Springer Proceedings in Mathematics & Statistics

Clément Cancès
Pascal Omnes *Editors*

Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems

FVCA 8, Lille, France, June 2017

Springer

Springer Proceedings in Mathematics & Statistics

Volume 200

Springer Proceedings in Mathematics & Statistics

This book series features volumes composed of selected contributions from workshops and conferences in all areas of current research in mathematics and statistics, including operation research and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

More information about this series at http://www.springer.com/series/10533

Clément Cancès · Pascal Omnes
Editors

Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems

FVCA 8, Lille, France, June 2017

Springer

Editors
Clément Cancès
Equipe RAPSODI
Inria Lille - Nord Europe
Villeneuve-d'Ascq
France

Pascal Omnes
Commissariat à l'énergie atomique et aux énergies alternatives
Centre de Saclay
Gif-sur-Yvette
France

ISSN 2194-1009 ISSN 2194-1017 (electronic)
Springer Proceedings in Mathematics & Statistics
ISBN 978-3-319-57393-9 ISBN 978-3-319-57394-6 (eBook)
DOI 10.1007/978-3-319-57394-6

Library of Congress Control Number: 2017938633

Mathematics Subject Classification (2010): 65-06, 65Mxx, 65Nxx, 76xx, 86-08, 92-08

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The finite volume method consists in a space discretization technique for partial differential equations. It is based on the fundamental principle of local conservation (or more generally local balance), making it very natural and successful in many applications, including fluid dynamics, magnetohydrodynamics, structural analysis, nuclear physics, and semiconductor theory. Motivated by their large applicability for real-world problems, finite volumes have been the purpose of an intensive research effort in the last decades, yielding significant progresses in the design, the numerical analysis, and the practical implementation of the methods.

Research on finite volumes remains very active since the problems to solve are everyday more complex and demanding. Among the current challenges addressed by the scientific community, let us mention for instance the design of robust (with respect to the mesh and/or physical parameters) numerical methods, of high-order methods, and of methods preserving structural properties (positivity and dissipation of a prescribed quantity). The implementation of such methods on new architectures is also a crucial issue.

Previous conferences on this series have been held in Rouen (1996), Duisburg (1999), Porquerolles (2002), Marrakech (2005), Aussois (2008), Prague (2011), and Berlin (2014).

The present volumes contain the invited and contributed papers presented as posters or talks at the Eights International Symposium on Finite Volumes for Complex Applications held in Lille, June 12–16, 2017. It also contains a benchmark on discretizations for incompressible viscous flows governed by Stokes and Navier–Stokes equations.

The first volume contains the invited contributions, the benchmark on discretizations for incompressible viscous flows, and some contributed papers focusing on theoretical aspects of finite volumes, including discrete functional analysis tools, convergence proof, and error estimates for problems governed by partial differential equations.

The second volume is focused on the simulation of problems arising in real-world applications, such as complex fluid mechanics, elasticity problems, and complex porous media flows.

The volume editors thank the authors for their high-quality contributions, the member of the program committee for supporting the organization of the review process, and all reviewers for their thorough work on the evaluation of each of the contributions.

The organization of the conference was made possible thanks to the financial support of Lille 1 University, the Centre National pour la Recherche Scientifique (CNRS), Inria, Total, IFP Energies nouvelles, the CEA, the Labex CEMPI and AMIES, the Weierstrass Institute for Applied Analysis and Stochastics (WIAS), the Universities of Nice, Paris 13, and Paris-Est Marne-la-Vallée.

Finally, we warmly thank the local organization committee and staff for their precious help to make this conference a friendly moment.

Villeneuve-d'Ascq, France — Clément Cancès
Gif-sur-Yvette, France — Pascal Omnes
March 2017

The original version of the book was revised: Missed out corrections have been updated. The erratum to the book is available at *https://doi.org/10.1007/978-3-319-57394-6_58*

Organization

Program Chairs

Jürgen Fuhrmann, Weierstrass Institute, Germany
Clément Cancès, Inria Lille - Nord Europe
Pascal Omnes, CEA Saclay, DM2S - STMF, France

Program Committee

Jerome Droniou, Monash University, Australia
Raphaèle Herbin, Aix-Marseille Université, France
Marsha Berger, New York University, United States
Clément Cancès, Inria Lille, France
Carlos Parés, Universidad de Málaga, Spain
Martin Vohralik, Inria Paris, France
Franck Boyer, Université Paul Sabatier - Toulouse 3, France
Volker John, Weierstrass Institute, Germany
Vincent Couaillier, ONERA, France
Mario Ohlberger, Universität Münster, Germany
Jean-Marc Hérard, EDF R&D, France
Jiří Fürst, Czech Technical University in Prague, Czech Republic
Emmanuel Audusse, University Paris 13, France
Konstantin Lipnikov, Los Alamos National Laboratory, United States
Karol Mikula, Slovak University of Technology, Slovakia
Jean-Claude Latché, IRSN, France
Jan Martin Nordbotten, Bergen University, Norway
Arthur Moncorgé, TOTAL E&P UK, United Kingdom
Maria Lukacova, Johannes Gutenberg-Universität Mainz, Germany
Fayssal Benkhaldoun, University Paris 13, France
Pascal Omnes, CEA Saclay, DM2S - STMF, France

Claire Chainais-Hillairet, Université Lille 1, France
Roland Masson, University Nice Sophia Antipolis, France
Jürgen Fuhrmann, Weierstrass Institute, Germany
Peter Bastian, Universität Heidelberg, Germany
Michael Dumbser, University of Trento, Italy
Siegfried Müller, RWTH Aachen, Germany

Contents

Part I Hyperbolic Problems

A Weighted Splitting Approach for Low-Mach Number Flows 3
David Iampietro, Frédéric Daude, Pascal Galon and Jean-Marc Hérard

Weno Scheme for Transport Equation on Unstructured Grids with a DDFV Approach 13
Florence Hubert and Rémi Tesson

New Types of Jacobian-Free Approximate Riemann Solvers for Hyperbolic Systems 23
Manuel J. Castro, José M. Gallardo and Antonio Marquina

A Fractional Step Method to Simulate Mixed Flows in Pipes with a Compressible Two-Layer Model 33
Charles Demay, Christian Bourdarias, Benoît de Laage de Meux, Stéphane Gerbi and Jean-Marc Hérard

A Second Order Cell-Centered Scheme for Lagrangian Hydrodynamics 43
Théo Corot

An Implicit Integral Formulation for the Modeling of Inviscid Fluid Flows in Domains Containing Obstacles 53
Clément Colas, Martin Ferrand, Jean-Marc Hérard, Erwan Le Coupanec and Xavier Martin

A High-Order Discontinuous Galerkin Lagrange Projection Scheme for the Barotropic Euler Equations 63
Christophe Chalons and Maxime Stauffert

Sensitivity Analysis for the Euler Equations in Lagrangian Coordinates 71
Christophe Chalons, Régis Duvigneau and Camilla Fiorini

Semi-implicit Level Set Method with Inflow-Based Gradient in a Polyhedron Mesh 81
Jooyoung Hahn, Karol Mikula, Peter Frolkovič and Branislav Basara

A Staggered Scheme for the Euler Equations 91
Thierry Goudon, Julie Llobell and Sebastian Minjeaud

A Numerical Scheme for the Propagation of Internal Waves in an Oceanographic Model 101
Christian Bourdarias, Stéphane Gerbi and Ralph Lteif

A Splitting Scheme for Three-Phase Flow Models 109
Hamza Boukili and Jean-Marc Hérard

Modelling and Simulation of Non-hydrostatic Shallow Flows 119
M.J. Castro, C. Escalante and T. Morales de Luna

A Flux Splitting Method for the Baer-Nunziato Equations of Compressible Two-Phase Flow 127
Svetlana Tokareva and Eleuterio Toro

GPU Accelerated Finite Volume Methods for Three-Dimensional Shallow Water Flows 137
Mohamed Boubekeur, Fayssal Benkhaldoun and Mohammed Seaid

