and then the integral conservation law (15) will be used to define the central constant state.

Consider first Figure 5a. This shows a possible nine state configuration (Method 1) in which points on opposite edges of the cell have been joined by straight lines. The points on the edges will be defined by 1D HLLE speeds. This method is rather inefficient as it requires computation of intersections of lines and areas of quadrilaterals.

A second possibility (Method 2) is shown in Figure 5b. Again, the points on the edges are defined by 1D HLLE speeds, but now the corner regions

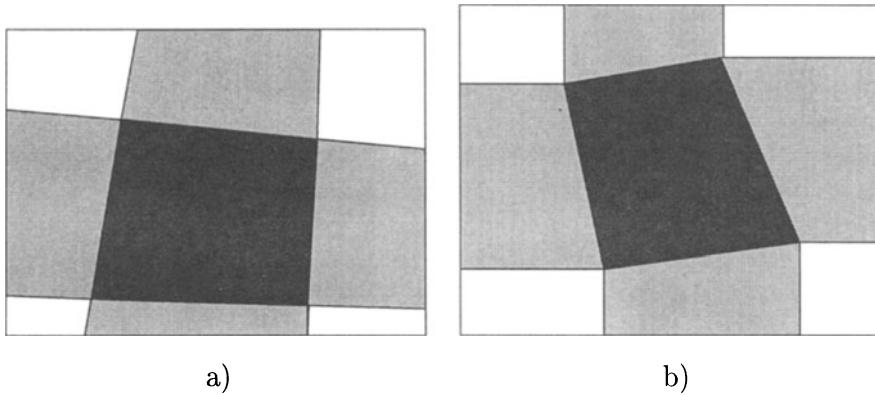


Figure 5. a): Method 1, direct connection to edges. b) Method 2, trapezoidal edge sets, quadrilateral central set

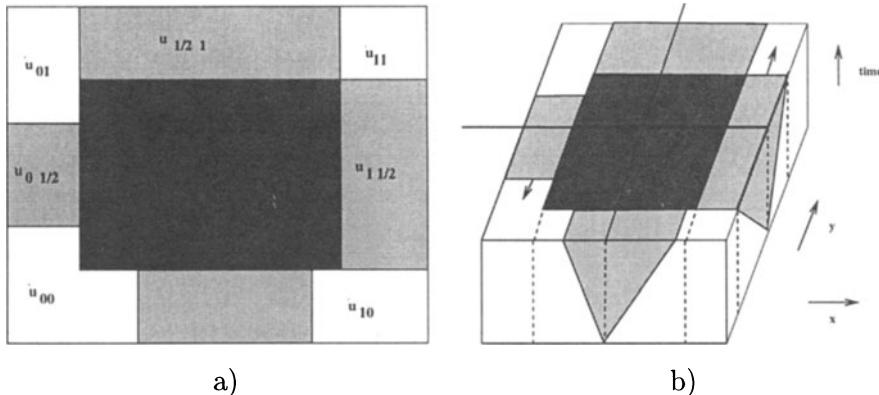


Figure 6. a): Method 3, rectangular sets. b) Space-time structure for Method 3

are rectangles and the edge strips are trapezoids. This is the construction used in (Wendroff, 1999).

A third arrangement (Method 3) is shown in Figure 6a. Here, the edge and central regions are rectangles. This is slightly less efficient than the second method above, but it has the advantage that there is a formula, (44), for the value of the central state that clearly is a generalization of (9). There also is a flux form, but it is very complicated and only of possible theoretical use. Figure 6b is a sketch of the time evolution of this approximate Riemann solver.

Consider only the latter two methods. The regions of constancy will

depend on the 1D Einfeldt speeds.

Consider the cell vertex at $(x_{\frac{1}{2}}, y_{\frac{1}{2}})$. The four states interacting at this vertex are (suppressing the time index) $u_{00} = v_{0,0}$, $u_{10} = v_{1,0}$, $u_{11} = v_{1,1}$, $u_{01} = v_{0,1}$. For Method 2 ideally we would like to use the following velocities to define the central region:

$$\begin{aligned} b_{\frac{1}{2}, \frac{1}{2}}^{01} &= (b_{\frac{1}{2}, 1}^0, b_{0, \frac{1}{2}}^1), \\ b_{\frac{1}{2}, \frac{1}{2}}^{11} &= (b_{\frac{1}{2}, 1}^1, b_{1, \frac{1}{2}}^1), \\ b_{\frac{1}{2}, \frac{1}{2}}^{10} &= (b_{\frac{1}{2}, 0}^1, b_{1, \frac{1}{2}}^0), \\ b_{\frac{1}{2}, \frac{1}{2}}^{00} &= (b_{\frac{1}{2}, 0}^0, b_{0, \frac{1}{2}}^0). \end{aligned} \quad (19)$$

(See Appendix A for the detailed definition of these speeds). For example, $b_{\frac{1}{2}, 1}^0$ and $b_{\frac{1}{2}, 1}^1$ are the left and right HLLE speeds for the one-dimensional Riemann problem in the x-direction (with flux f) with $u_0 = v_{0,1}$ and $u_1 = v_{1,1}$, while $b_{0, \frac{1}{2}}^0$ and $b_{0, \frac{1}{2}}^1$ are the left and right HLLE speeds for the one-dimensional Riemann problem in the y-direction (with flux g) with $u_0 = v_{0,0}$ and $u_1 = v_{0,1}$.

However, since it is possible that $b^0 > 0$ or $b^1 < 0$, the corner rectangles might not be contained in their respective quadrants, as they are in Figure 4. If that happens the picture might in some cases not be as simple as shown and the computer code would have to be quite complicated in order to allow for this situation. Although it might be worth the extra work, in order to simply show the feasibility and consistency of the 2D HLLE scheme, we have chosen to avoid this difficulty at the cost of introducing somewhat more diffusion than is present in the 1D HLLE method. What is needed is to use (11) and replace (19) with

$$\begin{aligned} b_{\frac{1}{2}, \frac{1}{2}}^{01} &= (b_{\frac{1}{2}, 1}^-, b_{0, \frac{1}{2}}^+), \\ b_{\frac{1}{2}, \frac{1}{2}}^{11} &= (b_{\frac{1}{2}, 1}^+, b_{1, \frac{1}{2}}^+), \\ b_{\frac{1}{2}, \frac{1}{2}}^{10} &= (b_{\frac{1}{2}, 0}^+, b_{1, \frac{1}{2}}^-), \\ b_{\frac{1}{2}, \frac{1}{2}}^{00} &= (b_{\frac{1}{2}, 0}^-, b_{0, \frac{1}{2}}^-). \end{aligned} \quad (20)$$

The vertices of the interior quadrilateral (refer to Figure 4) are

$$\begin{aligned} P^{00}(t) &= (x_{\frac{1}{2}}, y_{\frac{1}{2}}) + b_{\frac{1}{2}, \frac{1}{2}}^{00}(t - t^n), \\ P^{01}(t) &= (x_{\frac{1}{2}}, y_{\frac{1}{2}}) + b_{\frac{1}{2}, \frac{1}{2}}^{01}(t - t^n), \end{aligned}$$

$$\begin{aligned} P^{11}(t) &= (x_{\frac{1}{2}}, y_{\frac{1}{2}}) + b_{\frac{1}{2}, \frac{1}{2}}^{11}(t - t^n), \\ P^{10}(t) &= (x_{\frac{1}{2}}, y_{\frac{1}{2}}) + b_{\frac{1}{2}, \frac{1}{2}}^{10}(t - t^n). \end{aligned}$$

For method 3, define new velocities associated with $(x_{\frac{1}{2}}, y_{\frac{1}{2}})$

$$\begin{aligned} d^{x-} &= d_{\frac{1}{2}, \frac{1}{2}}^{x-} = \min(b_{\frac{1}{2}, 1}^-, b_{\frac{1}{2}, 0}^-), \\ d^{x+} &= d_{\frac{1}{2}, \frac{1}{2}}^{x+} = \max(b_{\frac{1}{2}, 1}^+, b_{\frac{1}{2}, 0}^+), \\ d^{y-} &= d_{\frac{1}{2}, \frac{1}{2}}^{y-} = \min(b_{1, \frac{1}{2}}^-, b_{0, \frac{1}{2}}^-), \\ d^{y+} &= d_{\frac{1}{2}, \frac{1}{2}}^{y+} = \max(b_{1, \frac{1}{2}}^+, b_{0, \frac{1}{2}}^+) \end{aligned}$$

and then the four corners of the interior rectangle are

$$\begin{aligned} P^{00}(t) &= (x_{\frac{1}{2}}, y_{\frac{1}{2}}) + (d^{x-}, d^{y-})(t - t^n), \\ P^{01}(t) &= (x_{\frac{1}{2}}, y_{\frac{1}{2}}) + (d^{x-}, d^{y+})(t - t^n), \\ P^{11}(t) &= (x_{\frac{1}{2}}, y_{\frac{1}{2}}) + (d^{x+}, d^{y+})(t - t^n), \\ P^{10}(t) &= (x_{\frac{1}{2}}, y_{\frac{1}{2}}) + (d^{x+}, d^{y-})(t - t^n). \end{aligned}$$

We can now define the edge states (in the light shaded regions of Figs. 5a,b and 6a); namely $u_{0, \frac{1}{2}}$, $u_{\frac{1}{2}, 0}$, $u_{1, \frac{1}{2}}$, and $u_{\frac{1}{2}, 1}$. For example, $u_{0, \frac{1}{2}}$ is the HLLE intermediate state for the Riemann problem in the y-direction with $u_0 = u_{00}$ and $u_1 = u_{01}$ and with left speed $b_{0, \frac{1}{2}}^-$ and right speed $b_{0, \frac{1}{2}}^+$, and the others are defined similarly.

We can state the complete solution. Let $S_{\alpha, \beta}$ for $\alpha = 0, \frac{1}{2}, 1$, $\beta = 0, \frac{1}{2}, 1$ be the nine regions. Then

$$\omega_{\frac{1}{2}, \frac{1}{2}}(x, y) = u_{\alpha, \beta} \text{ for } (x, y) \in S_{\alpha, \beta}. \quad (21)$$

The ninth state $u_{\frac{1}{2}, \frac{1}{2}}$ is yet to be defined. Let $|S|_{\alpha, \beta}$ be the area of $S_{\alpha, \beta}$ at time $t = t^{n+1}$. Then based on (15) we can write the following approximate integral conservation form:

$$\begin{aligned} \sum_{\substack{\alpha = 0, \frac{1}{2}, 1 \\ \beta = 0, \frac{1}{2}, 1}} |S|_{\alpha, \beta} u_{\alpha, \beta} &= \frac{\Delta x \Delta y}{4} \sum_{\substack{\alpha = 0, 1 \\ \beta = 0, 1}} u_{\alpha, \beta} \\ &\quad - \Delta y \Delta t (F_{1, \frac{1}{2}} - F_{0, \frac{1}{2}}) \\ &\quad - \Delta x \Delta t (G_{\frac{1}{2}, 1} - G_{\frac{1}{2}, 0}), \end{aligned} \quad (22)$$

from which $u_{\frac{1}{2}, \frac{1}{2}}$ can be obtained if $|S|_{\frac{1}{2}, \frac{1}{2}} \neq 0$. The flux integrals will also be approximated using the 1D HLLE solutions. Note that (15) is being formulated over the space-time cell $C = \{x, y, t; x_0 \leq x \leq x_1, y_0 \leq y \leq y_1, t^n \leq t \leq t^{n+1}\}$. The flux integrals are integrals over the time-like faces of this cell. The structure of such a face is just as in Figure 1. Thus,

$$\begin{aligned} F_{0, \frac{1}{2}} &= \frac{1}{\Delta t \Delta y} \left[\frac{1}{2} (\Delta y - b_{0, \frac{1}{2}}^+ \Delta t) \Delta t f(u_{01}) \right. \\ &\quad + \frac{1}{2} (\Delta y + b_{0, \frac{1}{2}}^- \Delta t) \Delta t f(u_{00}) \\ &\quad \left. + \frac{1}{2} (b_{0, \frac{1}{2}}^+ - b_{0, \frac{1}{2}}^-) (\Delta t)^2 f(u_{0, \frac{1}{2}}) \right], \end{aligned} \quad (23)$$

and

$$\begin{aligned} G_{\frac{1}{2}, 0} &= \frac{1}{\Delta t \Delta x} \left[\frac{1}{2} (\Delta x - b_{\frac{1}{2}, 0}^+ \Delta t) \Delta t g(u_{10}) \right. \\ &\quad + \frac{1}{2} (\Delta x + b_{\frac{1}{2}, 0}^- \Delta t) \Delta t g(u_{00}) \\ &\quad \left. + \frac{1}{2} (b_{\frac{1}{2}, 0}^+ - b_{\frac{1}{2}, 0}^-) (\Delta t)^2 g(u_{\frac{1}{2}, 0}) \right]. \end{aligned} \quad (24)$$

3.2. THE 2D HLLE-GODUNOV SCHEME

Just as with 1D in section 2.1, we use the approximate Riemann solution in place of the exact one in (18).

The computation can be organized into two phases

PHASE 1.

For each dual cell edge store the pair of speeds, the 1D intermediate state, and the fluxes F and G .

PHASE 2.

For each dual cell define the nine pieces, compute their areas, and then find the central intermediate state. Substitute the approximation given by (21) into (18) to obtain new constant states in the primary cells. This is quite easy since with (20) the corner sets Q_i in Fig. 4 lie entirely in their respective quadrants. For method 2 each edge set is a trapezoid which now splits into two trapezoids, each lying entirely in a quadrant. The area of the intersection of the dual cell central region with a quadrant (that is, with a neighboring primary cell) is simply what is left over from the other three pieces. Referring to the primary cell sketch in Figure 7, for method 3 the edge sets are rectangles which also split the same way, the corners of the primary cells are now simple rectangles (dark-shaded), and the undisturbed region of the primary cell is now just the complement of the union of the

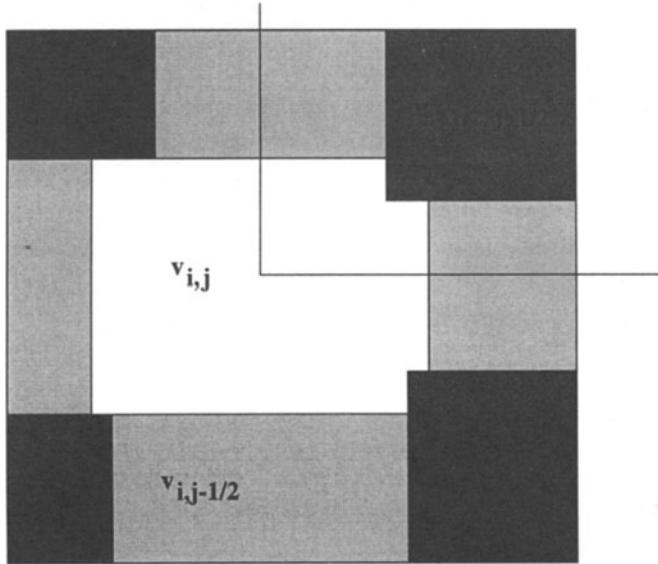


Figure 7. Primary cell

others. So the mapping onto the primary cells can be accomplished without a separate pass over the primary cells, by accumulating the contribution of each dual cell to its neighboring primary cells. In Figure 7, the states in the light-shaded regions are the 1D HLLE intermediate states determined by the adjacent cells, while the corner dark-shaded states come from the 2D interactions in the dual cells.

3.3. STABILITY

In Appendix B we present a flux form for method 3, that is, we show that

$$v_{i,j}^{n+1} = v_{i,j}^n - \frac{\Delta t}{\Delta x} (\mathcal{F}_{i+\frac{1}{2},j} - \mathcal{F}_{i-\frac{1}{2},j}) - \frac{\Delta t}{\Delta y} (\mathcal{G}_{i,j+\frac{1}{2}} - \mathcal{G}_{i,j-\frac{1}{2}}). \quad (25)$$

These fluxes are complicated and we do not see a way to perform the kind of stability analysis done in 1D in (Einfeldt, 1988). However, we can look at the linear case in which f and g have constant gradient, matrices, $f' = A$, $g' = B$. Even this is difficult if these matrices don't commute. If we assume that $AB = BA$, with common eigenvalues $\beta_1 < \beta_2 = \beta_3 < \beta_4$, then all the left speeds are the same, and all the right speeds are the same, that is we can set

$$d^- = d_{i+\frac{1}{2},j+\frac{1}{2}}^{x-} = d_{i+\frac{1}{2},j+\frac{1}{2}}^{y-} = \min(\beta_1, 0),$$

$$d^+ = d_{i+\frac{1}{2}, j+\frac{1}{2}}^{x+} = d_{i+\frac{1}{2}, j+\frac{1}{2}}^{y+} = \max(\beta_4, 0).$$

In this case the three 2D HLLE schemes form the same nine-point scheme that is, in fact, the product of the 1D schemes. Let

$$(T_x v)_{i,j} = v_{i+1,j}, \quad (T_y v)_{i,j} = v_{i,j+1},$$

and

$$\begin{aligned} H_y &= \left(\frac{1}{2} \lambda_y (\sigma_y - B) T_y + (1 - \sigma_y) I + \frac{1}{2} \lambda_y (B - \sigma_y) T_y^{-1} \right), \\ H_x &= \left(\frac{1}{2} \lambda_x (\sigma_x - A) T_x + (1 - \sigma_x) I + \frac{1}{2} \lambda_x (A - \sigma_x) T_x^{-1} \right), \end{aligned}$$

where $\lambda_x = \Delta t / \Delta x$, $\lambda_y = \Delta t / \Delta y$, and σ_x and σ_y are defined by (12) with the appropriate substitutions. Then

$$v_{i,j}^{n+1} = (H_x H_y v^n)_{i,j}$$

and in this very special case the 2D HLLE scheme is positive if (13) is satisfied.

For the numerical examples we used an adaptive time step determined by

$$\max(\lambda_x \kappa_x, \lambda_y \kappa_y) \leq \text{CFL} \leq 1,$$

where κ_x, κ_y are the wave speeds in each direction. The maximum is taken over the grid as well as over the pair.

3.4. ENTROPY AND POSITIVITY

There is little that can be said at this time about these difficult issues for the 2D HLLE scheme for gas dynamics, for several reasons. The computer program performing the operations described in section 3.2 is fairly simple, but the flux form of the finite difference scheme in terms of the speeds and the hydrodynamic states is very complicated, making an analysis of positivity extremely difficult. Concerning entropy, if we were using exact Riemann solvers for the 1D problems, exact integrals for the face fluxes and average intermediate states, and if the central region were guaranteed to contain all the 2D interactions, the convexity argument of (Harten et al, 1983) could presumably be used to establish an entropy inequality. Since none of these is the case, we are much worse off here concerning entropy than in 1D.

4. Numerical Examples

We have computed several of the 2D Riemann problems that were done in (Schulz-Rinne et al, 1993). This set of problems was arranged so that the 1D interactions in each of them produce just one wave. We have chosen configurations 4 and 3 which are each four-shock interaction problems. The data are, using the indexing of (16), for configuration 4, $\gamma=1.4$, and

$$\begin{aligned} p_{00} &= 1.1 & \rho_{00} = 1.1 & q_{00}^x = .8939 & q_{00}^y = .8939 \\ p_{01} &= .35 & \rho_{01} = .5065 & q_{01}^x = .8939 & q_{01}^y = 0. \\ p_{11} &= 1.1 & \rho_{11} = 1.1 & q_{11}^x = 0. & q_{11}^y = 0. \\ p_{10} &= .35 & \rho_{10} = .5065 & q_{10}^x = 0. & q_{10}^y = .8939. \end{aligned}$$

For configuration 3 the data are

$$\begin{aligned} p_{00} &= .029 & \rho_{00} = .138 & q_{00}^x = 1.206 & q_{00}^y = 1.206 \\ p_{01} &= .3 & \rho_{01} = .5323 & q_{01}^x = 1.206 & q_{01}^y = 0. \\ p_{11} &= 1.5 & \rho_{11} = 1.5 & q_{11}^x = 0. & q_{11}^y = 0. \\ p_{10} &= .3 & \rho_{10} = .5323 & q_{10}^x = 0. & q_{10}^y = 1.206. \end{aligned}$$

All versions of 2D HLLE gave the same result. Density contours for configuration 4, using a grid 400×400 , are shown in Figure 8. For comparison we show the same result for the composite scheme CFLF4, developed in (Liska and Wendroff, 1998). This configuration probably has the least complicated 2D structure of them all, and 2D HLLE does very well on it.

Configuration 3 with the same grid is shown in Figure 9, along with the result of the composite CFLF4. This case has a more complex interaction, and some resolution is lost by the HLLE scheme. All computations were done with $CFL=.9$.

For case 4 all the left speeds b^0 were negative and all the right speeds b^1 were positive, but for case 3 some left speeds were positive (this is true for the initial data).

5. Contact Correction in 1D

We now take up Part II of this paper, which, as indicated in the introduction, is a return to problems with 1D symmetry. However, since this will also be used in a 2D splitting scheme, we will use the 2D flux f from (14). We are going to present 1D versions of the Lax-Friedrichs (LF) and Lax-Wendroff (LW) difference schemes for the gas dynamic equations that have an upwind feature, in the sense that at each grid point the flow direction is used to modify the difference equations. The idea is based on the

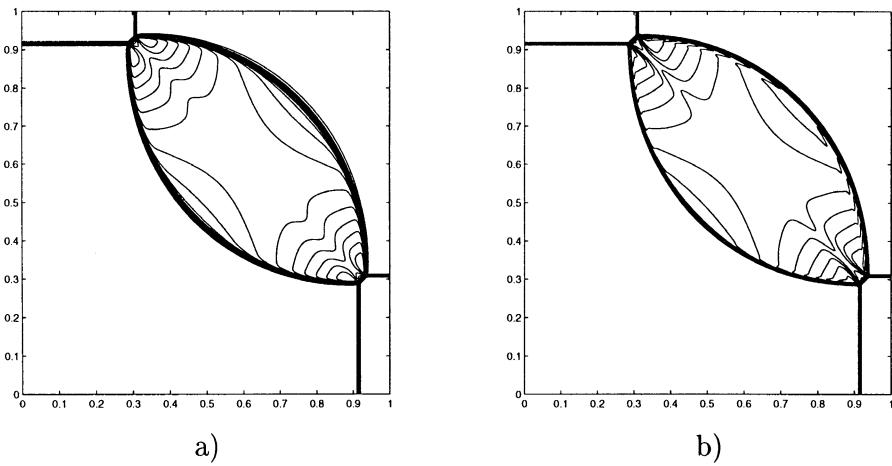


Figure 8. Configuration 4. 400×400 grid, time=.25, cfl=.9, 30 density contours: a) 2D HLLE. b) CFLF4

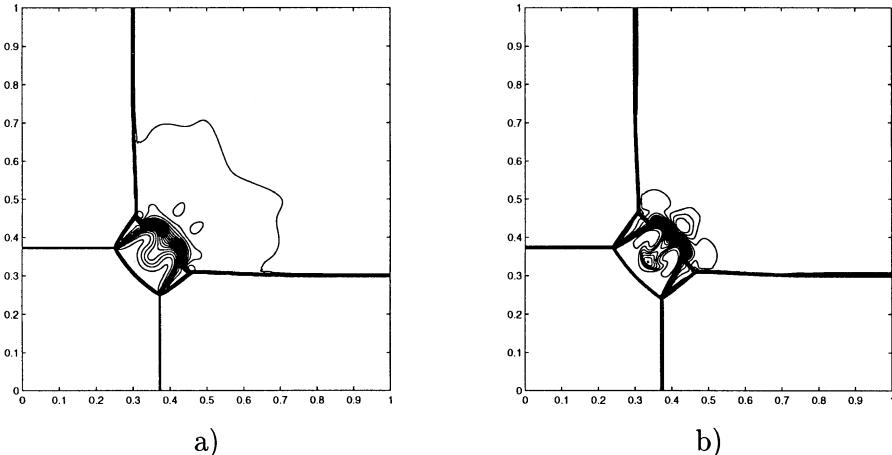


Figure 9. Configuration 3. 400×400 grid, time=.3, cfl=.9, 30 contours: a) 2D HLLE. b) CFLF4

approximate Riemann solver HLLC described in (Toro, 1999), Ch. 10, and we call these schemes LFC and LWC, respectively.

With the notation as in section 2, the two step LF scheme obtains the mean values v_i^{n+1} from the v_i^n according to the equations

$$v_{i+\frac{1}{2}}^{n+\frac{1}{2}} = .5(v_i^n + v_{i+1}^n) - .5\lambda(f(v_{i+1}^n) - f(v_i^n)) \quad (26)$$

and

$$v_i^{n+1} = .5(v_{i+\frac{1}{2}}^{n+\frac{1}{2}} + v_{i-\frac{1}{2}}^{n+\frac{1}{2}}) - .5\lambda(f(v_{i+\frac{1}{2}}^{n+\frac{1}{2}}) - f(v_{i-\frac{1}{2}}^{n+\frac{1}{2}})). \quad (27)$$

Note that this is just HLLE using maximal speeds and two half steps.

Let q^{x*} be the x -component of velocity extracted from $v_{i+\frac{1}{2}}^{n+\frac{1}{2}}$. That is

$$q^{x*} = m_{i+\frac{1}{2}}^{n+\frac{1}{2}} / \rho_{i+\frac{1}{2}}^{n+\frac{1}{2}}. \quad (28)$$

Define left and right densities ρ_l and ρ_r according to

$$\rho_l(\Delta x + q^{x*}\Delta t) = \rho_i^n(\Delta x + (q^x)_i^n\Delta t), \quad (29)$$

$$\rho_r(\Delta x - q^{x*}\Delta t) = \rho_{i+1}^n(\Delta x - (q^x)_{i+1}^n\Delta t). \quad (30)$$

This just amounts to assuming a contact discontinuity moving with speed q^{x*} as in Figure 10 and then splitting the mass flux into a left and right part. In the same way, the simply advected y -component of velocity splits according to

$$\begin{aligned} \rho_l q_l^y q^{x*} \Delta t + \Delta x &= \rho_i^n (q^y)_i^n (\Delta x + (q^x)_i^n \Delta t), \\ \rho_r q_r^y (\Delta x - q^{x*} \Delta t) &= \rho_{i+1}^n (q^y)_{i+1}^n (\Delta x - (q^x)_{i+1}^n \Delta t). \end{aligned}$$

But then it follows from (30) that

$$q_l^y = (q^y)_i^n \quad \text{and} \quad q_r^y = (q^y)_{i+1}^n.$$

The pressure p^* will also be extracted from $v_{i+\frac{1}{2}}^{n+\frac{1}{2}}$. However, special care has to be taken with the dependence of energy and pressure on the velocity component orthogonal to the direction of integration. This velocity is just advected in that direction with the fluid, but if the kinetic energy is computed using the average slip velocity then the pressure will be incorrect. This means we can't extract the pressure from the average energy using the kinetic energy of the average velocity field. Instead, we propose

$$\begin{aligned} p^* &= (\gamma - 1)(E_{i+\frac{1}{2}}^{n+\frac{1}{2}} - .5\rho_{i+\frac{1}{2}}^{n+\frac{1}{2}}(q^{x*})^2 - .25((1. + q^{x*}\lambda)\rho_l(q_l^y)^2 \\ &\quad + (1. - q^{x*}\lambda)\rho_r(q_r^y)^2)). \end{aligned} \quad (31)$$

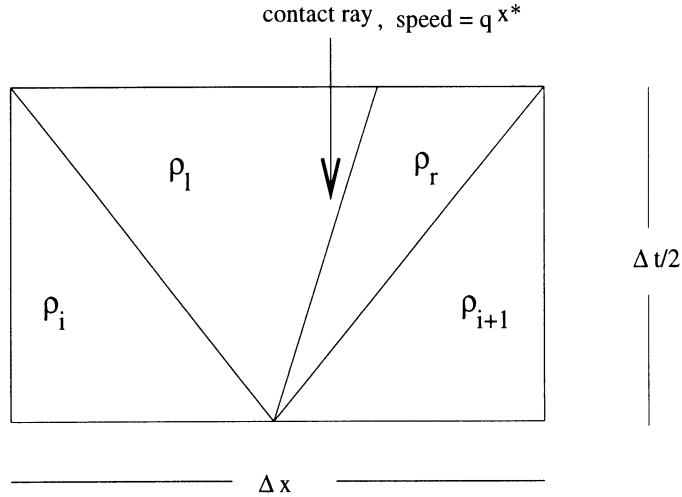


Figure 10. Contact ray and density distribution for LFC

Now replace (26) by

$$v_{i+\frac{1}{2}}^{n+\frac{1}{2}}(x) = \begin{cases} v_l, & x_i \leq x \leq x_{i+\frac{1}{2}} + q^{x*}\Delta t/2 \\ v_r, & x_{i+\frac{1}{2}} + q^{x*}\Delta t/2 \leq x \leq x_{i+1} \end{cases}, \quad (32)$$

where

$$\begin{aligned} v_l &= \begin{pmatrix} \rho_l \\ \rho_l q_l^{x*} \\ \rho_l q_l^y \\ \frac{p^*}{(\gamma-1)} + .5\rho_l[(q^{x*})^2 + (q_l^y)^2] \end{pmatrix}, \\ v_r &= \begin{pmatrix} \rho_r \\ \rho_r q_r^{x*} \\ \rho_r q_r^y \\ \frac{p^*}{(\gamma-1)} + .5\rho_r[(q^{x*})^2 + (q_r^y)^2] \end{pmatrix}. \end{aligned} \quad (33)$$

Then replace (27) by

$$\begin{aligned} v_i^{n+1} &= \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_i} v_{i-\frac{1}{2}}^{n+\frac{1}{2}}(x) dx + \frac{1}{\Delta x} \int_{x_i}^{x_{i+\frac{1}{2}}} v_{i+\frac{1}{2}}^{n+\frac{1}{2}}(x) dx \\ &\quad - \frac{1}{2} \lambda (f(v_{i+\frac{1}{2}}^{n+\frac{1}{2}}(x_{i+\frac{1}{2}})) - f(v_{i-\frac{1}{2}}^{n+\frac{1}{2}}(x_{i-\frac{1}{2}}))). \end{aligned} \quad (34)$$

Then (32) and (34) define the scheme LFC. For LWC we simply set

$$v_i^{n+1} = v_i^n - \lambda (f(v_{i+\frac{1}{2}}^{n+\frac{1}{2}}(x_{i+\frac{1}{2}})) - f(v_{i-\frac{1}{2}}^{n+\frac{1}{2}}(x_{i-\frac{1}{2}}))). \quad (35)$$

5.1. 1D NUMERICAL RESULTS

We have computed Toro's six test problems from Chapter 10 of (Toro, 1999) in Figures 11-13. The data are given below in the notation of (2). Recall that $q \equiv q^x$ is velocity, and here $q^y = 0$. For all problems, $\gamma = 1.4$.

Test 1:

$$\begin{aligned} p_0 &= 1. & \rho_0 &= 1. & q_0 &= .75 \\ p_1 &= .1 & \rho_1 &= .125 & q_1 &= 0. \end{aligned}$$

Test 2:

$$\begin{aligned} p_0 &= .4 & \rho_0 &= 1. & q_0 &= -2. \\ p_1 &= .4 & \rho_1 &= 1. & q_1 &= 2. \end{aligned}$$

Test 3:

$$\begin{aligned} p_0 &= 1000. & \rho_0 &= 1. & q_0 &= 0. \\ p_1 &= .01 & \rho_1 &= 1. & q_1 &= 0. \end{aligned}$$

Test 4:

$$\begin{aligned} p_0 &= 460.894 & \rho_0 &= 5.99924 & q_0 &= 460.894 \\ p_1 &= 46.095 & \rho_1 &= 5.99242 & q_1 &= 46.095 \end{aligned}$$

Test 5:

$$\begin{aligned} p_0 &= 1. & \rho_0 &= 1.4 & q_0 &= 0. \\ p_1 &= 1. & \rho_1 &= 1. & q_1 &= 0. \end{aligned}$$

Test 6:

$$\begin{aligned} p_0 &= 1. & \rho_0 &= 1.4 & q_0 &= .1 \\ p_1 &= 1. & \rho_1 &= 1. & q_1 &= .1 \end{aligned}$$

In all figures the solid curves are the exact solutions.

For Test 1 we compare LFC with HLLC in Figure 11a. Around the shock and the contact they behave in the same way, but the contact-corrected LF is less accurate than HLLC in the rarefaction wave. In Figure 11b we compare LW with LWC. LWC significantly reduces the oscillations of LW behind the contact, but has a larger overshoot at the shock.

For Test 2 LW and LWC fail. In Figure 12a we show the internal energy resulting from the scheme resulting from composing LFC and LWC, as described in (Liska and Wendroff, 1998). Call this LWCn. We do n-1 time

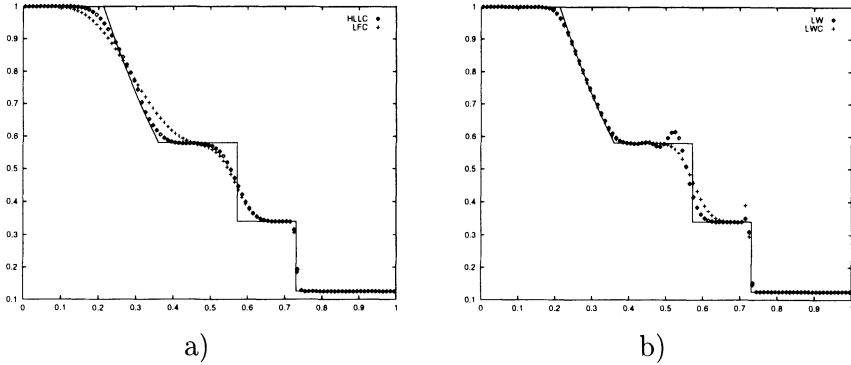


Figure 11. Test 1, density, time = .2. a) LFC and HLLC; b) LW and LWC.