Projective Integration for Nonlinear BGK Kinetic Equations 145
Ward Melis, Thomas Rey and Giovanni Samaey

Asymptotic Preserving Property of a Semi-implicit Method 155
Lei Zhang, Jean-Michel Ghidaglia and Anela Kumbaro

A Finite-Volume Discretization of Viscoelastic Saint-Venant Equations for FENE-P Fluids 163
Sébastien Boyaval

Palindromic Discontinuous Galerkin Method 171
David Coulette, Emmanuel Franck, Philippe Helluy, Michel Mehrenberger and Laurent Navoret

IMEX Finite Volume Methods for Cloud Simulation 179
M. Lukáčová-Medvid'ová, J. Rosemeier, P. Spichtinger and B. Wiebe

Hybrid Stochastic Galerkin Finite Volumes for the Diffusively Corrected Lighthill-Whitham-Richards Traffic Model 189
Raimund Bürger and Ilja Kröker

The RS-IMEX Scheme for the Rotating Shallow Water Equations with the Coriolis Force 199
Hamed Zakerzadeh

Analysis of Apparent Topography Scheme for the Linear Wave Equation with Coriolis Force 209
Emmanuel Audusse, Minh Hieu Do, Pascal Omnes and Yohan Penel

Application of a Combined Finite Element—Finite Volume Method to a 2D Non-hydrostatic Shallow Water Problem 219
Nora Aïssiouene, Marie-Odile Bristeau, Edwige Godlewski, Anne Mangeney, Carlos Parés and Jacques Sainte-Marie

A Relaxation Scheme for the Simulation of Low Mach Number Flows 227
Emanuela Abbate, Angelo Iollo and Gabriella Puppo

Comparison of Wetting and Drying Between a RKDG2 Method and Classical FV Based Second-Order Hydrostatic Reconstruction 237
Stefan Vater, Nicole Beisiegel and Jörn Behrens

A Discontinuous Galerkin Method for Non-hydrostatic Shallow Water Flows 247
Anja Jeschke, Stefan Vater and Jörn Behrens

Design of a Second-Order Fully Explicit Residual Distribution Scheme for Compressible Multiphase Flows 257
Rémi Abgrall and Paola Bacigaluppi

An Unstructured Forward-Backward Lagrangian Scheme for Transport Problems 265
Martin Campos Pinto

A Godunov-Type Scheme for Shallow Water Equations Dedicated to Simulations of Overland Flows on Stepped Slopes 275
Nicole Goutal, Minh-Hoang Le and Philippe Ung

Two Models for the Computation of Laminar Flames in Dust Clouds 285
Dionysios Grapsas, Raphaèle Herbin and Jean-Claude Latché

High Order Finite Volume Scheme and Conservative Grid Overlapping Technique for Complex Industrial Applications 295
Grégoire Pont and Pierre Brenner

Part II Elliptic and Parabolic Problems

Discontinuous Finite Volume Element Methods for the Optimal Control of Brinkman Equations 307
Sarvesh Kumar, Ricardo Ruiz-Baier and Ruchi Sandilya

Non-isothermal Compositional Two-Phase Darcy Flow: Formulation and Outflow Boundary Condition 317
L. Beaude, K. Brenner, S. Lopez, R. Masson and F. Smai

Numerical Scheme for a Stratigraphic Model with Erosion Constraint and Nonlinear Gravity Flux 327
Clément Cancès, Didier Granjeon, Nicolas Peton, Quang Huy Tran and Sylvie Wolf

Comparison of Adaptive Non-symmetric and Three-Field FVM-BEM Coupling 337
Christoph Erath and Robert Schorr

On the Conditions for Coupling Free Flow and Porous-Medium Flow in a Finite Volume Framework 347
Thomas Fetzer, Christoph Grüninger, Bernd Flemisch and Rainer Helmig

Non-conforming Localized Model Reduction with Online Enrichment: Towards Optimal Complexity in PDE Constrained Optimization 357
Mario Ohlberger and Felix Schindler

Combining the Hybrid Mimetic Mixed Method and the Eulerian Lagrangian Localised Adjoint Method for Approximating Miscible Flows in Porous Media 367
Hanz Martin Cheng and Jérôme Droniou

Mixed Finite Volume Methods for Linear Elasticity 377
I. Ambartsumyan, E. Khattatov and I. Yotov

A Nonlinear Domain Decomposition Method to Couple Compositional Gas Liquid Darcy and Free Gas Flows 387
Nabil Birgle, Roland Masson and Laurent Trenty

Hybrid Finite-Volume/Finite-Element Schemes for $p(x)$-Laplace Thermistor Models 397
Jürgen Fuhrmann, Annegret Glitzky and Matthias Liero

Finite Volume Scheme for Coupling Two–Phase Flow with Reactive Transport in Porous Media 407
E. Ahusborde, B. Amaziane and M. El Ossmani

Nonlinear Finite-Volume Scheme for Complex Flow Processes on Corner-Point Grids 417
Martin Schneider, Dennis Gläser, Bernd Flemisch and Rainer Helmig

Consistent Nonlinear Solver for Solute Transport in Variably Saturated Porous Media 427
Daniil Svyatskiy and Konstantin Lipnikov

A Two-Dimensional Complete Flux Scheme in Local Flow Adapted Coordinates 437
Jan ten Thije Boonkkamp, Martijn Anthonissen and Ruben Kwant

hp–Adaptive Discontinuous Galerkin Methods for Porous Media Flow 447
Birane Kane, Robert Klöfkorn and Christoph Gersbacher

A Nonlinear Flux Approximation Scheme for the Viscous Burgers Equation 457
N. Kumar, J.H.M. ten Thije Boonkkamp, B. Koren and A. Linke

Mimetic Staggered Discretization of Incompressible Navier–Stokes for Barycentric Dual Mesh 467
René Beltman, Martijn J.H. Anthonissen and Barry Koren

A Reduced-Basis Approach to Two-Phase Flow in Porous Media 477
Sébastien Boyaval, Guillaume Enchéry, Riad Sanchez and Quang Huy Tran

On the Capillary Pressure in Basin Modeling 487
Laurent Quaglia

A Finite Volume Scheme for Nernst-Planck-Poisson Systems with Ion Size and Solvation Effects 497
Jürgen Fuhrmann and Clemens Guhlke

A Nonlinear Correction FV Scheme for Near-Well Regions 507
Vasiliy Kramarenko, Kirill Nikitin and Yuri Vassilevski

A Hybrid High-Order Method for the Convective Cahn–Hilliard Problem in Mixed Form 517
Florent Chave, Daniele A. Di Pietro and Fabien Marche

A Hybrid Finite Volume—Finite Element Method for Modeling Flows in Fractured Media 527
Alexey Chernyshenko, Maxim Olshahskii and Yuri Vassilevski

A Nonconforming High-Order Method for Nonlinear Poroelasticity 537
Michele Botti, Daniele A. Di Pietro and Pierre Sochala

New Criteria for Mesh Adaptation in Finite Volume Simulation of Planar Ionization Wavefront Propagation 547
Hanen Amor, Fayssal Benkhaldoun, Tarek Ghoudi, Imad Kissami and Mohammed Seaid

Erratum to: Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems E1
Clément Cancès and Pascal Omnes

Author Index 557

Part I
Hyperbolic Problems

A Weighted Splitting Approach for Low-Mach Number Flows

David Iampietro, Frédéric Daude, Pascal Galon and Jean-Marc Hérard

Abstract In steady-state regimes, water circulating in the nuclear power plants pipes behaves as a low Mach number flow. However, when steep phenomena occur, strong shock waves are produced. Herein, a fractional step approach allowing to decouple the convective from the acoustic effects is proposed. The originality is that the splitting between these two parts of the physics evolves dynamically in time according to the Mach number. The first one-dimensional explicit and implicit numerical results on a wide panel of Mach numbers show that this approach is as accurate and CPU-consuming as a state of the art Lagrange-Projection-type method.

Keywords Low Mach number flows · Fractional step · Operator splitting · Hyperbolic · Relaxation schemes

1 Introduction

Even if it is intrinsically quasi-incompressible, water flowing inside nuclear plants can generate strong shock waves through which pressure can vary by dozens of bar.