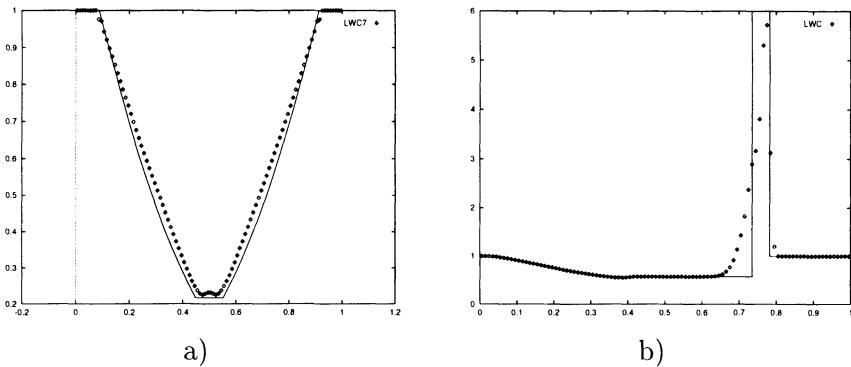


Figure 12. a) Internal energy, Test 2, LWC7, time = .15. b) Density: Test 3, LWC, time = .012

steps of LWC followed by one step of LFC. In this case $n=7$ gave the best result. This is a difficult problem, as there is a near vacuum around the center. Most methods computed in (Toro, 1999) have a far greater error in the internal energy than does our composite. However, this is not entirely satisfactory since there is a large variation in behavior around $n=7$.

For Test 3 we show the density for LWC.

LWC runs Test 4, but has severe oscillations. Instead, we show the result of LWC3 together with HLLC. Note that Toro points out that LW fails on Test 3 and Test 4.

For Tests 5 and 6, LFC and LWC produce identical results and these are identical to those obtained from HLLC in Figure 10.9 of (Toro, 1999). We show Test 5 only.

There are some other test problems that show that being exact for stationary contact discontinuities can be a mixed blessing. For the following

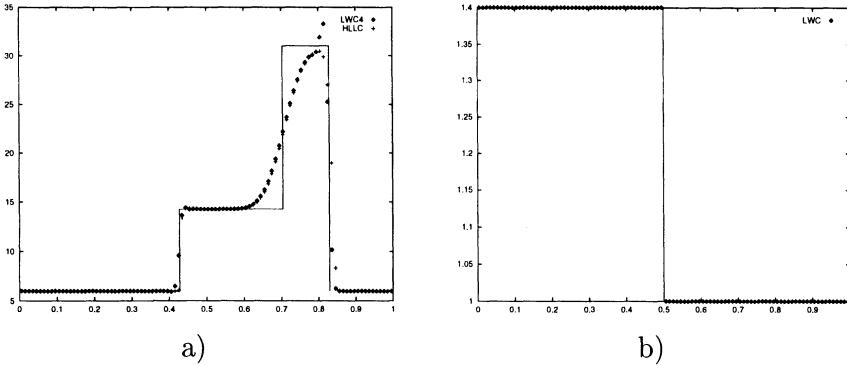


Figure 13. Density: Test 4, LWC4 and HLLC, time = .035: b) Density, Test 5, LWC, time = .2

data there is a shock with the fluid at rest behind the shock.

Simple shock:

$$\begin{aligned} p_0 &= .029 & \rho_0 &= .138 & q_0 &= 1.206 \\ p_1 &= .3 & \rho_1 &= .5323 & q_1 &= 0. \end{aligned}$$

Figure 14a shows a significant error near the position of the initial discontinuity at time $t = .5$. It seems here that there is no mechanism to dissipate the error left near the stationary trivial contact discontinuity, although HLLC works better on this problem than LWC. This phenomenon seems to be well known (E.F. Toro, private communication). See, for example, (Wada and Liou, 1997), (Roberts, 1990), and (Obayashi and Wada, 1994).

Part of the difficulty here is that these difference schemes are not Galilean invariant. For all of these Riemann problems adding a constant to the initial velocities does not change the exact final thermodynamic states, but it definitely affects the numerical results. Toro has suggested the following Riemann problem which has a stationary contact.

Toro's high flow velocity problem:

$$\begin{aligned} p_0 &= 30. \times 10^6 & \rho_0 &= 2. & q_0 &= -3019.5071 \\ p_1 &= 10^5 & \rho_1 &= 1. & q_1 &= -3019.5071 \end{aligned}$$

This is apparently difficult for some methods, but LWC4 does quite well with it, see Figure 14b.

6. Contact Correction in 2D

This is the third part of this work. The previous section shows that there may be sufficient improvement when contact discontinuities are modeled

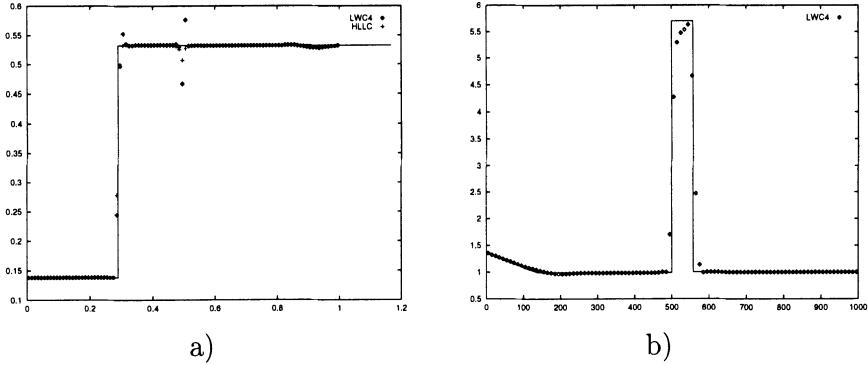


Figure 14. Density: a) Simple shock showing error for LWC4 and HLLC. b) High flow velocity stationary contact problem

in HLL-type schemes that it might be worthwhile to try to do this in two dimensions.

6.1. DIMENSIONALLY SPLIT SCHEME USING LFC AND LWC

As an easy preliminary to a full development, and to have a basis for comparison, we present the result of using a dimensionally split 2D version of the contact-corrected composite LWC_n. The composition is of LWC_xLWC_y applied $n - 1$ times followed by LFC_xLFC_y.

We present a computation of Configuration 15a which is a modification of Configuration 15 of (Schulz-Rinne et al, 1993)). The data are

$$\begin{aligned} p_{00} &= .3 & \rho_{00} &= .8 & q_{00}^x &= .0 & q_{00}^y &= 0. \\ p_{01} &= .4 & \rho_{01} &= .5197 & q_{01}^x &= -.7259 & q_{01}^y &= 0. \\ p_{11} &= 1. & \rho_{11} &= 1. & q_{11}^x &= .0 & q_{11}^y &= 0. \\ p_{10} &= .4 & \rho_{10} &= .5313 & q_{10}^x &= 0. & q_{10}^y &= .7276. \end{aligned}$$

This is arranged so that there are stationary contact discontinuities, with slip, at the bottom and left edges, a shock at the right edge, and a rarefaction wave at the top. Figure 15 compares the split composite with CFLF4. The composite exactly resolves the initial contact discontinuities, but there is noise (about a 6% error) left in the stationary flow behind the shock at the right, just as there was in Figure 14a.

For the original Configuration 15 of (Schulz-Rinne et al, 1993)) the data are

$$\begin{aligned} p_{00} &= .3 & \rho_{00} &= .8 & q_{00}^x &= .1 & q_{00}^y &= -.3 \\ p_{01} &= .4 & \rho_{01} &= .5197 & q_{01}^x &= -.6259 & q_{01}^y &= -.3 \end{aligned}$$

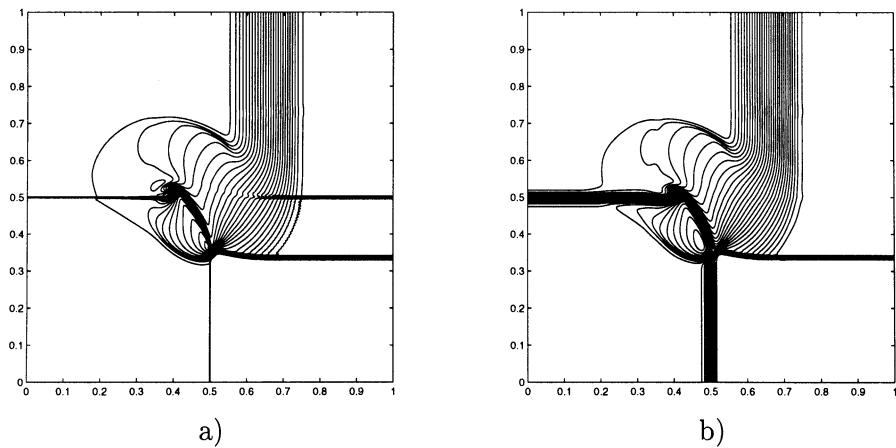


Figure 15. Configuration 15a. 400×400 grid, time=.2, cfl=.9, 30 density contours: a) Split composite. b) CFLF4

$$\begin{array}{llll} p_{11} = 1. & \rho_{11} = 1. & q_{11}^x = .1 & q_{11}^y = -.3 \\ p_{10} = .4 & \rho_{10} = .5313 & q_{10}^x = 1 & q_{10}^y = .4276. \end{array}$$

This differs from 15a only in that the contact discontinuities are not stationary. The result is shown in Figure 16a for the split composite, and in Figure 16b for CFLF4. The composite shows slightly better resolution than CFLF4. The understandably poor performance of 2DHLL on this problem is shown in Figure 17.

6.2. 2D CONTACT-CORRECTED LAX-FRIEDRICH

Just as in 1D, a 2D version of Lax-Friedrichs can be obtained as a special case of HLLE using maximal speeds and two half steps. The LF scheme in (Liska and Wendroff, 1998) is the result; it has one state instead of nine. In spite of the fact that there are difficulties with the notion of contact correction we have attempted to create a 2D version of LFC. If that were successful it would be fairly easy to extend it to a 2D HLLC, since whatever procedure is used in the dual cell for LFC could be used in the central state of HLLE, while the 1D edge states would just be split in two by the 1D contact discontinuities. We have been able to construct a four-state 2D LFC, but one that is not entirely satisfactory. For this reason, and also because it would involve 16 states and be rather inefficient, we have not at this time followed up with a 2D HLLC.

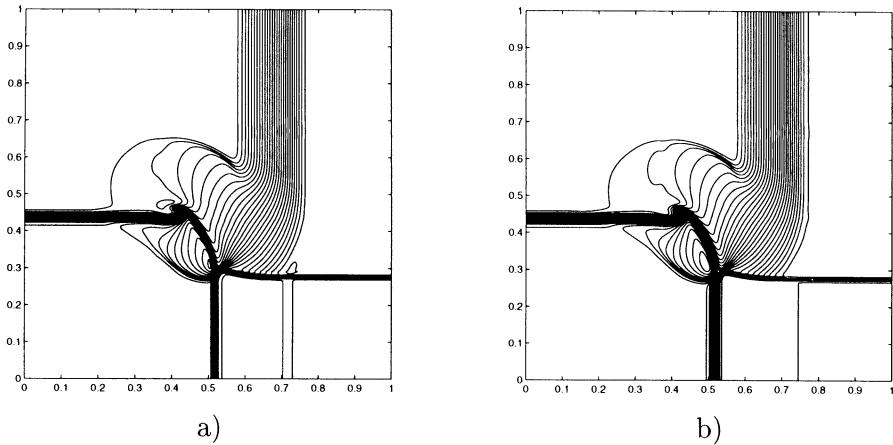


Figure 16. Configuration 15. 400×400 grid, time=.2, cfl=.9, 30 density contours: a) Split composite. b) CFLF4

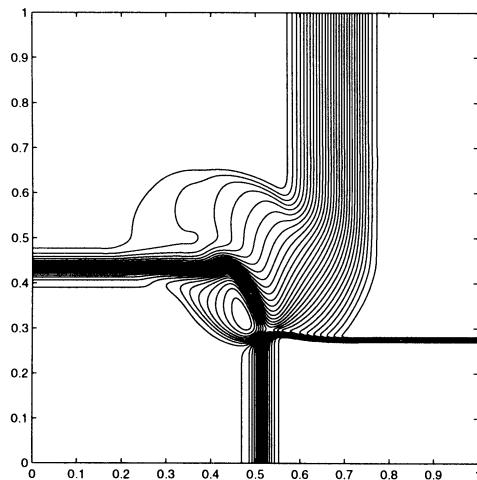


Figure 17. Configuration 15. 400×400 grid, time=.2, cfl=.9, 30 density contours: 2DHLL

The 2D LFC will be a two-step process, as in 1D. We give a brief partial description of the first half-step.

We require of a 2D LFC two properties: It should reduce to 1D LFC if the data is 1D, and it should be as simple as possible. The latter requirement means that we should not have to compute intersections of arbitrary

quadrilaterals. We have arrived at the following procedure to accomplish this. The idea is to create two internal (to the dual cell) space time surfaces that will be treated as moving grid surfaces, but not necessarily as contact surfaces. These will divide the dual cell into four rectangular sub-cells, in analogy with the two intervals in LFC. The surfaces are defined by a velocity (U, V) , to be chosen. Let the dual cell center be at $(0,0)$. One surface is obtained by translating the line $x = 0$ horizontally with speed U , the other by translating the line $y = 0$ vertically with speed V . The lines $x = U\Delta t/2$, $y = V\Delta t/2$ together with the boundaries of the dual cell form the four regions of the dual cell at the half step. Each region will have a different density, but there will be one horizontal velocity in each horizontal strip, one vertical velocity in each vertical strip. Then each quadrant will have a density, and two momenta. Total mass and momenta in the dual cell are conserved. The pressure is obtained from the total energy and the kinetic energy in each quadrant.

First, let us define U and V . Refer to Figure3. Let U_N , respectively U_S , be the mean LF horizontal velocity connecting the states with indices (01) and (11) , respectively (00) and (10) . N and S refer to the north (upper) and south (lower) portions of the dual cell, at time $n + \frac{1}{2}$. Similarly, also at $n + \frac{1}{2}$, let V_E , respectively V_W , be the mean LF vertical velocity connecting the states with indices (10) and (11) , respectively (00) and (01) . E and W refer to the east (right) and west (left) portions of the dual cell. Then set

$$U \equiv \frac{1}{2}(U_N + U_S) \quad V \equiv \frac{1}{2}(V_E + V_W).$$

Another possibility is to take U and V to be the 2D LF velocity components. In either case, if the data is independent of y then $U_N = U_S = U$, $V_E = q_{11}^y = q_{10}^y$, and $V_W = q_{01}^y = q_{00}^y$. An analogous statement holds if the data is independent of x .

We apply the integral conservation law for density in each quadrant, taking into account the moving internal surface. For example, for the northeast quadrant, after dividing by $\frac{1}{4}\Delta x\Delta y$, we have an equation of the form

$$(1 - \lambda U)(1 - \mu V)\rho_{NE} = \rho_{11} + \lambda((U_N - U)\rho_N - F_{NE}^1) + \mu((V_E - V)\rho_E - G_{NE}^1), \quad (36)$$

where F_{NE}^1 and G_{NE}^1 are the (normalized) integrated exterior partial boundary mass fluxes, and ρ_N and ρ_E are approximate integrated densities at the internal surfaces. ($\lambda = \Delta x/\Delta t$, $\mu = \Delta y/\Delta t$.) In order to maintain conservation the exact form of the fluxes depends on the relative sizes of U , U_N , U_S , and of V , V_E , and V_W . Suppose for example that $U \geq U_N$, $U \geq U_S$,

and $V \geq V_E$. Then

$$G_{NE}^1 = (\rho q^y)_{11} \frac{1 - \frac{1}{2}\lambda q_{11}^x}{1 - \frac{1}{2}\lambda U_N} (1 - \frac{1}{2}\lambda U),$$

$$\rho_E = \frac{1}{2} [\rho_{11} \frac{1 - \frac{1}{2}\lambda q_{11}^x}{1 - \frac{1}{2}\lambda U_N} (1 - \frac{1}{2}\lambda U) + \rho_{10} \frac{1 - \frac{1}{2}\lambda q_{10}^x}{1 - \frac{1}{2}\lambda U_S} (1 - \frac{1}{2}\lambda U)],$$

and

$$F_{NE}^1 = (\rho q^x)_{11} \frac{1 - \frac{1}{2}\mu q_{11}^y}{1 - \frac{1}{2}\mu V_E} (1 - \frac{1}{2}\mu V).$$

Suppose the data are independent of y . Then $V_E \rho_E = G_{NE}^1$ and

$$(1 - \lambda U)(1 - \mu V)\rho_{NE} = \rho_{11} - \lambda F_{NE}^1 - \mu V \rho_E$$

$$= \rho_{11} [1 - \lambda(1 - \frac{1}{2}\mu V)q_{11}^x - \mu V(1 - \frac{1}{2}\lambda q_{11}^x)]$$

$$= \rho_{11}(1 - \lambda q_{11}^x)(1 - \mu V)$$

which is the desired result.

The equation for the horizontal momentum in the north strip has the form

$$(1 - \mu V)(\bar{\rho} \bar{q}^x)_N = \frac{1}{2}(\rho q^x)_{01} + (\rho q^x)_{11} + \frac{1}{2}\lambda(F_{NW}^2 - F_{NE}^2)$$

$$+ \mu(\frac{1}{2}[(V_E - V)(\rho U)_E$$

$$+ (V_W - V)(\rho U)_W]) - G_N^2 \quad (37)$$

where F^2 and G^2 are the integrated x-component momentum fluxes. Again assuming $U \geq U_N$, $U \geq U_S$, $V \geq V_E$, and $V \geq V_W$ we have

$$F_{NE}^2 = [(\rho(q^x)^2)_{11} \frac{1 - \frac{1}{2}\mu q_{11}^y}{1 - \frac{1}{2}\mu V_E} + p_E^*](1 - \frac{1}{2}\mu V),$$

$$F_{NW}^2 = [(\rho(q^x)^2)_{01} \frac{1 - \frac{1}{2}\mu q_{01}^y}{1 - \frac{1}{2}\mu V_W} + p_W^*](1 - \frac{1}{2}\mu V),$$

$$G_N^2 = U_N(\rho q^y)_N,$$

where p_E^* , p_W^* are average pressures for the east and west pairs of states, $(\rho q^y)_N$ is the LF average y -momentum for the north pair of states at $n + \frac{1}{2}$ and

$$\begin{aligned}
(\rho U)_E &= \frac{1}{2} [U_N \rho_{11} \frac{1 - \frac{1}{2}\lambda q_{11}^x}{1 - \frac{1}{2}\lambda U_N} (1 - \frac{1}{2}\lambda U) \\
&\quad + U_S \rho_{10} \frac{1 - \frac{1}{2}\lambda q_{10}^x}{1 - \frac{1}{2}\lambda U_S} (1 - \frac{1}{2}\lambda U)], \\
(\rho U)_W &= \frac{1}{2} [U_N (\rho_{01} (1 + \frac{1}{2}\lambda q_{01}^x) + \frac{1}{2}\lambda (U - U_N) \rho_{11} \frac{1 - \frac{1}{2}\lambda q_{11}^x}{1 - \frac{1}{2}\lambda U_N}) \\
&\quad + U_S (\rho_{00} (1 + \frac{1}{2}\lambda q_{00}^x) + \frac{1}{2}\lambda (U - U_S) \rho_{10} \frac{1 - \frac{1}{2}\lambda q_{10}^x}{1 - \frac{1}{2}\lambda U_N})].
\end{aligned}$$

The horizontal velocity in the north strip is then

$$\bar{q}_N^x = \frac{(\bar{\rho} \bar{q}^x)_N}{\frac{1}{2}[(1 - \lambda U)\rho_{NE} + (1 + \lambda U)\rho_{NW}]}.$$

The complete description of all the quantities in all cases is too lengthy to include here.

For the second half step the masses and momenta are projected onto the primary cells. A single pressure, p^* , in the dual cell is obtained by subtracting the kinetic energy \mathcal{K} from the total energy of the cell, where, in analogy with 1D LFC,

$$\begin{aligned}
\mathcal{K} &= .125 [(1 - \mu V)(1 - \lambda U)\rho_{NE}((\bar{q}_N^x)^2 + (\bar{q}_E^y)^2) \\
&\quad + (1 + \mu V)(1 - \lambda U)\rho_{SE}((\bar{q}_S^x)^2 + (\bar{q}_E^y)^2) \\
&\quad + (1 - \mu V)(1 + \lambda U)\rho_{NW}((\bar{q}_N^x)^2 + (\bar{q}_W^y)^2) \\
&\quad + (1 + \mu V)(1 + \lambda U)\rho_{SW}((\bar{q}_S^x)^2 + (\bar{q}_W^y)^2)]
\end{aligned}$$

The fluxes at the center of the dual cell are obtained from the quantities in whichever quadrant that point lies. Finally, the projection of the energy is $p^*/(\gamma - 1)$ plus half [the projection of ρq^x]² plus half [the projection of ρq^y]² all divided by the projection of ρ .

We show the result of computing with this 2D LFC on the two examples, configurations 15a and 15, compared with 2D Lax-Friedrichs, in Figures 18 and 19. There are clearly false signals propagating from the center out along the 1D contact discontinuities. We did a test, not shown, with 1D data and a small perturbation at the center, and similar signals were seen.

7. Conclusion

We have presented a truly two-dimensional version of Godunov's method using an approximate 2D Riemann solver inspired by (Harten et al, 1983).

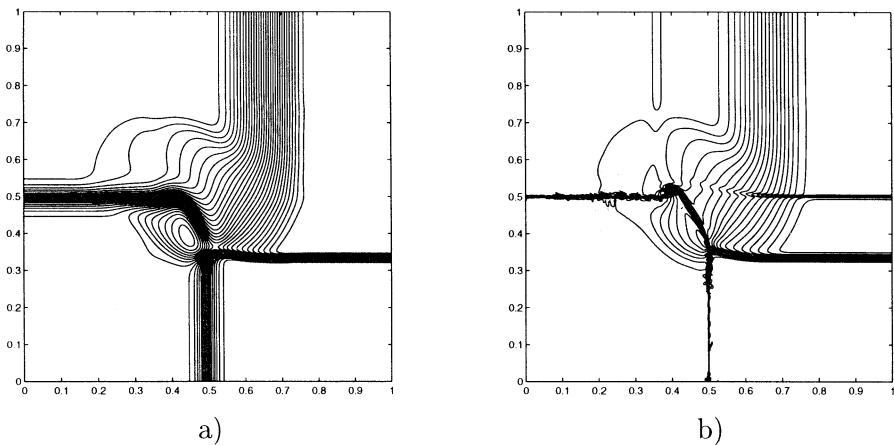


Figure 18. Configuration 15a. 400×400 grid, time=.2, cfl=.9, 30 density contours: a) LF. b) 2DLFC

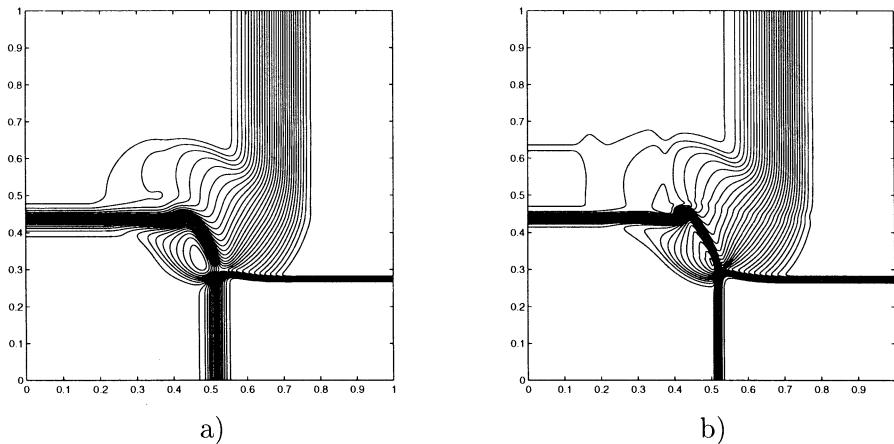


Figure 19. Configuration 15. 400×400 grid, time=.2, cfl=.9, 30 density contours: a) LF. b) 2DLFC

In its present form it is not really competitive with other methods for 2D compressible flow, but we hope it will stimulate research in this area. We have also offered 1D and 2D versions of the Lax-Friedrichs scheme designed to improve the resolution of contact discontinuities, with mixed results, but also with the hope of suggesting further research.

Acknowledgement

Work supported by the U. W. Department of Energy, under contract W-7405-ENG-36 and the CHAMMP program.

Appendix A - 2D Einfeldt Speeds

The 1D Roe averages and speeds for the 2D equations can be found in (Toro, 1999) along with original references. We present them here for completeness.

Let

$$\begin{aligned} H_{i,j} &= \frac{1}{\rho_{i,j}}(E_{i,j} + p_{i,j}), \\ c_{i,j} &= \sqrt{(\gamma - 1)(H_{i,j} - .5((q_{i,j}^x)^2 + (q_{i,j}^y)^2))}. \end{aligned}$$

Then

$$\begin{aligned} \bar{q}_{i+\frac{1}{2},j}^x &= \frac{\sqrt{\rho_{i,j}}q_{i,j}^x + \sqrt{\rho_{i+1,j}}q_{i+1,j}^x}{\sqrt{\rho_{i,j}} + \sqrt{\rho_{i+1,j}}}, \\ \bar{q}_{i+\frac{1}{2},j}^y &= \frac{\sqrt{\rho_{i,j}}q_{i,j}^y + \sqrt{\rho_{i+1,j}}q_{i+1,j}^y}{\sqrt{\rho_{i,j}} + \sqrt{\rho_{i+1,j}}}, \\ \bar{H}_{i+\frac{1}{2},j} &= \frac{\sqrt{\rho_{i,j}}H_{i,j} + \sqrt{\rho_{i+1,j}}H_{i+1,j}}{\sqrt{\rho_{i,j}} + \sqrt{\rho_{i+1,j}}}, \\ \bar{c}_{i+\frac{1}{2},j} &= \sqrt{(\gamma - 1)(\bar{H}_{i+\frac{1}{2},j} - .5(((\bar{q}_{i+\frac{1}{2},j}^x)^2 + (\bar{q}_{i+\frac{1}{2},j}^y)^2))).} \end{aligned}$$

Then define the speeds:

$$\begin{aligned} b_{i+\frac{1}{2},j}^1 &= \max(\bar{q}_{i+\frac{1}{2},j}^x + \bar{c}_{i+\frac{1}{2},j}, q_{i+1,j}^x + c_{i+1,j}) \\ b_{i+\frac{1}{2},j}^0 &= \min(\bar{q}_{i+\frac{1}{2},j}^x - \bar{c}_{i+\frac{1}{2},j}, q_{i,j}^x - c_{i,j}). \end{aligned}$$

The speeds $b_{i,j+\frac{1}{2}}^1, b_{i,j+\frac{1}{2}}^0$ are obtained by interchanging the indices in all the doubly indexed quantities above, then interchanging i and j and then interchanging q^x and q^y .

Appendix B - Flux Form

We present here the flux form for method 3,

$$v_{i,j}^{n+1} = v_{i,j}^n - \frac{\Delta t}{\Delta x}(\mathcal{F}_{i+\frac{1}{2},j} - \mathcal{F}_{i-\frac{1}{2},j}) - \frac{\Delta t}{\Delta y}(\mathcal{G}_{i,j+\frac{1}{2}} - \mathcal{G}_{i,j-\frac{1}{2}})$$

without proof. This is a little easier to do in terms of lengths and areas rather than speeds. So, let

$$L_{i+\frac{1}{2},j}^\bullet = |b_{i+\frac{1}{2},j}^\bullet| \Delta t, \quad L_{i,j+\frac{1}{2}}^\bullet = |b_{i,j+\frac{1}{2}}^\bullet| \Delta t, \quad L_{i+\frac{1}{2},j+\frac{1}{2}}^\bullet = |d_{i+\frac{1}{2},j+\frac{1}{2}}^\bullet| \Delta t,$$

and

$$\begin{aligned} A_{i+\frac{1}{2},j+\frac{1}{2}}^{00} &= L_{i+\frac{1}{2},j+\frac{1}{2}}^{x-} L_{i+\frac{1}{2},j+\frac{1}{2}}^{y-}, & A_{i+\frac{1}{2},j+\frac{1}{2}}^{01} &= L_{i+\frac{1}{2},j+\frac{1}{2}}^{x-} L_{i+\frac{1}{2},j+\frac{1}{2}}^{y+}, \\ A_{i+\frac{1}{2},j+\frac{1}{2}}^{11} &= L_{i+\frac{1}{2},j+\frac{1}{2}}^{x+} L_{i+\frac{1}{2},j+\frac{1}{2}}^{y+}, & A_{i+\frac{1}{2},j+\frac{1}{2}}^{10} &= L_{i+\frac{1}{2},j+\frac{1}{2}}^{x+} L_{i+\frac{1}{2},j+\frac{1}{2}}^{y-}. \end{aligned}$$

The area of the central rectangle of the dual cell at $(i + \frac{1}{2}, j + \frac{1}{2})$ is

$$A_{i+\frac{1}{2},j+\frac{1}{2}} = A_{i+\frac{1}{2},j+\frac{1}{2}}^{00} + A_{i+\frac{1}{2},j+\frac{1}{2}}^{01} + A_{i+\frac{1}{2},j+\frac{1}{2}}^{11} + A_{i+\frac{1}{2},j+\frac{1}{2}}^{10}$$

Also

$$A_{i+\frac{1}{2},j+\frac{1}{2}}^{00} A_{i+\frac{1}{2},j+\frac{1}{2}}^{11} = A_{i+\frac{1}{2},j+\frac{1}{2}}^{01} A_{i+\frac{1}{2},j+\frac{1}{2}}^{10} \quad (38)$$

Let us first define cell fluxes to put into

$$v_{i,j}^{n+1} = v_{i,j}^n - \frac{\Delta t}{\Delta x} (\mathcal{F}_{i,j}^r - \mathcal{F}_{i,j}^l) - \frac{\Delta t}{\Delta y} (\mathcal{G}_{i,j}^r - \mathcal{G}_{i,j}^l).$$

Then we will just verify that

$$\mathcal{F}_{i,j}^r = \mathcal{F}_{i+1,j}^l, \quad \mathcal{G}_{i,j}^r = \mathcal{G}_{i,j+1}^l,$$

so that

$$\begin{aligned} \mathcal{F}_{i+\frac{1}{2},j} &= \mathcal{F}_{i,j}^r, & \mathcal{F}_{i-\frac{1}{2},j} &= \mathcal{F}_{i,j}^l, \\ \mathcal{G}_{i,j+\frac{1}{2}} &= \mathcal{G}_{i,j}^r, & \mathcal{G}_{i,j-\frac{1}{2}} &= \mathcal{G}_{i,j}^l. \end{aligned}$$

Consider $\Delta y \Delta t \mathcal{F}_{i,j}^r$, which can be split into three parts:

$$\Delta y \Delta t \mathcal{F}_{i,j}^r = Z_1^r + Z_2^r + Z_3^r, \quad (39)$$

and correspondingly,

$$\Delta y \Delta t \mathcal{F}_{i,j}^l = Z_1^l + Z_2^l + Z_3^l. \quad (40)$$

Z_1^r is the flux at $(i + \frac{1}{2}, j)$ coming from the normal Riemann problem at this edge. There is a contribution to it from the dual cells at $(i + \frac{1}{2}, j + \frac{1}{2})$ and at $(i + \frac{1}{2}, j - \frac{1}{2})$. It is

$$\begin{aligned} Z_1^r &= \frac{\Delta y - L_{i+\frac{1}{2},j+\frac{1}{2}}^{y-} - L_{i+\frac{1}{2},j-\frac{1}{2}}^{y+}}{L_{i+\frac{1}{2},j}^{x+} + L_{i+\frac{1}{2},j}^{x-}} [L_{i+\frac{1}{2},j}^{x+} L_{i+\frac{1}{2},j}^{x-} (v_{i,j} - v_{i+1,j}) \\ &\quad + \Delta t (L_{i+\frac{1}{2},j}^{x+} f_{i,j} + L_{i+\frac{1}{2},j}^{x-} f_{i+1,j})]. \end{aligned} \quad (41)$$

Then Z_1^l is just Z_1^r with i replaced by $i - 1$ and x^- interchanged with x^+ . These clearly have the needed left right symmetry.