D. Iampietro (✉) · F. Daude
EDF Lab Saclay, 7 Boulevard Gaspard Monge, 92120 Palaiseau, France
e-mail: david.iampietro@edf.fr

F. Daude
e-mail: frederic.daude@edf.fr

J.-M. Hérard
EDF Lab Chatou, 6 Quai Watier, 78400 Chatou, France
e-mail: jean-marc.herard@edf.fr

D. Iampietro · J.-M. Hérard
I2M, UMR CNRS, 7373 Technople Château-Gombert 39, Rue F. Joliot Curie, 13453 Marseille, France

P. Galon
CEA DEN/DANS/DM2S/SEMT/DYN, 91190 Saclay, France
e-mail: pascal.galon@cea.fr

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_1

From a numerical point of view this diversity of behaviors raises a dilemma. Indeed, an efficient way to capture shocks in a fluid is to use exact or approximate Riemann solvers. However, stationary cases shown in [9] and theory developed in [7] prove that these latter are unable to maintain the approximated solution in the initial low-Mach phase space also called "well-prepared space". What is more, they suffer from a serious loss of accuracy in case of low-Mach number flows. Eventually, the CFL condition inherent to explicit schemes requires very demanding non-dimensional time steps bounded by the Mach number. One way to bypass these difficulties is to decouple convection from acoustic waves production by splitting the original conservation laws into two subsystems. Then, they can be successively solved and a specific low-Mach number treatment or a direct time-implicit scheme can be done on the acoustic subsystem. Such a strategy has been tested in [4] where a transport and an acoustic subsystems are exhibited, the latter being reformulated in Lagrange variables. Inspired by the pioneering work of [1], the present approach introduces a splitting weighted by a parameter related to the instantaneous flow Mach number. By doing so, it becomes sensitive to any change of Mach regime, allows to capture shocks and may be accurate in the case of low-Mach number flows.

2 A Weighted Splitting Approach

Our work focuses on the compressible Euler system whom differential structure is similar to these of two-phase homogeneous models. In one dimension, the mass, momentum and energy conservation laws read:

$$\partial_t \rho + \partial_x (\rho u) = 0, \tag{1a}$$

$$\partial_t (\rho u) + \partial_x (\rho u^2 + p) = 0, \tag{1b}$$

$$\partial_t (\rho e) + \partial_x ((\rho e + p) u) = 0. \tag{1c}$$

Here, $e = u^2/2 + \varepsilon$ is the specific total energy made of the kinetic contribution plus the specific internal energy ε related to pressure and density by the equation of state $\varepsilon = \varepsilon^{EOS}(\rho, p)$. Eventually, one can introduce c the sound speed such that $(\rho c)^2 = \left(\partial_p \varepsilon_{|\rho}\right)^{-1} \left(p - \rho^2 \partial_\rho \varepsilon_{|p}\right)$ which strongly depends on the fluid equation of state and governs the acoustic waves speed. Following [1], let us introduce $\mathscr{C}$ (respectively $\mathscr{A}$) a convective (respectively an acoustic) subsystem, namely:

$$\mathscr{C} : \begin{cases} \partial_t \rho + \partial_x (\rho u) = 0, \\ \partial_t (\rho u) + \partial_x \left(\rho u^2 + \mathscr{E}_0^2(t)\, p\right) = 0, \\ \partial_t (\rho e) + \partial_x \left((\rho e + \mathscr{E}_0^2(t)\, p) u\right) = 0. \end{cases} \quad \mathscr{A} : \begin{cases} \partial_t \rho = 0, \\ \partial_t (\rho u) + \partial_x \left((1 - \mathscr{E}_0^2(t))\, p\right) = 0, \\ \partial_t (\rho e) + \partial_x \left((1 - \mathscr{E}_0^2(t))\, p\, u\right) = 0. \end{cases}$$

Here, $\mathscr{E}_0(.)$ is a time-dependent weighting factor belonging to interval $]0,\ 1]$. It is directly related to the maximal Mach number of the flow by the expression below:

$$\mathscr{E}_0(t) = min\,(M_{max}(t),\ 1)\,;\ \text{with: } M_{max}(t) = \sup_{x\in\Omega}\left(M(x,t) = \frac{|u(x,t)|}{c(x,t)}\right), \tag{2}$$

Ω being the computational domain. One can notice that formally summing conservative subsystems $\mathscr{C}$ and $\mathscr{A}$ allows to recover the original Euler system (1). In the case of a globally low-Mach number flow, $M_{max}(t) \approx \mathscr{E}_0(t) \ll 1$, and pressure terms completely disappear from $\mathscr{C}$ which turns out to be a pure "convective" subsystem. Pressure terms are treated afterwards in $\mathscr{A}$ which becomes an "acoustic" subsystem. Actually, a low-Mach correction or a straight time-implicit resolution applied on its flux would allow to reduce the numerical diffusion or remove the most constraining part of the CFL condition. However, suppose that at instant t the flow is such that $M_{max}(t)$ jumps to 1 suddenly. Then, $\mathscr{E}_0(t)$ will be close to 1, $\mathscr{C}$ formally converges towards the full Euler system while $\mathscr{A}$ is a degenerated stationary subsystem. Hence, if $\mathscr{C}$ is solved using a time-explicit Godunov-like scheme, Euler shocks would be optimally captured. This strategy relies on the hypothesis that the waves produced by $\mathscr{C}$ and $\mathscr{A}$ are real and also that their asymptotic behavior in terms of $\mathscr{E}_0$ is the expected one. The proposition below clarifies this point (see [10]):

Proposition 1 (Hyperbolicity of convective and acoustic subsystems)
Let us introduce $c_{\mathscr{C}}(\rho,\ p)$ and $c_{\mathscr{A}}(\rho,\ p)$ two modified sound speeds such that:

$$\begin{aligned}(\rho\, c_{\mathscr{C}}(\rho,\ p))^2 &= \left(\partial_p\, \varepsilon_{|\rho}\right)^{-1}\left(\mathscr{E}_0^2\, p - \rho^2\, \partial_\rho\, \varepsilon_{|p}\right),\\ (\rho\, c_{\mathscr{A}}(\rho,\ p))^2 &= \left(\partial_p\, \varepsilon_{|\rho}\right)^{-1} p.\end{aligned} \tag{3}$$

In case of a stiffened gas thermodynamics, $c^2_{\mathscr{C}} \geq 0$. Besides, if pressure remains positive, $c^2_{\mathscr{A}} \geq 0$. Under this condition, the subsystems $\mathscr{C}$ and $\mathscr{A}$ are hyperbolic. The eigenvalues of $\mathscr{C}$ and $\mathscr{A}$ are:

$$\begin{aligned}\lambda_1^{\mathscr{C}} &= u - \mathscr{E}_0\, c_{\mathscr{C}} &&\leq \lambda_2^{\mathscr{C}} = u \leq \lambda_3^{\mathscr{C}} = u + \mathscr{E}_0\, c_{\mathscr{C}},\\ \lambda_1^{\mathscr{A}} &= -\left(1-\mathscr{E}_0^2\right) c_{\mathscr{A}} &&\leq \lambda_2^{\mathscr{A}} = 0 \leq \lambda_3^{\mathscr{A}} = \left(1-\mathscr{E}_0^2\right) c_{\mathscr{A}},\end{aligned} \tag{4}$$

the 1*-wave and* 3*-wave of both subsystems are associated to genuinely non-linear fields whereas the* 2*-wave field are linearly degenerate.*