Z_2^r comes from the same dual cells, but it contains the terms arising from the 2D intermediate dual states that are independent of f and g . It is

$$\begin{aligned} Z_2^r = & \frac{A_{i+\frac{1}{2},j+\frac{1}{2}}^{00}}{A_{i+\frac{1}{2},j+\frac{1}{2}}} \left[\frac{1}{2} A_{i+\frac{1}{2},j+\frac{1}{2}}^{11} (v_{i,j+1} - v_{i+1,j+1} + v_{i,j} - v_{i+1,j}) \right. \\ & \left. + A_{i+\frac{1}{2},j+\frac{1}{2}}^{10} (v_{i,j} - v_{i+1,j}) \right] \\ & + \frac{A_{i+\frac{1}{2},j-\frac{1}{2}}^{01}}{A_{i+\frac{1}{2},j-\frac{1}{2}}} \left[\frac{1}{2} A_{i+\frac{1}{2},j-\frac{1}{2}}^{10} (v_{i,j-1} - v_{i+1,j-1} + v_{i,j} - v_{i+1,j}) \right. \end{aligned} \quad (42)$$

$$\left. + A_{i+\frac{1}{2},j+\frac{1}{2}}^{11} (v_{i,j} - v_{i+1,j}) \right] \quad (43)$$

For Z_2^l replace i by $i - 1$, and in the first superscript of the A 's interchange 0 and 1. Then (38) provides the left right symmetry.

Z_3^l contains the remaining contributions from the two dual cells. First, let

$$\begin{aligned} \hat{F}_{0,\frac{1}{2}}^+ &= \frac{1}{2} (2d_{\frac{1}{2},\frac{1}{2}}^{y+} - b_{0,\frac{1}{2}}^+ \Delta t) \Delta t f(u_{01}) + \frac{1}{2} b_{0,\frac{1}{2}}^+ (\Delta t)^2 f(u_{0,\frac{1}{2}}), \\ \hat{F}_{0,\frac{1}{2}}^- &= \frac{1}{2} (2|d_{\frac{1}{2},\frac{1}{2}}^{y-}| + b_{0,\frac{1}{2}}^- \Delta t) \Delta t f(u_{00}) + \frac{1}{2} (-b_{0,\frac{1}{2}}^-) (\Delta t)^2 f(u_{0,\frac{1}{2}}), \\ \hat{F}_{0,\frac{1}{2}} &= \hat{F}_{0,\frac{1}{2}}^+ + \hat{F}_{0,\frac{1}{2}}^- \end{aligned}$$

and

$$\begin{aligned} \hat{G}_{\frac{1}{2},0}^+ &= \frac{1}{2} (2d_{\frac{1}{2},\frac{1}{2}}^{x+} - b_{\frac{1}{2},0}^+ \Delta t) \Delta t g(u_{10}) + \frac{1}{2} b_{\frac{1}{2},0}^+ (\Delta t)^2 g(u_{\frac{1}{2},0}), \\ \hat{G}_{\frac{1}{2},0}^- &= \frac{1}{2} (2|d_{\frac{1}{2},\frac{1}{2}}^{x-}| + b_{\frac{1}{2},0}^- \Delta t) \Delta t g(u_{00}) + \frac{1}{2} (-b_{\frac{1}{2},0}^-) (\Delta t)^2 g(u_{\frac{1}{2},0}), \\ \hat{G}_{\frac{1}{2},0} &= \hat{G}_{\frac{1}{2},0}^+ + \hat{G}_{\frac{1}{2},0}^-. \end{aligned}$$

Then in analogy with (9) it can be shown that

$$\begin{aligned} A_{\frac{1}{2},\frac{1}{2}} u_{\frac{1}{2},\frac{1}{2}} &= A_{\frac{1}{2},\frac{1}{2}}^{00} u_{00} + A_{\frac{1}{2},\frac{1}{2}}^{01} u_{01} + A_{\frac{1}{2},\frac{1}{2}}^{10} u_{10} + A_{\frac{1}{2},\frac{1}{2}}^{11} u_{11} \\ &+ [\hat{F}_{0,\frac{1}{2}} - \hat{F}_{1,\frac{1}{2}}] + [\hat{G}_{\frac{1}{2},0} - \hat{G}_{\frac{1}{2},1}] \end{aligned} \quad (44)$$

Then Z_3^r has a part from the dual cell at $(i + \frac{1}{2}, j + \frac{1}{2})$, namely,

$$Z_{3a}^r = \frac{|dy_{i+\frac{1}{2},j+\frac{1}{2}}^-|}{(|dy_{i+\frac{1}{2},j+\frac{1}{2}}^-| + dy_{i+\frac{1}{2},j+\frac{1}{2}}^+) A_{i+\frac{1}{2},j+\frac{1}{2}}} [(A_{i+\frac{1}{2},j+\frac{1}{2}}^{00} + A_{i+\frac{1}{2},j+\frac{1}{2}}^{01})$$

$$\begin{aligned}
& (\hat{F}_{i+1,j+\frac{1}{2}} - \hat{G}_{i+\frac{1}{2},j}^+ + \hat{G}_{i+\frac{1}{2},j+1}^+) \\
& + (A_{i+\frac{1}{2},j+\frac{1}{2}}^{10} + A_{i+\frac{1}{2},j+\frac{1}{2}}^{11})(\hat{F}_{i,j+\frac{1}{2}} - \hat{G}_{i+\frac{1}{2},j}^- + \hat{G}_{i+\frac{1}{2},j+1}^-)] \quad (45)
\end{aligned}$$

The part from the dual cell at $(i + \frac{1}{2}, j - \frac{1}{2})$, Z_3^r , is obtained from Z_{3a}^r by interchanging dy^- with dy^+ , and replacing j by $j - 1$. Then $Z_3^r = Z_{3a}^r + Z_{3b}^r$. To get Z_3^l interchange 0 with 1 in the first superscript of the A 's, interchange \hat{G}^+ with \hat{G}^- , and replace i by $i - 1$. Again, the necessary left right symmetry is present.

To define the \mathcal{G} fluxes, it is only necessary to make appropriate substitutions. First, in (39,40) replace \mathcal{F} by \mathcal{G} . In (41,43,45) interchange y with x (e.g., dy becomes dx , L^x becomes L^y), interchange f with g and F with G . Then interchange subscripts, and i with j (e.g. $v_{i,j+1}$ becomes $v_{i+1,j}$). Next, interchange superscripts in the A 's. These interchanges must also be made when obtaining the Z^l quantities.

References

- Batten P, Clarke N, Lambert C, and Clausen D M (1997). On the Choice of Wavespeeds for the HLLC Riemann Solver. *SIAM J. Sci. Comput.* **18**, pp 1553-1570.
- Einfeldt B (1988). On Godunov-type Methods for Gas Dynamics. *SIAM J. Numer. Anal.* **25**, pp 294-318.
- Einfeldt B, Munz C D, Roe P L and Sjøgreen B (1991). On Godunov-type Methods Near Low Densities. *J. Comp. Phys.* **92**, pp 273-295.
- Harten A, Lax P D, and van Leer B (1983). On Upstream Differencing and Godunov-type Schemes for Hyperbolic Conservation Laws. *SIAM Rev.* **25**, pp 35-61.
- Liska R and Wendroff B (1998). Composite Schemes for Conservation Laws. *SIAM J. Numer. Anal.* **35**, pp 2250-2271.
- Obayashi S and Wada Y (1994). Practical Formulation of a Positively Conservative Scheme. *AIAA J.* **32**, pp 1093-1095.
- Roberts T W (1990). The Behavior of Flux Difference Splitting Schemes Near Slowly Moving Shock Waves, *J. Comp. Phys.* **90**, pp 2250-2271.
- Schulz-Rinne C W, Collins J P and Glaz H M (1993). Numerical solution of the Riemann problem for two-dimensional gas dynamics, *SIAM J. Sci. Comput.* **14**, pp 1394-1414.
- Smoller J (1994). Shock waves and Reaction Diffusion Equations. Springer.
- Toro E F (1999). Riemann Solvers and Numerical Methods for Fluid Dynamics. Second Edition, Springer-Verlag.
- Wada Y and Liou M-S (1997). An Accurate and Robust Flux Splitting Scheme for Shock and Contact Discontinuities. *SIAM J. Sci. Comput.* **18**, pp 633-657.
- Wendroff B (1999). A Two-dimensional HLLE Riemann Solver and Associated Godunov-Type Difference Scheme for Gas Dynamics, *Computers Math Applic.* **38**, pp 175-185.

A UNIFIED METHOD FOR COMPRESSIBLE AND INCOMPRESSIBLE FLOWS WITH GENERAL EQUATION OF STATE

P. WESSELING AND D.R. VAN DER HEUL

J.M. Burgers Center and

Faculty of Information Technology and Systems

Delft University of Technology

Mekelweg 4, 2628 CD Delft, The Netherlands

Abstract. A unified method for computing incompressible and compressible flows is presented. The method is an extension of a staggered scheme with pressure correction for incompressible flows. The method is extended to fluids with arbitrary equations of state, and compared with the Osher scheme for Riemann problems. An application is shown to a hydrodynamic flow with cavitation, in which the Mach number varies between 10^{-3} in the water and 20 in the transition zone between water and vapor.

1. Introduction

To obtain a unified method for computing compressible and incompressible flows, one may extend methods for compressible flows to the low Mach number regime by preconditioning, see (Guillard and Viozat, 1999) and references quoted there. Alternatively, one may extend methods for incompressible flows to the compressible case. Departing from the incompressible staggered scheme of (Harlow and Welch, 1965), this is done in (Bijl and Wesseling, 1998), and references quoted there. The last approach is adopted here and extended to a nonconvex equation of state.

2. Extension of incompressible method to the compressible case

As $M \downarrow 0$, pressure variations become small compared to the ambient pressure, so that there is a danger of loss of significant figures if the customary units are employed. Choosing units \tilde{x}_r , \tilde{u}_r , \tilde{T}_r and $\tilde{\rho}_r$, and defining (assuming a perfect gas) $\tilde{p}_r = \tilde{\rho}_r R \tilde{T}_r$, the one-dimensional dimensionless inviscid

momentum equation takes the following form:

$$(\rho u)_t + (\rho uu)_x + \frac{1}{\varepsilon} p_x = 0 , \quad \varepsilon = \gamma M_r^2 ,$$

This equation becomes singular as $M_r \downarrow 0$. To avoid this, we introduce the following dimensionless pressure variable: $p = (\tilde{p} - \tilde{p}_r)/\tilde{\rho}_r \tilde{u}_r^2$. This makes the irksome factor $1/\varepsilon$ disappear. The dimensionless equation of state is given by $\rho = (1 + \varepsilon p)/T$. We will generalize the incompressible staggered scheme of (Harlow and Welch, 1965) to the compressible case. To obtain Mach-uniform efficiency, we use the pressure correction method to solve the discrete system. For this it is convenient to use the following non-conservative form of the dimensionless Euler equations:

$$\rho_t + m_x = 0 , \quad m = \rho u , \quad (1)$$

$$m_t + (m^2/\rho + p)_x = 0 , \quad (2)$$

$$M_r^2 \{ p_t + (up)_x + (\gamma - 1) p u_x \} + u_x = 0 . \quad (3)$$

As $M_r \downarrow 0$, (3) reduces to the familiar solenoidality constraint of incompressible flow. For $M_r = 0$ the pressure acts as a degree of freedom making it possible to satisfy this constraint. Therefore the constraint (i.e. the $\text{div } \mathbf{u}$ term in (3)) and the free parameter (p in (2)) have to be treated implicitly in time stepping methods. Pressure correction is an efficient solution method for the resulting implicit system.

Following the usual approach to get second order accuracy in gasdynamics, we use Runge-Kutta time stepping and a second order upwind-biased slope limited scheme in space. The grid is uniform and staggered, with the (ρ, p) nodes located at $x_j = (j - 1/2)h$ and the nodes for m at $x_{j+1/2} = jh$. The time step is denoted by τ . The $(k+1)^{\text{th}}$ Runge-Kutta stage is given by

$$\begin{aligned} \rho_j^{(k+1)} - \rho_j^n + \alpha_{k+1} \lambda \{ (u\rho)_{j+1/2} - (u\rho)_{j-1/2} \}^{(k)} &= 0 , \quad \lambda = \tau/h , \\ m_{j+1/2}^{(k+1)} - m_{j+1/2}^n + \alpha_{k+1} \lambda \{ (um)_{j+1}^{(k)} - (um)_j^{(k)} + p_{j+1}^n - p_j^n \} &= 0 , \end{aligned}$$

where $\rho_{j+1/2}^{(k)}$ and $(um)_j^{(k)}$ are approximated with a slope limited scheme, using the van Albada limiter (van Albada et al., 1982). The Runge-Kutta scheme is a four stage method proposed in (Sommeijer et al., 1994) with $\alpha_1 = 1/4$, $\alpha_2 = 1/3$, $\alpha_3 = 1/2$, $\alpha_4 = 1$. To save computing time, pressure correction is applied only after the Runge-Kutta step has been completed. We put

$$\rho_j^{n+1} = \rho_j^{(4)} , \quad m_{j+1/2}^{n+1} = m_{j+1/2}^{(4)} - \frac{1}{2} \lambda (\delta p_{j+1} - \delta p_j) , \quad \delta p = p^{n+1} - p^n , \quad (4)$$

where the factor $1/2$ is inserted because for accuracy we wish to approximate p_x in (2) at $t = t^n + \tau/2$. The pressure equation (3) is discretized as follows:

$$M_r^2 \{ \delta p_j + \lambda(u^{n+1} p^n)|_{j-1/2}^{j+1/2} + \lambda(\gamma - 1)p^n u^{n+1}|_{j-1/2}^{j+1/2} \} + \lambda u^{n+1}|_{j-1/2}^{j+1/2} = 0 , \quad (5)$$

where $p_{j+1/2}$ is approximated with the slope limited scheme. Substitution of $u_{j+1/2}^{n+1} = (m/\rho)_{j+1/2}^{n+1}$, with m^{n+1} given by (4) gives a linear system for δp .

The equivalence as $M_r \downarrow 0$ with the classical incompressible scheme of (Harlow and Welch, 1965) is obvious, but the scheme needs validation for compressible flows, especially because (3) is not in conservation form. Fig. 1 gives results for Sod's (1978) shocktube problem. For comparison, the solution obtained with the Osher scheme using the same slope-limited state interpolation and Runge-Kutta method is shown in Fig. 2. For this and

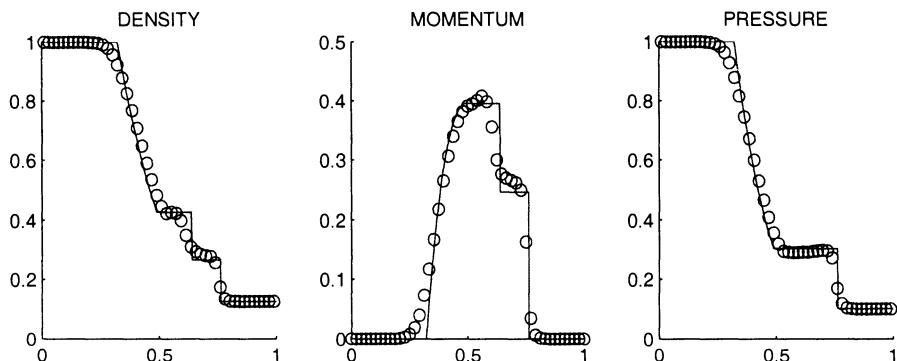


Figure 1. Exact and numerical solution of shocktube problem with staggered scheme; $\lambda = 0.45$, $t = 0.15$.

many other Riemann problems that we tried we found satisfactory agreement with the exact solution, with accuracy of the same quality as for established schemes (such as the Osher, Roe, AUSM and Jameson schemes). It seems that the staggered scheme converges to genuine weak solutions that satisfy the entropy condition. Note that only simple central and upwind differences are used. Hence, the numerical fluxes employed by the staggered scheme are much simpler to compute than those of the established schemes mentioned above. Accuracy and efficiency of the staggered scheme with the pressure correction method are found to be approximately Mach-uniform for two-dimensional applications in (Bijl and Wesseling, 1998).

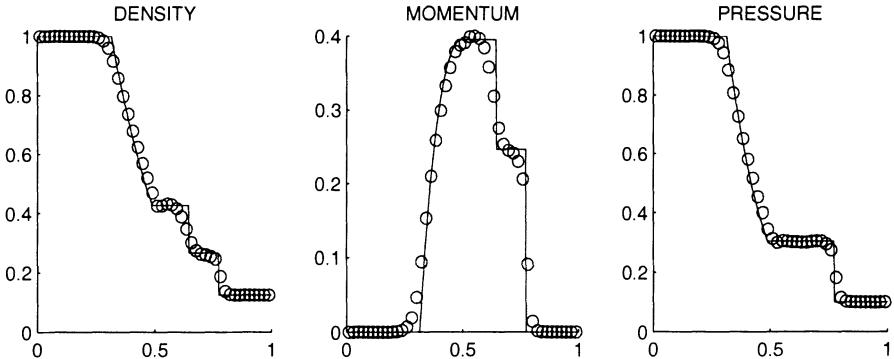


Figure 2. Solution of shocktube problem with Osher scheme.

3. Application to the barotropic Euler equations with a nonconvex equation of state

The governing equations are

$$\rho_t + m_x = 0, \quad m_t + (um + p)_x = 0, \quad p = p(\rho).$$

For nonperfect gases or fluids the equation of state is nonconvex, and/or is not available in simple analytic form, but given only in tables determined experimentally. Many established methods for computing compressible flows make explicit use of the perfect gas law in such a way, that their generalization to a general equation of state is far from straightforward, except for the Osher scheme.