3 Suliciu-like Relaxation Schemes To Solve $\mathscr{C}$ and $\mathscr{A}$

Relaxation schemes emerge from the theory of kinetic schemes described in [2, 3]. As shown in [5], such a method can be applied on a rather general *fluid model* endowed with a set of conservation laws, a strictly convex entropy and basic thermodynamical constraints linking state variables. Following the Suliciu-like relaxation method, also used in [4], let us introduce $\mathscr{C}^\mu$ (respectively $\mathscr{A}^\mu$) the relaxation convective (respectively the relaxation acoustic) subsystem:

$$\mathscr{C}^{\mu}:\begin{cases}\partial_t\,\rho+\partial_x\,(\rho\,u)=0,\\ \partial_t\,(\rho\,u)+\partial_x\,(\rho\,u^2)+\partial_x\,\left(\mathscr{E}_0^2(t)\,\Pi\right)=0,\\ \partial_t\,(\rho\,e)+\partial_x\,\left((\rho\,e+\mathscr{E}_0^2(t)\,\Pi)\,u\right)=0,\\ \partial_t\,(\rho\,\Pi)+\partial_x\,\left((\rho\,\Pi+a_{\mathscr{C}}^2)\,u\right)=\dfrac{\rho}{\mu}\,(p-\Pi)\,,\end{cases}$$

$$\mathscr{A}^{\mu}:\begin{cases}\partial_t\,\rho=0,\\ \partial_t\,(\rho\,u)+\partial_x\,\left((1-\mathscr{E}_0^2(t))\,\Pi\right)=0,\\ \partial_t\,(\rho\,e)+\partial_x\,\left((1-\mathscr{E}_0^2(t))\,\Pi\,u\right)=0,\\ \partial_t\,(\rho\,\Pi)+\partial_x\,\left((1-\mathscr{E}_0^2(t))\,a_{\mathscr{A}}^2\,u\right)=\dfrac{\rho}{\mu}\,(p-\Pi)\,.\end{cases}$$

Here, $\frac{\rho}{\mu}\,(p-\Pi)$ can be formally interpreted as a correction term of time scale μ forcing the relaxed pressure Π to converge towards the physical pressure instantaneously if μ tends to zero. Besides, $a_{\mathscr{C}}$ and $a_{\mathscr{A}}$ are the constant relaxation coefficients encapsulating the thermodynamical nonlinearity. What is more, under the subcaracteristic condition $a_{\mathscr{C}}>\rho\,c_{\mathscr{C}}$ (respectively $a_{\mathscr{A}}>\rho\,c_{\mathscr{A}}$), $\mathscr{C}^{\mu}$ (respectively $\mathscr{A}^{\mu}$) converges formally towards $\mathscr{C}$ (respectively $\mathscr{A}$) at order one in μ. Then, the augmented set of conservation laws is still hyperbolic and all its fields are linearly degenerate. Hence, it is possible to derive an exact Godunov solver for these relaxed subsystems. The eigenvalues of $\mathscr{C}^{\mu}$ are $u-\mathscr{E}_0\,a_{\mathscr{C}}\,\tau, u$ and $u+\mathscr{E}_0\,a_{\mathscr{C}}\,\tau$ with $\tau=1/\rho$ the specific volume. The ones of $\mathscr{A}^{\mu}$ are $-\left(1-\mathscr{E}_0^2\right)\,a_{\mathscr{A}}\,\tau$, 0, and $\left(1-\mathscr{E}_0^2\right)\,a_{\mathscr{A}}\,\tau$. The numerical flux related to $\mathscr{C}$ (respectively $\mathscr{A}$) is derived by solving a convective (respectively an acoustic) Riemann problem associated to $\mathscr{C}^{\mu}$ (respectively $\mathscr{A}^{\mu}$). In the end, for an explicit time integration, the convective flux at face $i+1/2$ and time t^n reads:

$$\mathbf{H}_{\mathbf{c}\,i+1/2}^{\,n}=\begin{cases}\frac{1}{2}\,\left(\mathbf{F}_{\mathscr{C}}\left(\mathbf{U}_i^n\right)+\mathbf{F}_{\mathscr{C}}\left(\mathbf{U}_{i+1}^n\right)\right)\\ -\frac{1}{2}\,\left|u_i^n-\mathscr{E}_0^n\,(a_{\mathscr{C}}^n)_{i+1/2}\,\tau_i^n\right|\,(\mathbf{U}_{i+1/2}^{*,\,n}-\mathbf{U}_i^n)\\ -\frac{1}{2}\,\left|(u_{\mathscr{C}}^*)_{i+1/2}^n\right|\,(\mathbf{U}_{i+1/2}^{**,\,n}-\mathbf{U}_{i+1/2}^{*,\,n})\\ -\frac{1}{2}\,\left|u_{i+1}^n+\mathscr{E}_0^n\,(a_{\mathscr{C}}^n)_{i+1/2}\,\tau_{i+1}^n\right|\,(\mathbf{U}_{i+1}^n-\mathbf{U}_{i+1/2}^{**,\,n}),\end{cases}\tag{5}$$

with $\mathbf{F}_{\mathscr{C}}\left(\underline{U}\right)=\left[\rho\,u,\ \rho\,u^2+\mathscr{E}_0^2\,p,\ (\rho\,e+\mathscr{E}_0^2\,p)\,u\right]^T$, and:

$$\mathbf{U}_{i+1/2}^{*,\,n}=\begin{bmatrix}(\rho_{i,\,\mathscr{C}}^*)^n\\ (\rho_{i,\,\mathscr{C}}^*)^n\,(u_{\mathscr{C}}^*)_{i+1/2}^n\\ (\rho_{i,\,\mathscr{C}}^*)^n\,(e_{i,\,\mathscr{C}}^*)^n\end{bmatrix},\ \mathbf{U}_{i+1/2}^{**,\,n}=\begin{bmatrix}(\rho_{i+1,\,\mathscr{C}}^*)^n\\ (\rho_{i+1,\,\mathscr{C}}^*)^n\,(u_{\mathscr{C}}^*)_{i+1/2}^n\\ (\rho_{i+1,\,\mathscr{C}}^*)^n\,(e_{i+1,\,\mathscr{C}}^*)^n\end{bmatrix},$$

$$(a_{\mathscr{C}}^n)_{i+1/2}=K\,max\left(\rho_i^n\,(c_{\mathscr{C}})_i^n,\ \rho_{i+1}^n\,(c_{\mathscr{C}})_{i+1}^n\right),\ K>1.$$

The expressions of intermediate quantities like $(u^*_{\mathscr{C}})^n_{i+1/2}$, $(\rho^*_{k,\mathscr{C}})^n$ and $(e^*_{k,\mathscr{C}})^n$, $k \in \{i,\ i+1\}$ are close to these derived in [4, 6]. More details are given in [10]. The acoustic flux is simpler because of the zero eigenvalue:

$$\mathbf{H_{ac}}^{n,\theta}_{i+1/2} = \left(1 - (\mathscr{E}_0^n)^2\right) \begin{bmatrix} 0 \\ (\Pi^*_{\mathscr{A}})^{n,\theta}_{i+1/2} \\ (\Pi^*_{\mathscr{A}})^{n,\theta}_{i+1/2}\,(u^*_{\mathscr{A}})^n_{i+1/2} \end{bmatrix}, \tag{6}$$

$$\text{with:} \begin{cases} (u^*_{\mathscr{A}})^n_{i+1/2} = \dfrac{u^n_{i+1}+u^n_i}{2} - \dfrac{1}{2\,(a_{\mathscr{A}})^n_{i+1/2}}\left(p^n_{i+1}-p^n_i\right), \\ (\Pi^*_{\mathscr{A}})^{n,\theta}_{i+1/2} = \dfrac{p^n_{i+1}+p^n_i}{2} - \dfrac{(a_{\mathscr{A}}\,\theta)^n_{i+1/2}}{2}\left(u^n_{i+1}-u^n_i\right), \\ (a^n_{\mathscr{A}})_{i+1/2} = K\ max\left(\rho^n_i\,(c_{\mathscr{A}})^n_i,\ \rho^n_{i+1}\,(c_{\mathscr{A}})^n_{i+1}\right),\ K>1. \end{cases} \tag{7}$$

Following [7], the parameter $\theta^n_{i+1/2}$ introduced in (7) is a low-Mach number correction term. Indeed, $\theta^n_{i+1/2}=1$ is the original formula with no correction, $\theta^n_{i+1/2} = \left|(u^*_{\mathscr{A}})^n_{i+1/2}\right| / \max\left(c^n_i,\ c^n_{i+1}\right)$ prevents an initial well-prepared solution from leaving the well-prepared space after one iteration. It also diminishes the numerical diffusion in the low-Mach number configurations. See [4, 7, 9] for more details. The discrete expression of the weighting parameter $\mathscr{E}_0$ follows the continuous definition written in (2):

$$\mathscr{E}_0^n = \max\left(\mathscr{E}_{inf},\ \min\left(M^n_{max},\ 1\right)\right);\ \text{with: } M^n_{max} = \underset{i\in[1,\,N_{cells}]}{max}\left(\frac{\left|u^n_i\right|}{c^n_i}\right). \tag{8}$$

Here, $0 < \mathscr{E}_{inf} \ll 1$ is only a lower bound preventing $\mathscr{E}_0^n$ from being exactly equal to zero if velocity is initially null everywhere. Finally, the overall dynamical fractional step approach can be summed up in the following equations:

$$\begin{aligned} \mathscr{C}:&\begin{cases} \mathbf{U}^{n+}_i = \mathbf{U}^n_i - \frac{\Delta t}{\Delta x}\left(\mathbf{H_c}_{i+1/2}\left(\mathbf{U}^n_i,\ \mathbf{U}^n_{i+1}\right) - \mathbf{H_c}_{i-1/2}\left(\mathbf{U}^n_{i-1},\ \mathbf{U}^n_i\right)\right), \\ \Pi^{n+}_i = p^{EOS}\left(\mathbf{U}^{n+}_i\right) = p^{n+}_i, \end{cases} \\ \mathscr{A}:&\begin{cases} \mathbf{U}^{n+1}_i = \mathbf{U}^{n+}_i - \frac{\Delta t}{\Delta x}\left(\mathbf{H_{ac}}_{i+1/2}\left(\mathbf{U}^{n+}_i,\ \mathbf{U}^{n+}_{i+1}\right) - \mathbf{H_{ac}}_{i-1/2}\left(\mathbf{U}^{n+}_{i-1},\ \mathbf{U}^{n+}_i\right)\right), \\ \Pi^{n+1}_i = p^{EOS}\left(\mathbf{U}^{n+1}_i\right) = p^{n+1}_i. \end{cases} \end{aligned} \tag{9}$$

We also have the following results (see [10] for a proof):

Proposition 2 (Conservativity, Positivity, Low-Mach Accuracy)

- *Conservativity: The overall scheme (9) is conservative.*
- *Positivity: Assume* $\forall i:\ \rho^n_i > 0,\ \varepsilon^n_i > 0$. *Then,* $\rho^{n+1}_i > 0,\ \varepsilon^{n+1}_i > 0$ *is ensured under modified subcaracteristic conditions:*

$$(a^n_{\mathscr{C}})_{i+1/2} = K\, max\left(\rho^n_i\,(c_{\mathscr{C}})^n_i,\ \rho^n_{i+1}\,(c_{\mathscr{C}})^n_{i+1},\ a_i^{\rho;\varepsilon,n},\ a_{i+1}^{\rho;\varepsilon,n}\right) \tag{10a}$$

$$(a^n_{\mathscr{A}})_{i+1/2} = K\, max\left(\rho^n_i\,(c_{\mathscr{A}})^n_i,\ \rho^n_{i+1}\,(c_{\mathscr{A}})^n_{i+1},\ a_i^{\rho;\varepsilon,n},\ a_{i+1}^{\rho;\varepsilon,n}\right) \tag{10b}$$

where the non-dimensional expressions associated with $a_i^{\rho;\varepsilon,n}$ *and* $a_{i+1}^{\rho;\varepsilon,n}$ *are of order* $O(1)$*; and under a global CFL condition:* $\Delta t^n = min\left(\Delta t^n_E,\ \Delta t^n_{\mathscr{C}},\ \Delta t^n_{\mathscr{A}}\right)$*, with* Δt^n_E *(respectively* $\Delta t^n_{\mathscr{C}}$*,* $\Delta t^n_{\mathscr{A}}$*) the timestep bounded by the Euler (respectively* $\mathscr{C}$*,* $\mathscr{A}$*) CFL condition guaranteeing no interaction between waves produced by the face Riemann problems.*

- *Low-Mach accuracy: Assume that the initial conditions belong to the well-prepared space (see [7] for a definition) and that* $\mathscr{E}^n_0$ *is given by (8). Then, the non-dimensional numerical diffusion of a smooth solution computed thanks to the scheme is a* $O(\Delta x)$ *instead of* $O(\Delta x/M)$ *if the above global CFL condition holds, and if the discrete low-Mach correction* $\theta^n_{i+1/2}$ *is triggered.*

4 Numerical Results

We perform a one-dimensional Sod-type shock tube. The fluid is endowed of an ideal gas thermodynamics with $\gamma = 7/5$. The initial data are: $\rho^0_L = 1\ kg.m^{-3}$, $u^0_L = 0\ m.s^{-1}$, $p_{0,L} = p_{0,R}\ (1+\varepsilon)$, $\rho^0_R = 0.125\ kg.m^{-3}$, $u^0_R = 0\ m.s^{-1}$, $p_{0,R} = 0.1\ bar$. By tuning ε, the maximal flow Mach number can be modified. Figure 1 shows the pressure convergence curves for three different Mach values: $M = 0.92$, $M = 9.5 \times 10^{-2}$ and $M = 4.2 \times 10^{-3}$. The cells number varies from 10^2 to 9×10^4. Five different schemes have been tested: "no-Sp" corresponds to the case where $\mathscr{E}^n_0 = 1$ is imposed along the simulation. Thus, the splitting is not triggered. "Sp-$(\sqrt{M})$" is the weighted splitting approach with $\mathscr{E}^n_0 = \max\left(\mathscr{E}_{inf},\ \min\left(\sqrt{M^n_{max}},\ 1\right)\right)$ while "Sp-(M)" involves $\mathscr{E}^n_0$ defined in formula (8) which is optimal, because, as proven in [10], it minimizes the numerical diffusion of the subsystem $\mathscr{C}$ in the low-Mach number case. Eventually, "LP" is the Lagrange Projection splitting method, fully described in [4] and taken as a benchmark. "-corr" means that the low-Mach correction is triggered.

One can notice that, when $M \approx 1$, all the schemes are equivalent in terms of accuracy. In the sequel, as the Mach number decreases the low-Mach corrected schemes become the most accurate ones. Particularly, Sp-$(\sqrt{M})$-corr seems to be more precise than Sp-LP-corr at $M = 4.2 \times 10^{-3}$. However, one should notice that for *every* schemes, the order of convergence is depreciated as the Mach number decreases. Indeed for pressure, it passes from 0.87 at $M = 0.92$ (the expected order already obtained in [8]) to 0.82 at $M = 9.5 \times 10^{-2}$ and 0.56 in the low-Mach case.