On the staggered grid, we discretize as follows:

$$\rho_j^{n+1} - \rho_j^n + \lambda(m_{j+1/2} - m_{j-1/2})^{n+1} = 0, \quad (6)$$

$$m_{j+1/2}^{n+1} - m_{j+1/2}^n + \lambda\{(um)_{j+1} - (um)_j\}^n + \quad (7)$$

$$\lambda(p_{j+1} - p_j)^{n+1} = 0.$$

In order to introduce irreversibility so that (hopefully) solutions satisfy the entropy condition, in (7) the first order upwind scheme is used: $(um)_j = (um)_{j-1/2}$ (assuming $u > 0$). In one dimension, solving the implicit system (6), (7) takes little computing time, but in more dimensions this is expensive. To achieve Mach-uniform efficiency, we use the pressure correction method. A prediction m^* is made using the old pressure:

$$m_{j+1/2}^* - m_{j+1/2}^n + \lambda\{(um)_{j+1} - (um)_j\}^n + \lambda(p_{j+1} - p_j)^n = 0.$$

Next, we put $m_{j+1/2}^{n+1} = m_{j+1/2}^* - \lambda(\delta p_{j+1} - \delta p_j)$, $\delta p = p^{n+1} - p^n$. Substitution in (6) gives

$$\delta p_j + (\lambda c_j^n)^2(-\delta p_{j+1} + 2\delta p_j - \delta p_{j-1}) = -\lambda(m_{j+1/2}^* - m_{j-1/2}^*) ,$$

where the following approximation has been made:

$$\rho^{n+1} - \rho^n \cong \frac{1}{c^2} \delta p , \quad c^2 = \frac{dp}{d\rho} .$$

Results for a Riemann problem are shown in Fig. 4. The equation of the state is plotted as $p = p(V)$, $V = 1/\rho$ in Fig. 3. The equation of state

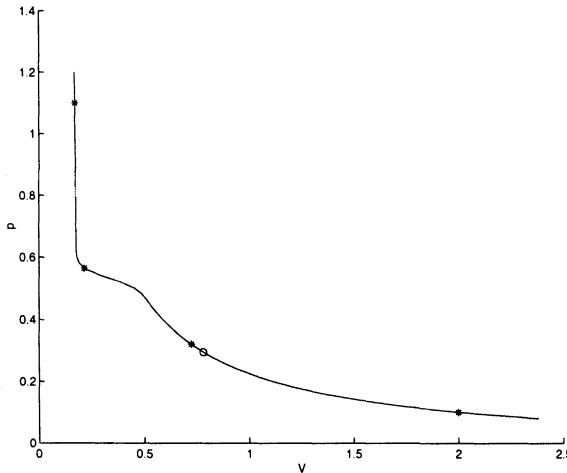


Figure 3. Graph of $p(V)$ with intermediate states in Riemann solution.

is nonconvex with two inflection points. The symbols give from right to left states 1 to 5. States 1 and 5 are the given right and left states. State 2, indicated by a circle, separates right and left running waves. According to Oleinik's (1957) entropy condition, right running waves must follow the upper concave hull of the $p - V$ diagram, whereas left running waves must follow the lower convex hull. Chords correspond to shocks. Accordingly, we have a compression shock between states 1 and 2, a fan between 2 and 3, an expansion shock between 3 and 4, and a fan between 4 and 5. Fig. 4 shows that the staggered scheme seems to converge to a genuine weak solution that satisfies the entropy condition. The very steep expansion fan between states 4 and 5 is smeared, as is to be expected for a first order scheme.

Because a staggered scheme for computing compressible flows is unorthodox, for comparison we have also used the Osher scheme. Since $p = p(\rho)$ is

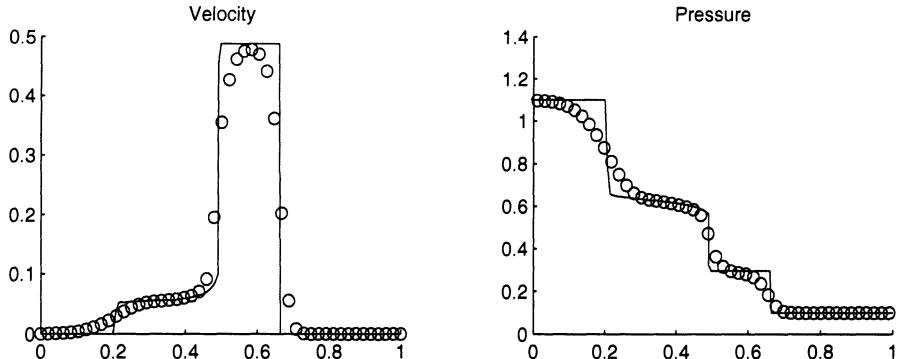


Figure 4. Solution of Riemann problem with a nonconvex equation of state with the staggered scheme. Solid line: exact solution; symbols: numerical solution with $\lambda = 0.4$, $h = 1/48$.

arbitrary, evaluation of the Osher flux is computing-intensive, so that the Osher scheme requires significantly more computing time than the staggered scheme. Results are shown in Fig. 5. The accuracy of both schemes is seen to be about the same.

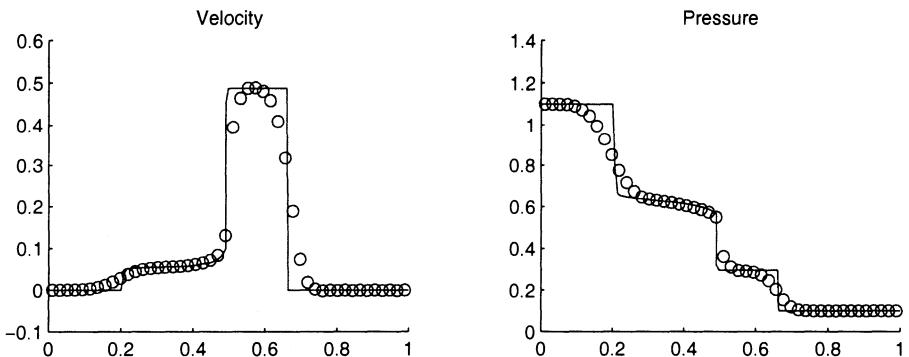


Figure 5. As Fig. 4, but with Osher scheme.

We briefly discuss application of the staggered scheme to flow with cavitation with the homogeneous equilibrium model. A hypothetical medium is adopted with properties of water above a certain pressure and of vapor below another pressure value, with a smooth transition in between. This approach is followed in (van der Heul and Wesseling, 1998), (Dellanoy and Kueny, 1990), (Hoeijmakers et al., 1998), (Merkle et al., 1998), (Song and He, 1998). A nonconvex equation of state results of the kind shown in Fig. 3. Extension of the staggered scheme to multidimensional curvilinear-

ear coordinates is described in (Wesseling et al., 1998), (Wesseling et al., 1999). A result for two-dimensional cavitating flow is shown in Fig. 6. This is a snapshot of isodensity contours. Low density regions corresponding

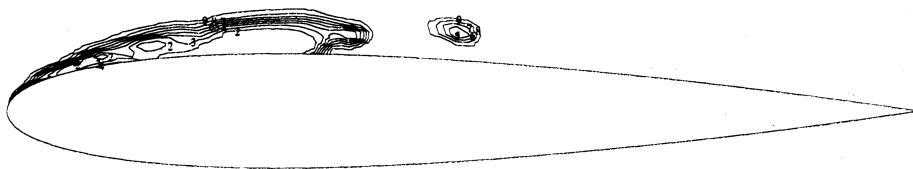


Figure 6. Iso-density contours for cavitating flow around a hydrofoil.

to cavitation bubbles are clearly delineated, quite similar to what is found in experiment. The equation of state gives a low speed of sound in the transitional regime between water and vapor. As a consequence, the maximum Mach number is found to be about 20. On the other hand, in the water phase, $M \approx 10^{-3}$. Clearly, this kind of mathematical model requires a uniformly effective numerical method, that not only allows an arbitrary equation of state, but also has more or less Mach-uniform accuracy and efficiency.

4. Final remarks

Although for computing compressible flows the use of colocated schemes is prevalent, the results obtained indicate that with staggered schemes comparable accuracy can be obtained, and extension to an arbitrary equation of state is easy. The Osher scheme is more computing-intensive, whereas other well-established schemes for computing compressible flows do not generalize easily to an arbitrary equation of state. Furthermore, by combining the staggered scheme with a pressure correction method and choosing a suitable dimensionless pressure variable, a unified scheme for compressible and incompressible flows is easily obtained, with accuracy and efficiency uniform in the Mach number.

Acknowledgements

The second author was supported by the Netherlands Organization for Scientific Research (NWO).

References

- Bijl H and Wesseling P (1998). A unified method for computing incompressible and compressible flows in boundary-fitted coordinates. *J. Comp. Phys.* **141**, pp 153-173.
- Dellanoy Y and Kueny J (1990). Two phase flow approach in unsteady cavitation modelling. Cavitation and Multiphase Flow Forum, pp. 153-158. Furuya O (Editor). New York, ASME.
- Guillard H and Viozat C (1999). On the behavior of upwind schemes in the low Mach number limit. *Computers and Fluids* **28**, pp 63-86.
- Harlow F and Welch J (1965). Numerical calculation of time-dependent viscous incompressible flow of fluid with a free surface. *The Physics of Fluids* **8**, pp 2182-2189.
- Hoeijmakers H, Janssens M and Kwan W (1998). Numerical simulation of sheet cavitation. Proc. Third International Symposium on Cavitation, April 7-10,1998, Grenoble. Michel and Kato (Editors). Volume 2, pp 257-262.
- Merkle C, Feng J and Buelow P (1998). Computational modeling of the dynamics of sheet cavitation. Proc. Third International Symposium on Cavitation, April 7-10,1998, Grenoble. Michel and Kato (Editors). Volume 2, pp 307-311.
- Oleinik O. (1957). Discontinuous solutions of nonlinear differential equations. *Uspekhi Mat. Nauk* **12**, pp 3-73.
- Osher S and Solomon F (1982). Upwind difference schemes for hyperbolic systems of conservation laws. *Math. Comp.* **38**, pp 339-374.
- Sod G (1978). A survey of several finite difference methods for systems of nonlinear conservation laws. *J. Comp. Phys.* **27**, pp 1-31.
- Sommeijer B, van der Houwen P and Kok J (1994). Time integration of three-dimensional numerical transport models. *Appl. Numer. Math.* **16**, pp 201-225.
- Song C and He J (1998). Numerical simulation of cavitating flows by single-phase flow approach. Proc. Third International Symposium on Cavitation, April 7-10,1998, Grenoble. Michel and Kato (Editors). Volume 2, pp. 295-300.
- van Albada G, van Leer B and Roberts W (1982). A comparative study of computational methods in cosmic gas dynamics. *Astron. Astrophys.* **108**, pp 76-84.
- van der Heul D and Wesseling P (1998). A staggered scheme for hyperbolic conservation laws. Computational Fluid Dynamics '98, Volume 1, pp 730-735. K Papailiou, D Tsahalis, J Périaux, C Hirsch and M Pandolfi (Editors). Wiley.
- Wesseling P, Segal A and Kassels C (1999). Computing flows on general three-dimensional nonsmooth staggered grids. *J. Comp. Phys.* **149**, pp 333-362.
- Wesseling P, Segal A, Kassels C and Bijl H (1998). Computing flows on general two-dimensional nonsmooth staggered grids. *J. Eng. Math.* **34**, pp 21-44.

WAVE INTERACTIONS IN NONLINEAR STRINGS

ROBIN YOUNG

University of Massachusetts, Amherst

Email: young@math.umass.edu

Abstract. We study the interactions of waves in a system for nonlinear elastic strings. The system can be written

$$\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}_t + \begin{pmatrix} -\mathbf{v} \\ -T(r)\frac{\mathbf{u}}{r} \end{pmatrix}_x = 0,$$

where $\mathbf{u} \in \mathbf{R}^3$ and $r = |\mathbf{u}|$. This is a 6×6 system of nonstrictly hyperbolic conservation laws in one space dimension. There are four wave families, two being genuinely nonlinear and two being degenerate. The eigenvalues corresponding to the degenerate families have multiplicity two, and the locus of points which can be connected through a jump discontinuity is a surface, rather than a curve as in the classical case. There is a highly nonlinear coupling between the different families, which we describe geometrically. We describe the interactions of elementary waves, and speculate on the nature of long-term solutions.

1. The Riemann problem

We consider a nonlinear elastic string living in 3-space. The position \mathbf{w} of a string element satisfies the equation

$$\mathbf{w}_{tt} = \left(T(|\mathbf{w}_x|) \frac{\mathbf{w}_x}{|\mathbf{w}_x|} \right)_x, \quad (1)$$

where T is the scalar tension. Writing this as a first order system, we have

$$\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}_t + \begin{pmatrix} -\mathbf{v} \\ -T(r)\frac{\mathbf{u}}{r} \end{pmatrix}_x = 0, \quad (2)$$

where we have written $r = |\mathbf{u}|$. This is a 6×6 system of conservation laws in one space dimension, which can be written in quasilinear form with flux matrix

$$DF = \begin{pmatrix} 0 & -I \\ -A & 0 \end{pmatrix} \quad \text{with} \quad A = \nabla_{\mathbf{u}} \left(T(r) \frac{\mathbf{u}}{r} \right). \quad (3)$$

It is convenient to make some specific assumptions on the scalar tension T . First, we assume that the string is elastic, so that $T(r) > 0$ for $r > 1$ (corresponding to stretching). We also assume that T is ‘rapidly increasing’: that is $(T/r)' > 0$, or equivalently $T' > T/r$, which yields a unique solution to the Riemann problem in the region $r > 1$ (Ta-Tsien, et. al., 1992).

It is useful to write $\mathbf{d} = \mathbf{u}/r \in \mathbf{S}^2$ for the direction vector of \mathbf{u} , and use r and \mathbf{d} to parameterise our wave curves. The eigensystem of the system (3) is

$$\begin{aligned} \lambda_i : & \quad -\sqrt{T'} \quad \quad \quad -\sqrt{T/r} \quad \quad \quad \sqrt{T/r} \quad \quad \quad \sqrt{T'} \\ \xi^i : & \quad \left(\begin{array}{c} \mathbf{d} \\ \sqrt{T'} \mathbf{d} \end{array} \right) \quad \left(\begin{array}{c} \mathbf{n} \\ \sqrt{T/r} \mathbf{n} \end{array} \right) \quad \left(\begin{array}{c} -\mathbf{n} \\ \sqrt{T/r} \mathbf{n} \end{array} \right) \quad \left(\begin{array}{c} -\mathbf{d} \\ \sqrt{T'} \mathbf{d} \end{array} \right) \quad (4) \\ \eta_i : & \quad \left(\frac{1}{2} \mathbf{d} \quad \frac{1}{2\sqrt{T'}} \mathbf{d} \right) \quad \left(\frac{1}{2} \mathbf{n} \quad \frac{1}{2\sqrt{T/r}} \mathbf{n} \right) \quad \left(-\frac{1}{2} \mathbf{n} \quad \frac{1}{2\sqrt{T/r}} \mathbf{n} \right) \quad \left(-\frac{1}{2} \mathbf{d} \quad \frac{1}{2\sqrt{T'}} \mathbf{d} \right) \end{aligned}$$

where λ_i , ξ^i and η_i are the eigenvalues and corresponding right and left eigenvectors, respectively. Note that the eigenvalues and eigenvectors of the 3×3 matrix A are T' and T/r and \mathbf{d} and \mathbf{n} , respectively. In our notation, the vector \mathbf{n} is *any* vector perpendicular to \mathbf{d} , i.e. any (unit) tangent to the sphere of radius r in \mathbf{u} -space. In particular, \mathbf{d} and any two independent \mathbf{n}_1 and \mathbf{n}_2 span R^3 , so unless T or T' vanishes, we have a full set of eigenvectors and the system is hyperbolic.

Because the middle eigenvalues have multiplicity 2, those families must necessarily be degenerate, and the integrals of the eigenvectors when taken together form a surface. In our case, this wave surface is parameterised by the 2-sphere $r = r_0$. We note that the degeneracy can be avoided by considering rotationally invariant solutions (Freistühler, 1991), but although this slightly reduces the work needed to solve the Riemann problem, it does result in the loss of some information about wave interactions.

We now solve the Riemann problem for system (2). First, to find the simple waves, we must integrate the eigenvectors. Rarefactions in the first family satisfy

$$\frac{d}{d\epsilon} \left(\begin{array}{c} \mathbf{u} \\ \mathbf{v} \end{array} \right) = \xi^1 = \left(\begin{array}{c} \mathbf{d} \\ \sqrt{T'} \mathbf{d} \end{array} \right),$$

and since $\nabla_u r = \mathbf{d}$, we can parameterise the curve by r and get $\mathbf{u} = \mathbf{u}(\mathbf{r}) = \mathbf{r}\mathbf{d}$, where \mathbf{d} is constant, and

$$\mathbf{v} - \mathbf{v}_0 = \int_{\mathbf{r}_0}^{\mathbf{r}} \sqrt{T'} d\mathbf{r} \mathbf{d}. \quad (5)$$

This solution works provided the wavespeed $\sqrt{T'}$ is larger on the right: if for definiteness we assume $T'' < 0$, this means $r_0 < r$. To find the shocks (and jump discontinuities) of the system, we solve the Rankine-Hugoniot equations

$$\sigma[\mathbf{u}] = -[\mathbf{v}] \quad \text{and} \quad \sigma[\mathbf{v}] = -[\mathbf{T}\mathbf{d}], \quad (6)$$

where again $\mathbf{u} = \mathbf{r}\mathbf{d}$. Solutions are easily seen to satisfy $\mathbf{d} = \mathbf{d}_0$ or $r = r_0$, with the former describing the shock waves. The 1-shock curve is given by

$$\begin{aligned} \mathbf{u} = \mathbf{r}\mathbf{d}, \quad \sigma &= -\sqrt{\frac{T(r)-T(r_0)}{r-r_0}} \\ \mathbf{v} - \mathbf{v}_0 = -\sigma(\mathbf{r} - \mathbf{r}_0)\mathbf{d} &= -\sqrt{(T(r_0) - T(r))(r_0 - r)}\mathbf{d}, \end{aligned} \quad (7)$$

where $u_0 = r_0\mathbf{d}$ and $r_0 > r$. We combine the two curves to find the wave curve for the first family: first, define the function $g : R^2 \rightarrow R$ by

$$g(r, s) = \begin{cases} \int_r^s \sqrt{T'(\rho)} d\rho, & \text{for } r \leq s, \\ -\sqrt{(T(r) - T(s))(r - s)}, & \text{for } r \geq s. \end{cases} \quad (8)$$

Now, given a state $(\mathbf{u}_0 \ \mathbf{v}_0)$, the set of all right states $(\mathbf{u} \ \mathbf{v})$ which can be reached by a 1-wave is given by

$$\mathbf{u} = \mathbf{r}\mathbf{d} = \frac{\mathbf{r}}{\mathbf{r}_0} \mathbf{u}_0 \quad \text{and} \quad \mathbf{v} - \mathbf{v}_0 = g(\mathbf{r}_0, \mathbf{r}) \mathbf{d}, \quad (9)$$

with \mathbf{d} constant. Here r is the parameter and can be used to measure wave strength. Similarly, the 4-wave curve is given by

$$\mathbf{u} = \mathbf{r}\mathbf{d} = \frac{\mathbf{r}}{\mathbf{r}_0} \mathbf{u}_0 \quad \text{and} \quad \mathbf{v} - \mathbf{v}_0 = g(\mathbf{r}, \mathbf{r}_0) \mathbf{d}, \quad (10)$$

where \mathbf{d} is again constant. We refer to these genuinely nonlinear waves as longitudinal waves. Note that the size of the quantity $g(r, s) + g(s, r)$ can be regarded as a measure of the degree of genuine nonlinearity of these waves.

The locus of states which connect to u_0 by a 2- or 3-wave can be found by solving the jump conditions (6). Parameterising the “wave surfaces” by the 2-sphere $r = r_0$ (.i.e allowing $\mathbf{d} \in \mathbf{S}^2$ to vary), we describe them as

$$\mathbf{u} = \mathbf{r}_0\mathbf{d}, \quad \mathbf{v} - \mathbf{v}_0 = \sqrt{\mathbf{T}(\mathbf{r}_0)\mathbf{r}_0} (\mathbf{d} - \mathbf{d}_0), \quad \text{and} \quad (11)$$

$$\mathbf{u} = \mathbf{r}_0\mathbf{d}, \quad \mathbf{v} - \mathbf{v}_0 = -\sqrt{\mathbf{T}(\mathbf{r}_0)\mathbf{r}_0} (\mathbf{d} - \mathbf{d}_0), \quad (12)$$

with associated wavespeeds $-\sqrt{T(r_0)/r_0}$ and $\sqrt{T(r_0)/r_0}$, respectively, and refer to them as transverse waves. We get the same locus of points by integrating the degenerate eigenvectors. What we are actually doing here is finding the integral manifold of the distribution of eigenvectors. In this case, we take the integral curves to be geodesics, and these sweep out the integral manifold.

We can now solve the Riemann problem: starting with a left reference state $(\mathbf{u}_0 \ \mathbf{v}_0) = (\mathbf{u}_L \ \mathbf{v}_L)$, we introduce intermediate states $(\mathbf{u}_1 \ \mathbf{v}_1)$, $(\mathbf{u}_2 \ \mathbf{v}_2)$, $(\mathbf{u}_3 \ \mathbf{v}_3)$ and $(\mathbf{u}_4 \ \mathbf{v}_4) = (\mathbf{u}_R \ \mathbf{v}_R)$, and connect them by elementary waves. We have

$$\mathbf{d}_1 = \mathbf{d}_0, \quad \mathbf{d}_4 = \mathbf{d}_3 \quad \text{and} \quad \mathbf{r}_1 = \mathbf{r}_2 = \mathbf{r}_3,$$

where r and \mathbf{d} are the wave strength parameters, and

$$\begin{aligned} \mathbf{v}_1 - \mathbf{v}_0 &= g(r_0, r_1) \mathbf{d}_0, \\ \mathbf{v}_2 - \mathbf{v}_1 &= \sqrt{T(r_1)r_1} (\mathbf{d}_2 - \mathbf{d}_1), \\ \mathbf{v}_3 - \mathbf{v}_2 &= \sqrt{T(r_2)r_2} (\mathbf{d}_2 - \mathbf{d}_3), \\ \mathbf{v}_4 - \mathbf{v}_3 &= g(r_4, r_3) \mathbf{d}_3, \end{aligned}$$

these being the equations for the individual waves. According to our construction, only r_2 and \mathbf{d}_2 need be found, and indeed, adding these equations yields

$$\begin{aligned} \mathbf{v}_R - \mathbf{v}_L &= \left(g(r_L, r_*) - \sqrt{T(r_*)r_*} \right) \mathbf{d}_L \\ &\quad + \left(g(r_R, r_*) - \sqrt{T(r_*)r_*} \right) \mathbf{d}_R \\ &\quad + 2\sqrt{T(r_*)r_*} \mathbf{d}_*, \end{aligned} \tag{13}$$

where we have set $r_* = r_2$ and $\mathbf{d}_* = \mathbf{d}_2$. This equation can be solved by first eliminating d_* (by squaring the quantity $2\sqrt{T_*r_*}\mathbf{d}_*$) to get a single nonlinear equation for r_* , and plugging in to get d_* , exactly as in (Ta-Tsien, et. al., 1992).

2. Interactions

When two elementary waves meet, we solve the interaction problem by posing a new Riemann problem and relating the emerging waves to the incident waves. Provided the incident waves are weak, we can describe the interaction to leading order as follows: suppose a j -wave of strength α_j interacts with a k -wave of strength β_k ; then the correction to the i -th family due to the interaction is, to leading order, $\alpha_j \beta_k \Lambda_i^{jk}$, where the *interaction*

coefficients are defined by the Lie brackets

$$[\xi^j, \xi^k] = \sum \Lambda_i^{jk} \xi^i. \quad (14)$$

The error in the interaction is the bracket, and its projection onto eigenvectors gives the nonlinear correction to each wave family.

Once we have these interaction coefficients, we can give a good qualitative description of interactions of weak waves. This in turn helps us deduce possible long-range behaviour of solutions. The coefficients can be calculated directly by differentiating the eigenvectors, or from the Hessian of the flux.

To calculate the coefficients from the flux, we first fix an appropriate pair of eigenvectors for the repeated eigenvalues: this amounts to a choice of two vectors $\mathbf{n} \in S^2$. The *symmetric flux coefficients*, defined by

$$\Gamma_{jk}^i = \eta_i \cdot D^2 F(\xi^j, \xi^k),$$

encode all leading order information about the nonlinearity, including the interaction coefficients and the degree $\xi^i \nabla_u \lambda_i$ (Young, 1993). See (Young, 2000) for the calculation of these coefficients.

It is somewhat more enlightening to calculate the coefficients directly. For brevity, we discuss only one case, the interaction of transverse waves of different families. Thus, consider the two eigenvectors

$$\xi^2 = \begin{pmatrix} \mathbf{n} \\ \sqrt{T/r} \mathbf{n} \end{pmatrix} \quad \text{and} \quad \xi^3 = \begin{pmatrix} -\mathbf{m} \\ \sqrt{T/r} \mathbf{m} \end{pmatrix}$$

where \mathbf{n} and \mathbf{m} are unit tangents to S^2 , with $\mathbf{n} \cdot \nabla_u \mathbf{m} = \mathbf{m} \cdot \nabla_u \mathbf{n} = -\cos \alpha \mathbf{d}$, where α is the angle between them. After simplifying we get

$$[\xi^3, \xi^2] = \begin{pmatrix} 0 \\ 2\sqrt{T/r} \cos \alpha \mathbf{d} \end{pmatrix} = \sqrt{\frac{T}{rT'}} \cos \alpha (\xi^1 + \xi^4).$$

This calculation tells us what happens when two opposite transverse waves meet, namely that longitudinal waves are reflected out from the point of interaction. These reflected waves have equal strength, and indeed, the secondary waves corresponding to all other interactions also occur in pairs of equal strength (to leading order). The appearance of the term $\cos \alpha$ is interesting in that it represents a weakening of the coupling in higher dimensions: if we restrict to $\mathbf{u} \in \mathbf{R}^2$, this coefficient is always 1. Physically, two transverse waves will produce a longitudinal effect which depends on the polarisation between them, with orthogonal waves crossing linearly and polarised waves interacting fully. Loosely speaking, a spatial string is more flexible than a planar string.

The remaining interaction coefficients are calculated similarly, and we can deduce the leading order effects of other wave interactions in the same way. We refer the reader to (Young, 2000) for a detailed analysis of these coefficients and interactions. Here we will only briefly describe the different interactions that may occur.

The interaction of any two longitudinal waves (i.e. 1- or 4-waves) does not generate any transverse component, and so if the solution initially lives on a ray, it will remain on the ray, and because these waves are genuinely nonlinear, we expect such a solution to decay to N -waves.

All brackets of the form $[\xi^l, \xi^t]$ of longitudinal and transverse waves lie in the span of the vector $(0, \mathbf{n})^T$, where ξ^t has polarisation \mathbf{n} . This says that when a transverse wave crosses a longitudinal wave, it is scattered into transmitted and reflected transverse waves of the same polarisation. To leading order, the longitudinal wave is unaffected.

3. Global behaviour

We can combine the foregoing to attempt to predict the long term behaviour of solutions. We are most interested in periodic boundary conditions. First we note some consequences of our calculations for the structure of the equations. Because a wave of each family can be generated (to leading order) by some nonlinear interaction, we know that there are no Riemann invariant coordinates for the system, so the system cannot be reduced. This says that no wave field decouples, even in a weak sense.

It is also interesting to note that the system is *not* symmetrizable. This follows from a direct calculation involving the interaction coefficients: if the system were symmetrizable, then equality of mixed partials implies that the expression

$$\frac{\Lambda_i^{jk}}{\lambda_j - \lambda_k} + \frac{\Lambda_j^{ki}}{\lambda_k - \lambda_i} + \frac{\Lambda_k^{ij}}{\lambda_i - \lambda_j} \quad (15)$$

vanishes for all triples $\{i, j, k\}$ (Young, 1993). This comes from the equality of mixed partials: by differentiating the eigenvector relation $DF \cdot r = \lambda r$, the Lie bracket coefficients Λ_i^{jk} can be given in terms of the Hessian D^2F ; if the system were symmetric hyperbolic, DF would be symmetric, so that $F = D\phi$, and in turn $D^2F = D^3\phi$. Resolving the symmetry of the trilinear form $D^3\phi$, the expression (15) vanishes for all triples. However, for the present system, the quantity (15) gives $\frac{1}{2\sqrt{T'}}(1 - \frac{rT'}{T})$, so that our system is not symmetrizable. One consequence of symmetrizability is that the equation possesses a convex entropy, which can be used to pick out admissible discontinuities, and which gives a formal bound for solutions. Thus we have an example of a “physical” system of equations, which does not possess a

convex entropy. However, according to (Keyfitz and Kranzer, 1980), this system does possess an energy functional. The difference here is that the energy functional is defined using derivatives, rather than just the state variables, and so an equation for the energy functional cannot be deduced directly from the conservation law.

We now attempt to put together many interactions and speculate as to the long-term behaviour of the system. If we consider periodic data, then the different wave families cannot separate, and interactions between different families will keep occurring for all time. As we have seen, the longitudinal waves would decay to N -waves in the absence of any transverse perturbations. However, when transverse waves are included, the situation becomes more complicated. The effect of longitudinal waves is to diffuse transverse waves by scattering them. Thus as few as three initial waves will lead to significant diffusion in the linearly degenerate fields. Moreover, when two transverse waves interact, new longitudinal waves are generated, and these will act as diffusers themselves. Thus we expect twists in our string to diffuse throughout the string, while longitudinal waves will vibrate up and down the length of the string, gradually decaying in strength.

Finally, we describe the weakly nonlinear geometric optics (WNGO) approximation to the spring system. It is well known that the symmetric flux coefficients appear in the expansion (Hunter, 1989). The WNGO approximation contains convolution terms associated with these coefficients. It is shown in (Hunter, 1989) that these convolution terms vanish unless a ratio of differences in wavespeeds, $(\lambda_i - \lambda_j)/(\lambda_i - \lambda_k)$ is rational. However, in our case this condition is simply the requirement that the quantity $\sqrt{T(r)/r T'(r)}$ be rational, where r is the unperturbed state. Thus the WNGO approximation takes on different forms depending on the background state. To my knowledge this is the first example of such a system. For further details we refer the reader to (Young, 2000).

References

- H. Freistühler (1991) Rotational degeneracy of hyperbolic systems of conservation laws. *Arch. Rational Mech. Anal.*, **113**: 39.
- John Hunter (1989) Strongly Nonlinear Hyperbolic Waves. *Nonlinear Hyperbolic Equations – Theory, Computation Methods, and Applications*, J. Ballman & R. Jeltsch, eds. Viewig: 257.
- B.L. Keyfitz & H. Kranzer (1980) A system of non-strictly hyperbolic conservation laws arising in elasticity theory. *Arch. Rational Mech. Anal.*, **72**: 219.
- Li Ta-Tsien, D. Serre and Zhang Hao (1992) The Generalized Riemann Problem for the motion of Elastic Strings. *SIAM J. Math Anal.*, **23**: .
- Robin Young (1993) On elementary interactions for hyperbolic conservation laws. *Preprint, available at <http://math.umass.edu/~young/Archive/elem.ps.gz>.*
- Robin Young (2000) Wave Interactions in Nonlinear Strings. *Preprint, draft available at <http://math.umass.edu/~young/Archives/string.ps.gz>.*