Velocity profiles are plotted on Fig. 2. It seems that the accuracy of Sp-(M) is higher than these of Sp-$(\sqrt{M})$ through the left rarefaction wave where the exact solution is continuous. Besides, the low-Mach correction applied on the weighted

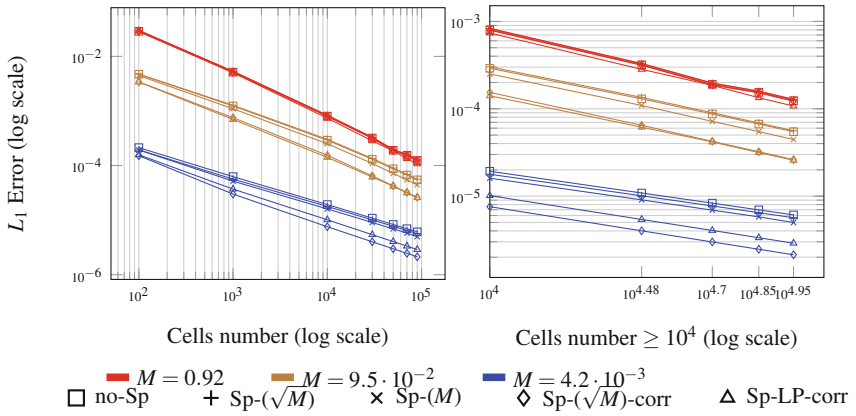


Fig. 1 Pressure convergence curves: explicit schemes

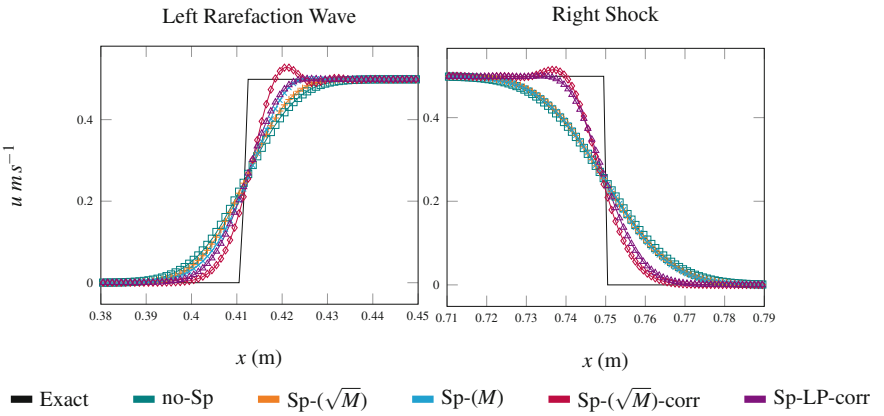


Fig. 2 Velocity profiles: $M = 4.2 \times 10^{-3}$, $N_{cells} = 10^3$

splitting approach results in small overshoots located in the tail of the left rarefaction wave and before the shock front.

Instead of using the low-Mach correction, one can directly apply an implicit approximation of the acoustic flux (6) using a method that relies on strong relaxed Riemann Invariants (see [4, 6] for more details). Pressure convergence curves for different implicit schemes at $M = 4.2 \times 10^{-3}$ are shown on Fig. 3. Here, $\forall k \in \{1/2,\ 5,\ 20\}$, the mention "-cfl[$k$]" indicates that the Courant number involved in the determination of Δt_E^n and $\Delta t_{\mathscr{A}}^n$ defined in Proposition 2 is equal to "k". As expected, the implicit techniques are more diffusive than explicit schemes. Besides, at a given mesh, CPU time diminishes considerably as the Courant number increases. For example, at $N_{cells} = 10^3$, Sp-(M)-Imp-RI-cfl0.5 takes 10.1 s whereas Sp-(M)-Imp-RI-cfl5 (respectively Sp-(M)-Imp-RI-cfl20) requires 1.9 s (respectively 0.7 s). Finally, one can notice that the present implicit weighted splitting approach is as

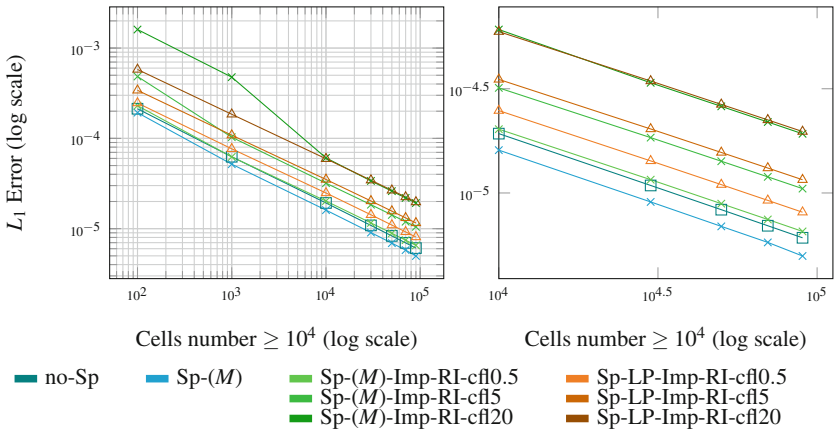


Fig. 3 Pressure convergence curves: implicit schemes, $M = 4.2 \cdot 10^{-3}$

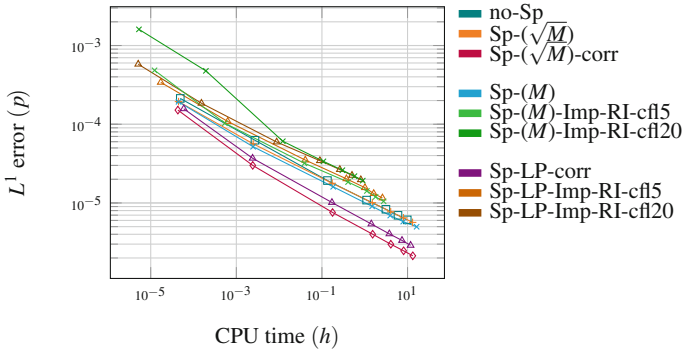


Fig. 4 Pressure efficiency curve: $M = 4.2 \times 10^{-3}$

accurate as the implicit Lagrange-Projection method. In the low-Mach regime, the trade-off between explicit-accuracy versus the implicit-CPU-rapidity is solved thanks to the efficiency curve plotted on Fig. 4. At a given precision, for low-Mach unsteady cases, explicit schemes are still less CPU-consuming than implicit techniques.

Acknowledgements D. Iampietro received a financial support by ANRT through an EDF-CIFRE contract 2015/0561. Numerical facilities were provided by EDF.

References

1. Baraille, R., Bourdin, G., Dubois, F., Le Roux, A.Y.: Une version à pas fractionnaires du schéma de Godunov pour l'hydrodynamique. C. R. de l'Académie des Sci. **314**, 147–152 (1992)
2. Bouchut, F.: Entropy satisfying flux vector splittings and kinetic BGK models. Numer. Math. **94**, 623–672 (2003)
3. Bouchut, F.: Nonlinear Stability of Finite Vol. Methods for Hyperbolic Conserv.Laws. Birkäser (2004)
4. Chalons, C., Girardin, M., Kokh, S.: An all-regime Lagrange-Projection like scheme for the gas dynamics equations on unstructured meshes. Commun. Comput. Phys. **20**, 188–233 (2016)
5. Coquel, F., Godlewski, E., Seguin, N.: Relaxation of fluid systems. Math. Models Methods Appl. Sci. **22**, 43–95 (2012)
6. Coquel, F., Nguyen, Q.L., Postel, M., Tran, Q.H.: Entropy-satisfying relaxation method with large time-steps for Euler IBVPS. Math. Comput. **79**, 1493–1533 (2010)
7. Dellacherie, S., Omnes, P., Jung, J., Raviart, P.: Construction of modified Godunov type schemes accurate at any Mach number for the compressible Euler system. Math. Models Methods Appl. Science **26**, 2525–2615 (2016)
8. Gallouët, T., Hérard, J.M., Seguin, N.: Some recent finite volume schemes to compute Euler equations using real gas EOS. Intern.J. Numer. Methods Fluids **39**, 1073–1138 (2002)
9. Guillard, H., Murrone, A.: On the behavior of upwind schemes in the low Mach number limit: II Godunov type schemes. Comput. Fluids **33**, 655–675 (2004)
10. Iampietro, D., Daude, F., Galon, P., Hérard, J.M.: A Mach-sensitive splitting approach for Euler-like systems (2017). https://hal.archives-ouvertes.fr/hal-01466827

Weno Scheme for Transport Equation on Unstructured Grids with a DDFV Approach

Florence Hubert and Rémi Tesson

Abstract In this paper we develop a DDFV approach for WENO scheme on unstructred grids for 2D transport equations. An order 2 scheme is presented using the DDFV diamond structure to define the different stencils. Numerical tests illustrate the accuracy and robustness of the method.

Keywords Weighted essentially non-oscillatory · Transport equation · Discrete duality finite volume scheme

MSC (2010): 65M08 · 65Z12 · 65D05

1 Introduction

The problems we are interested in are fluid-structure interaction problems in 2D, where we use a level-set approach. In such problems, we look at the behavior and displacement of a structure, that can be a solid or an elastic membrane, inside a fluid. The level-set approach consists, in this situation, in representing the interface between the fluid and the structure implicitly as the level-set of a function ϕ. The modelization of this situations often implies fluid mechanics equations, such as Stokes equations, coupled with transport equations.

Here we focus on numerical tools for the resolution of transport equations. Because the level-set function ϕ that captures the interface is the solution of a transport equation we want to be very sharp when solving this equation. Order one schemes are known to be very diffusive and inadapted in the level-set context. High order method, like WENO schemes, appear to be a good solution to solve precisely trans-

F. Hubert (✉) · R. Tesson (✉)
I2M, Marseille, 39 Rue F. Joliot Curie, 13453 Marseille, France
e-mail: florence.hubert@univ-amu.fr

R. Tesson
e-mail: remi.tesson@univ-amu.fr

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_2

port equation. First introduce by Harten, Osher and others [6–8, 13], WENO schemes are known to be efficient on convection problems.

The interest to use locally refined meshes is that it allows us to be accurate near the interface between fluid and structure although to be efficient in term of computational time and memory. In this paper, we will develop a DDFV approach for WENO scheme on locally refined grids. The Discrete Duality Finite Volume method (DDFV) is a Finite Volume method, that has been successfully used to solve Stokes equations [11] on various kind of meshes, including locally refined meshes. In Sect. 2 we will present the time and spatial discretization of the transport equation, in Sect. 3 we will expose the reconstruction procedure used in the WENO scheme in itself and then we will illustrate our statement with numerical tests in Sect. 4.

2 Discretization of the Transport Equation

2.1 Notations and DDFV Structure

In fluid-structure interaction problems, the velocity used in the transport equation is often given by fluid mechanics equations like Stokes equations. In such models we must couple the resolution of Stokes equations with the resolution of transport equation. In order to be able to deal with a large class of meshes and to release us from the orthogonality constraint imposed by VF4 methods (see [4]), we choose to use a DDFV strategy.

DDFV are Finite Volume methods introduced first in [3, 9]. They consist in a decomposition of the computing domain in a set of polygons. Those polygons form the primal mesh and one unknown is associated to the barycenter of each polygon. Then other unknowns are added on the vertices of the polygons. Those vertices are therefore seen as centers of other polygons that define a dual mesh as in Fig. 1. The interest of introducing new unknowns is that it allows us to compute an approximation of the gradient in every directions.

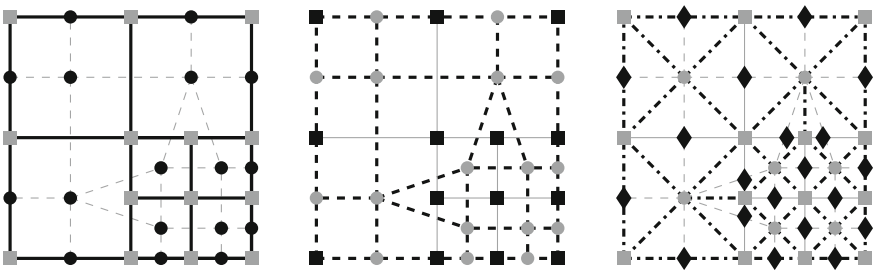


Fig. 1 DDFV structure. From the *left* to the *right* primal mesh, dual mesh and diamond structure

We denote by K a polygon of the primal mesh $\mathfrak{M}$ and K^* a polygon of the dual mesh $\mathfrak{M}^*$. In the following, we will denote an element of the primal or dual mesh by $C \in \mathfrak{M} \cup \mathfrak{M}^*$.

Another mesh, the diamond mesh, can be associated to the DDFV structure. This third mesh is very convenient when we have to implement the DDFV method because it is a link beetween primal and dual meshes in which they both play a symmetric role. In particular, this is on the diamond mesh that we define the discretized gradient. To create the diamond mesh, we construct quadrangle associated to each edge of the primal and dual mesh like in Fig. 1.

2.2 Time Discretization

The transport equation on a bounded open set $\Omega \subset \mathbb{R}^2$, with a divergence-free velocity u, can be written as:

$$\frac{d\phi}{dt} = -div(\phi u) := \mathscr{L}(\phi) \tag{1}$$

For the time discretization, we follow [5] and use a TVD Runge-Kutta of order k. The order k is then chosen to be in adequation with the order of the spatial discretization, that means here $k = 2$. Let Δt be the the time step of the method, we will denote by ϕ^n the approximation of function ϕ at time $t_n = n\Delta t$. The RK2 scheme is then given by the following steps:

$$\phi^{n,1} := \phi^n + \Delta t \mathscr{L}\left(\phi^n\right), \quad \phi^{n+1} := \frac{1}{2}\phi^n + \frac{1}{2}\phi^{n,1} + \frac{1}{2}\Delta t \mathscr{L}\left(\phi^{n,1}\right) \tag{2}$$

We will now focus on the space discretization of operator $\mathscr{L}$ by a WENO method.

2.3 Discretization of Operator $div(\phi U)$

Let $\phi^\tau = (\phi_C)_{C \in \mathfrak{M} \cup \mathfrak{M}^*}$, the vector of the approximations ϕ_C of the mean values $\bar{\phi}_C = \frac{1}{|C|}\int_C \phi$ of function ϕ on the cells $C \in \mathfrak{M} \cup \mathfrak{M}^*$ that we want to compute.

Following the Finite Volume strategy, we integrate the operator $\mathscr{L}$ on each cell:

$$\frac{1}{|C|}\int_C \mathscr{L}(\phi) = \frac{1}{|C|}\int_{\partial C} \phi u.n \tag{3}$$

where n is the outer unit normal to the boundary ∂C of C.

Because the cells are polygonal, we can rewrite the boundary integral as a sum over the edges:

$$\int_{\partial C} \phi u.n = \sum_{\sigma \subset \partial C} \int_\sigma \phi u.n_\sigma \tag{4}$$

The line integral of the right member can be approximated using a p point Gaussian quadrature. Taking $p = 1$ allows us to find back the classical DDFV formulation of divergence operator. Of course the same work can be done for $p > 1$

$$\int_{\partial C} \phi u.n \approx \sum_{\sigma \subset \partial C} |\sigma| \phi(x_\sigma) u(x_\sigma).n_\sigma \text{ where } x_\sigma \text{ is the middle of } \sigma \tag{5}$$

The WENO scheme consists in approximating for each cell C and each edge σ, the value $\phi(x_\sigma)$ by a convex combination of the value in x_σ of several polynomials whose mean values coincide with the mean values of ϕ on a set of selected cells. This set of cells is called the stencil of the method. The WENO procedure for polynomial reconstructions will be developed in Sect. 3. For the moment, let us assume that we dispose of such an approximation $\phi_{C,\sigma}$. Then we define the spatial discretization $\mathscr{L}_C$ of operator $\mathscr{L}$ using an upwind flux as:

$$\begin{aligned} \mathscr{L}_C(\phi^\tau) := &- \sum_{\sigma = C \cap \tilde{C}} |\sigma| \left[\phi_{C,\sigma}(u(x_\sigma).n_\sigma)^+ - \phi_{\tilde{C},\sigma}(u(x_\sigma).n_\sigma)^- \right] \\ &- \sum_{\sigma \in C \cap \partial\Omega} |\sigma| \left[\phi_{C,\sigma}(u(x_\sigma).n_\sigma)^+ - \phi_b(x_\sigma)(u(x_\sigma).n_\sigma)^- \right] \end{aligned} \tag{6}$$

where $\tilde{C}$ and C share the edge σ and ϕ_b prescribed through the boundary condition.

Let define $\phi^{n,\tau} = \left(\phi_C^n\right)_{C \in \mathfrak{M} \cup \mathfrak{M}^*}$ the vector of the approximation ϕ_C^n of the mean value of ϕ on the cells C at time t_n. The full discretization is then given by:

$$\phi_C^{n+1} := \phi_C^n + \frac{1}{2} \Delta t \left[\mathscr{L}(\phi_C^n) + \mathscr{L}\left(\phi_C^n + \Delta t \mathscr{L}(\phi_C^n)\right) \right] \tag{7}$$

The previous work is done in the same way on both primal mesh and dual mesh. If we take a look at Eq. (6), we can see that each cell is only linked with its neighbours. One can then think that primal and dual meshes are totally decoupled. In fact the coupling beetween the two meshes will be ensured by the reconstruction process as we are going to see in the next section and depends on the degree of the polynomial approximation.