INDEX

- Accretion disc, 525
- Adaptive mesh, 163, 527, 539, 812
- Adaptive numerical flux, 500
- Adaptive Riemann solver, 719, 722
- Adaptive scheme, 763
- Adaptive two-shock solver, 721
- Adaptive two-step method, 720
- ADER in 2D and 3D, 922, 923, 924
- ADER schemes, 907, 909, 915, 921
- ADER schemes for non-linear problems, 924, 925, 926, 927
- Advection scheme, 125, 833, 834
- Advection-dispersion, 603, 604
- Anomalous fluid behaviour, 717, 718, 722, 723
- Anomalous van der Waals gas, 718, 721
- A-priori estimate, 293, 527, 531, 536
- Approximate Riemann solver, 955, 1051
- Arbitrary Lagrangian Euler (ALE) method, 507, 993
- Artificial boundaries, 217
- Artificial compressibility, 135
- Artificial compression method, 595, 596, 601
- Artificial dissipation, 873
- Artificially compressible fluid, 962
- Astrophysical flows, 462, 519
- Astrophysical jet, 45
- Atmospheric chemistry and transport, 1015
- Atmospheric dispersion, 117
- AUSM, 713, 1059
- Balance laws, 745
- Bathymetry, 142
- BDF scheme, 335, 337, 340, 341
- Bicharacteristics, 573, 691
- Black hole, 524
- Boltzmann equation, 225
- Bottom friction, 864
- Boundary conditions, 164, 165, 167, 218, 1010
- Boundary-fitted curvilinear system, 142
- Boundary forces, 98
- Bow shock, 247
- BZT fluid, 419
- Capturing water waves, 949
- Carbuncle, 623, 639, 711, 711
- Cavitation, 506, 509, 785, 788, 970, 1057, 1062, 1063
- Centred scheme, 899, 891, 902, 903, 947
- Characteristics, 296
- Characteristic-based scheme, 263
- Characteristic scheme, 671, 672, 674
- Characteristic wave speed, 586
- Charge conservation equation, 647, 649, 651
- Charged particle flow, 648
- Cartesian cell, 170
- Cartesian grid, 161, 162, 169, 170, 175, 1008
- Cell vertex, 149
- CFL condition, 119, 167
- CFL number, 977
- CLAWPACK, 549, 550
- Combustion, 557
- Composite wave structure, 722, 723
- Compressible flow, 149
- Compressible fluid, 960
- Compressible pressure correction, 972
- Computational efficiency, 665
- Conservation, 310
- Conservation form, 580, 1059
- Conservation laws, 97, 233, 327, 399, 497, 557, 833, 834, 941
- Conservative method, 125, 186
- Conservation principle, 167, 172
- Conservative scheme, 162, 167, 175
- Conserved variables, 701
- Consistency condition, 125
- Contact discontinuity, 310, 312, 313, 315, 595, 600, 601, 623, 631, 1045, 1052
- Control volume, 162, 167, 168, 175
- Convection-diffusion equation, 474
- Convection-dominated nonlinear PDE, 469
- Convex cone, 229
- Convex entropy, 205
- Corrected operator splitting, 474
- Courant number, 946
- De Saint-Venant equations, 809
- Degrees of freedom, 825
- Dense gas, 422
- Derivative Riemann problems, 913, 920
- Detonation, 445, 785
- Difference approximation, 425
- Diffusion-advection equation, 837
- Diffusion term, 663
- Diffusive relaxation, 655, 660
- Diphasic equation, 228

- Dimensionally split, 1046
 Dimensionless diffusion number, 663
 Discontinuity, 445, 447, 449, 450
 Discontinuous Galerkin finite element method, 691, 987, 994
 Discrete Fourier Transform, 61
 Distribution of sources, 464
 Divergence constraint, 649
 Divergence wave, 192
 Double hotspot, 464
 Downwinding, 946
 Drift flux model, 319, 320, 322
 Dual-time stepping, 263
 E-condition, 1, 2, 3, 5, 6, 14, 15, 22, 23, 24
 Elliptical constraint, 647, 648
 Energy-momentum tensor, 487
 Energy relaxation method, 181, 185
 ENO scheme, 254, 263, 579, 581, 890, 891, 908, 909, 928, 929
 Ensemble of clouds, 466
 Entropy, 29, 312, 314, 315, 1029, 1038
 Entropy characterization, 203, 205
 Entropy condition, 2, 4, 5, 687, 1059, 1061
 Entropy fix, 2, 4, 6, 14, 24, 92, 590, 685, 686, 687, 688
 Entropy increase, 567
 Entropy inequality, 2, 5, 6, 23, 187, 233
 Entropy production, 558
 Entropy shock, 205
 Entropy solution, 2, 4, 5, 6, 23, 24
 Entropy stable scheme, 38
 Entropy variables, 32
 Equilibrium manifold, 726
 Error estimate, 815
 Error indicator, 122
 Euler equations, 78, 155, 234, 311, 312, 313, 314, 412, 472, 575, 823, 993, 1059, 1061
 Exact linearization, 611
 Exact Riemann solver, 203
 Expansion shock, 1061
 Expert system, 977
 Explicit scheme, 290
 Extragalactic jet, 461
 Finite differences, 162, 163, 174, 309, 316
 Finite volume, 179, 184, 603, 604, 605, 606, 866, 1008
 Finite volume method, 161, 285, 478, 498, 505, 507, 571, 573
 Finite volume scheme, 119, 120, 528
 Flood map, 864
 Fluctuation splitting, 190, 192, 193, 194
 Fluctuation splitting scheme, 135
 Flux corrected transport (FCT), 886
 Flux correction procedure, 598
 Flux difference splitting, 89, 866
 Flux form, 1028, 1053
 Flux function, 611
 Flux-limited scheme, 833
 Flux limiter, 886
 Flux separation, 446, 448
 Flux vector splitting, 581, 623, 671, 672, 675, 842
 FORCE scheme, 899, 900, 901, 902
 Four-shock interaction, 1039
 Free-boundary problem, 161, 172
 Free surface, 445
 Free-surface flow, 253
 Free-surface model, 125
 Front tracking, 445, 446, 471
 Fundamental derivative, 717, 718, 721, 722, 723
 Galerkin method, 571, 572, 573
 Galilean invariance, 233
 Gamma-ray burst, 53
 Garcia–Navarro, 809
 Gas dynamics, 78, 155
 Gas-liquid flow, 319
 Gauss law, 647
 Gauss–Seidel iteration, 118
 General equation of state, 1060, 1063
 Generalised Riemann problem, 915
 Geophysical flow, 833, 834
 Global error, 119
 Godunov method, 2, 4, 5, 6, 8, 77, 117, 118, 119, 169, 170, 179, 182, 203, 204, 209, 218, 248, 263, 327, 328, 329, 505, 527, 528, 647, 677, 717, 718, 720, 721, 785, 809, 815, 825, 879, 880, 907, 922, 941, 946, 1024, 1026, 1051
 Godunov theorem, 907, 908
 Grid alignment, 707
 Grid convergence, 200
 GRP scheme, 909
 Hancock, 889
 Hancock scheme, 720
 Harten and Hyman entropy fix, 687
 Harten, Lax and van Leer (HLL), 885
 High altitude flight, 740
 High-order Godunov scheme, 722, 891
 High-order methods, 263, 907
 High-order reconstruction, 77
 Higher-order nonoscillatory scheme, 838
 High-resolution method, 722
 HLL, 1024, 1026
 HLLC Riemann solver, 69, 70, 71, 72, 73, 74, 75, 507, 885, 1024
 HLLE scheme, 588, 685, 687, 1036
 HLLM scheme, 687
 Hydraulic jump, 288, 289
 Hyperbolic conservation laws, 167, 377, 378, 379, 571, 572, 578, 899, 900, 901, 941, 943, 945, 969

- Hyperbolic diffusion, 725
 Hyperbolic partial differential equations, 987, 993
 Hyperbolic relaxation, 725
 Hyperbolic relaxation approximation, 655
 Hyperbolic systems, 293, 691
 Hybrid Finite Element, 603, 604, 605, 606
 Hypersonic flow, 453

 Ideal MHD, 191, 192
 Implicit Godunov method, 611
 Implicit Runge–Kutta method, 174
 Implicit scheme, 69, 70
 Implicit upwind method, 977
 Incompressible fluid, 963
 Inhomogeneous medium, 109
 Inhomogeneous transport, 837
 Initial-boundary value problem, 425
 Interaction coefficient, 1069
 Interaction problem, 1068
 Interface, 309, 310, 312, 313, 314, 315, 785
 Interface capturing, 367, 1009
 Interface reconstruction, 367, 374
 Interface tracking, 367, 368, 1007
 Intermediate density, 420
 Intermediate shock, 247
 Irregular boundaries, 163, 174

 Jet-cloud collision, 461

 Kinematic frontogenesis, 1018
 Kinetic-fluid coupling, 230
 Kinetic scheme, 225, 226
 Kinetic system, 655, 656, 657, 659

 Lagrangian coordinates, 234, 377, 378, 383, 384, 385, 388
 Large-eddy simulation, 834, 838
 Large-time step, 1015, 1016
 Laval nozzle, 83
 Lax–Friedrichs, 4, 5, 6, 14, 248, 599, 1024, 1047
 Lax inequality, 205
 Lax–Wendroff, 1024
 Lax–Wendroff theorem, 30
 Lax–Wendroff TVD scheme, 1009
 LDA scheme, 36
 Level set, 368, 445, 446, 447
 Lie brackets, 1069
 Limiter, 412
 Linearised stability, 378
 Linearization, 347, 348
 Liou’s Method (AUSM+), 677
 Local error, 120
 - high-order reconstruction, 773
 Low Mach number, 161, 1057
- ick scheme, 198
 ction of bores, 905
- Magnetic field, 191, 192
 Magnetohydrodynamics, 189, 209, 217, 247
 Marquina solver, 590
 Mass flux, 623
 Mass fraction, 205, 209
 Material interface, 161, 445, 448
 Maxwell equations, 648, 649, 691
 Maxwell solver, 647
 Measure-valued solution, 97
 Mesh quality, 121
 Mesh refinement, 1007
 Meshless methods, 97
 Meteorological analysis, 1015, 1020
 Method of lines, 584, 863
 Method of transport, 671, 672
 MHD, 247
 MHD Riemann problem, 211
 MHD Riemann solver, 209
 Microscale devices, 740
 Mie–Gruneisen, 509
 Minkowskian coordinates, 492
 Modified equation, 163, 166, 170
 Modified GRP scheme, 912, 914
 Monotone transport scheme, 837
 Moving contact discontinuity, 456
 Multi-dimensional decomposition, 825
 Multi-dimensional linearisation, 672, 673
 Multi-dimensional upwinding, 189, 190, 343, 344, 345, 355, 359
 Multi-dimensional wave model, 671, 672, 673
 Multi-grid method, 263, 1009
 Multi-level algorithm, 155, 156, 158, 159
 Multi-phase flow, 185, 367, 785, 789
 Multi-phase media, 756
 Multi-resolution analysis, 497
 MUSCL, 887
 MUSCL-Hancock scheme, 903, 911, 912, 945
 MUSCL scheme, 319, 322, 323, 650
 MUSCL-TVD scheme, 199

 N scheme, 34
 Naval hydrodynamics, 253
 Navier–Stokes equations, 707
 Negative nonlinearities, 419
 Nessyahu–Tadmor method, 700
 Nine-state Riemann solver, 1031
 Noh’s problem, 226, 231
 Non-classical shock wave, 247
 Non-conservative formulation, 826
 Non-conservative scheme, 168
 Non-conservative terms, 790
 Non-convex equation of state, 1057, 1060, 1062
 Non-convex isentrope, 717, 718, 721
 Non-convex region, 718
 Non-linear elastic string, 1065
 Non-oscillatory schemes, 907
 Non-strictly hyperbolic, 247, 1065

- Northern polar vortex. 1020
 Numerical diffusion. 513
 Numerical dissipation. 685
 Object oriented programming. 446
 Oleinik, I., 2, 3, 4, 5, 7
 Open channel flow. 478
 Operator splitting. 179, 184, 470, 816
 Osher scheme. 285, 287, 288, 707, 884, 956, 958,
 1057, 1059, 1061, 1062, 1063
 O-variant. 956
 Outward normal. 162
 Parabolic equation. 655, 656
 Parabolized Navier–Stokes (PNS). 335
 Particle-in-cell method. 648
 Perfect gases. 89, 453, 1060
 Phase transition. 717, 718
 Piecewise Hyperbolic Method (PHM). 584
 Piecewise-parabolic reconstruction. 590
 Plasma. 247
 Poisson’s equation. 163, 164, 165
 Porous–Fischer equation. 655, 659, 660
 Positivity. 631, 1038
 Positivity-conserving. 1029
 Potential vorticity. 1021
 Powell source term. 249
 PPM. 891
 Pressure correction. 1057, 1058, 1059, 1060, 1063
 Pressurised water reactor. 179
 Primitive-conservative scheme. 319
 Pseudo-acoustic shear wave. 736
 PSI scheme. 37
 P-variant. 956
 Quadtree grid. 141
 Quasar. 461
 Quasi-conservative approach. 629, 637
 Radio galaxy. 461
 Radio jet. 462
 Random walk. 666
 Rankine–Hugoniot relations. 250, 311, 313, 314,
 316, 702
 Rarefactive shock. 717, 721, 722
 Rayleigh–Taylor instability. 1011
 Reconstruction. 149
 Relativistic blast wave. 587
 Relativistic Euler equations. 54, 584
 Relativistic fluid dynamics. 486
 Relativistic hydrodynamics. 46, 54, 327, 328, 330,
 485, 699, 700
 Relativistic jet. 45, 53, 54, 55, 57, 58, 486, 521, 590
 Relativistic MHD. 519
 Relativistic Riemann solver. 491
 Relaxation. 818
 Relaxation approximation. 820
 Relaxation effect. 302
 Residual distribution. 189, 192
 Residual scheme. 27, 30
 Richtmyer–Meshkov instability. 69, 377, 378, 380,
 387
 Riemann problem. 203, 205, 286, 287, 288, 310, 311,
 316, 581, 587, 699, 703, 704, 705, 911,
 946, 951, 1015, 1017, 1025, 1026, 1057,
 1059, 1061
 Riemann problem with relaxation. 730
 Riemann solver. 69, 74, 179, 182, 199, 233, 579,
 699, 700, 703, 704, 707, 842, 883
 River-bed geometry. 864
 River flow. 863, 864
 Roe scheme. 142, 319, 322, 324, 325, 419, 477, 677,
 685, 883, 1053, 1059
 Rosenbrock–Wanner. 866
 Rotational invariance. 399
 Runge–Kutta method. 199, 285, 288, 584, 1058,
 1059
 Rusanov scheme. 203, 207
 Scalar transport. 125
 Semi-implicit scheme. 125
 Semi-Lagrangian. 296
 Semi-Lagrangian approximation. 837
 Separated boundary layer. 199
 Separated flow. 198
 Shallow water equations. 141, 348, 359, 477, 673,
 675, 676, 863, 864, 899, 903, 942, 943, 944
 Shear layer. 450
 Shock admissibility. 247
 Shock capturing. 189, 194
 Shock capturing scheme. 286
 Shock curve. 90
 Shock stability. 247
 Shock structure. 302, 304, 306, 307
 Shock tracking. 172
 Shock tube. 718, 721, 741, 1058
 Shock visualisation. 553
 Shock wave. 199, 639, 763
 Shocklets. 200
 Sign preserving advection. 839
 Signal velocity. 686, 700, 701, 703
 Simulation of jet. 54
 Sink term. 285, 286
 SLIC scheme. 899, 900, 902, 903
 Slope limiter. 330
 Slope limited scheme. 1058, 1059
 Sod’s problem. 226, 231
 Solar convection zone. 218
 Sonic point. 2, 8, 13, 23, 24, 672, 675, 676
 Source term. 78, 189, 190, 192, 195, 228, 230, 285,
 286, 288, 291, 345, 354, 477, 514, 557,
 565, 726, 745, 809, 815, 893, 941, 942,
 943, 945, 948
 Space-marching. 335, 341

- Space-time limited, 335, 337, 341
Space vehicle, 726
Spatial resolution, 663
Splitting error, 816
Splitting method, 185
Splitting technique, 286, 288, 291
Stability, 1028
Stability theory, 425
Staggered grid, 969
Staggered mesh, 738
Staggered scheme, 1057, 1058, 1059, 1061, 1062, 1063
Steady state, 77
Stiff source term, 817
Stratospheric ozone, 1020
Stratospheric transport, 1015
Strong rarefactions, 207
Sub-shock, 302, 306, 307
Sub-sonic flow, 823
Supersonic flow, 823, 831
Symmetric hyperbolic systems, 745
Symmetrization theorem, 233
Symmetrizable system, 1070

Taub adiabat, 702
Tetrahedral mesh, 117
Thermoelastic wave, 109
Thermodynamics, 399, 745
Time-dependent PDE, 425
Time-splitting, 603, 604, 605, 606, 607, 608
Transonic rarefaction, 686
Translational number, 61, 63
Travelling wave, 596, 597
Triple point, 200
Truncation error, 163, 164, 165, 166, 170
Turbulent kinetic energy, 203

Total variation diminishing (TVD), 2, 4, 6, 14, 197, 330, 337, 477, 886
TVD Lax–Wendroff scheme, 873
TVD-MUSCL scheme, 411
Two-dimensional Riemann problem, 1030
Two-fluid models, 843, 844
Two-phase flow, 179, 180, 474, 513, 677, 841, 965, 1007

Unconditionally stable, 471
Underwater shock wave, 505, 511
Upwind method, 639, 946, 947
Unsplit finite volume scheme, 899, 901, 902, 903
Unstructured mesh, 27, 119, 120, 135, 149, 450, 527, 603, 605, 763, 873

Vacuum, 92
Van Albada limiter, 1058
van der Waals gas, 419, 717, 718, 719, 721
Vazquez–Cendon, 809
Vfroe scheme, 203, 205, 207
Viscous flow, 197
Viscous shock tube, 197
Vlasov–Maxwell equation, 648
Volume of fluid method, 161, 174, 175, 368
Von Neumann–Richtmyer, 243

Wave equation, 296, 574
Wavelet theory, 497
Weakly nonlinear geometric optics (WNGO), 1071
Weighted Average Flux (WAF) method, 892, 900, 901, 1015
WENO schemes, 891, 909
Whitham criterion, 304, 308, 309

Young scheme, 370, 372, 373