3 Reconstruction Procedure

3.1 Problem Statement

Given a cell C and an edge σ, we want to reconstruct an approximation of $\phi(x_\sigma)$ though we only know the mean values $(\bar{\phi}_C)$ of ϕ on $\mathfrak{M} \cup \mathfrak{M}^*$. Following the WENO strategy, the approximation $\phi_{C,\sigma}$ is computed as a convex combination of several polynomial interpolations of ϕ.

To find those polynomial interpolations, we fix a subset $S \subset \mathfrak{M} \cup \mathfrak{M}^*$, depending on C and σ, and we choose the polynomial $P_S[\phi]$ among the polynomials of degree k as the solution of the following problem:

$$\frac{1}{|C|}\int_C P_S[\phi] = \bar{\phi}_C, \quad \forall C \in S \tag{8}$$

The degree k of polynomial P_S is fixed arbitrary and impacts the size of the stencil S.

For high degree k, interpolation often leads to oscillating polynomials. That is the reason why we compute a convex combination of several different interpolations of ϕ. The weights in the convex combination are choosen in order to favour non-oscillating polynomials

$$\phi_{C,\sigma} = \sum_S a_S P_S[\phi](x_\sigma) \tag{9}$$

In this paper, we choose to focus on the oscillating criterion proposed by Abgrall in [1] but other criterion and weights can be found in [5, 10]:

$$a_S = \frac{(\varepsilon + c_0(P_S[\phi]))^{-4}}{\sum_T (\varepsilon + c_0(P_T[\phi]))^{-4}}, \text{ with } c_0(P) = \sum_{|\alpha|=m} |p_\alpha| \text{ for } P = \sum_{|\alpha|\le m} p_\alpha X^\alpha \tag{10}$$

3.2 Polynomial Interpolation Procedure

Let us consider a stencil $S = \{C_0, ..., C_l\}$ and $(\bar{\phi}_{C_i})_{i=1..l}$ the mean values of ϕ on the cells C_i. We want to find a polynomial P_S that depends on the stencil S and such as: $\langle P_S \rangle_C = \bar{\phi}_C$, for each $C \in S$. With idea of computing an approximation P_S on the stencil S and to avoid spatial dependency, we use a barycentric representation with respect to a given cell C_0:

$$P_S = \sum_{|\alpha|\le n} p_\alpha (X - x_{C_0})^\alpha$$

where x_{C_0} is the barycenter of cell $C_0 \subset S$. When we rewrite the previous equations on P_S in an extended form

$$\sum_{|\alpha|\le n} p_\alpha \langle (X - x_{C_0})^\alpha \rangle_C = \phi_C, \quad \text{for each } C \in S$$

we can easily see that we have to solve a linear problem $\mathscr{A}X = b$, with $\mathscr{A}_{K,\alpha} = \langle (X - x_{C_0})^\alpha \rangle_C$, $X = (p_\alpha)_\alpha$ and $b = (\bar{\phi}_C)_{C \subset S}$. If the matrix $\mathscr{A}$ is invertible, the stencil S is called admissible. In practice, we don't have access to an easy way to know if a stencil is admissible. When a stencil is not admissible, then we have to change

the stencil, test again if it is admissible and repeat those operations until we find an admissible one.

3.3 Stencil Choice

The choice of the stencil is a crucial point in the construction of the scheme. Stencils will be composed of both primal and dual cells. As in classical WENO scheme stencils have to be centered in the smooth regions and one-sided near the shocks.

The usual strategy is to associate a given number of stencil to each cell, and then to evaluate the corresponding polynomials on each edges. Here, because the natural structure to use in DDFV schemes is the diamonds structure, we define stencils from this structure. We associate stencils to each couple (C, σ) and each couple is associated to an unique diamond $\mathscr{D}$, see Fig. 2 (left). Let us define $\mathscr{V}(\mathscr{D}) = \{\mathscr{D}' / \text{ such that } \mathscr{D} \cap \mathscr{D}' \neq \emptyset\}$. In order to construct the stencils, we will use the unknowns provided by $\mathscr{D}$ but also by $\mathscr{V}(\mathscr{D})$ and $\mathscr{V}(\mathscr{V}(\mathscr{D}))$. This choice allows us to have access to enough unknowns on the boundary and to construct centered as one-sided stencils.

Then we will construct the stencils associated to (C, σ) as follows. First, we set C as the first cell of the stencil. Then we will choose randomly the other ones (if needed) in $\mathscr{D} \cup \mathscr{V}(\mathscr{D}) \cup \mathscr{V}(\mathscr{V}(\mathscr{D}))$.

Because diamonds are quadrangles and two neighbours share an edge, the number $\mathscr{N}_u$ of potential unknowns is then given by $\mathscr{N}_u \leq 4 + 4 \times 2 + 3 \times 4 \times 2 = 36$.

For a reconstruction of order 0, we only need one point in the stencil and so we only have one potential stencil. One can easily see that in that case we find back the classical upwind scheme and both primal and dual meshes are totally decoupled. For a reconstruction of order greater than 0, primal and dual mesh are coupled. For example in the case of order 2, we have at most $\mathscr{N}_S = \binom{36}{5}$ different stencils. We can however mention that in practice, the maximal number of potential unknowns is not achieved (see Fig. 2 for examples).

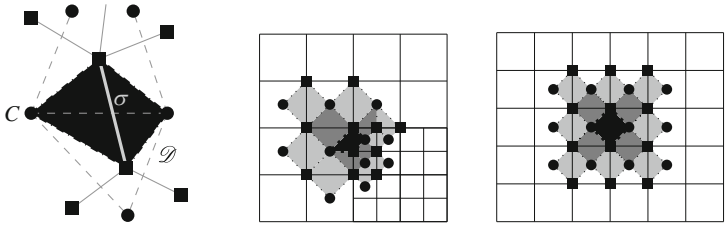


Fig. 2 Diamond cells. $\mathscr{D}$ is in *black*, $\mathscr{V}(\mathscr{D})$ in *gray* and $\mathscr{V}(\mathscr{V}(\mathscr{D}))$ in *light gray*

4 Numerical Tests

In all the following tests, we use a locally refined mesh like in Fig. 1. The previous WENO scheme is implemented in each case with 15 stencils.

4.1 Sinus Translation

First, we test our WENO scheme on the equation

$$\frac{\partial \phi}{\partial t} + \frac{\partial \phi}{\partial x} + \frac{\partial \phi}{\partial y} = 0, \quad (x, y) \in [-2; 2] \times [-2; 2]$$

with the initial condition $\phi_0(x, y) = \sin(\frac{\pi}{2}(x + y))$. One can refer to [10] for comparison of the results. Tests are done with a time step equal to $\Delta t = 0.01$ and in each case we compute the L_1 error at time $t = 2$.The results for the error and the order of the scheme are presented in Table 1 (mesh size refers for the minimal size of the square cells). We obtain an order 2 for the method, which is in adequation with the degree of the polynomial reconstruction.

4.2 Solid Body Rotation (SBR)

Solid body rotation is a classical test used in the literature for advection equation. Zalesak proposed in [14] the rotation of a slotted cylinder. The width of the slot as well as the "bridge" connecting the two half must be about 5 cells. Here, we choose an adaptation of this test introduced in [12] and used in [2]. It consists in the rotation of three body shapes, a hump, a cone and a the slotted cylinder of Zalesak. The overvalue of the initial condition is given on Fig. 3 (left). We choose $\Delta t = 0.005$ and a mesh size $h = 1/128$. As it is mention in [2], a way to measure the accuracy of the scheme is to count the number of isolines outside of the slot. Figure 3, show the isolines at $t = 2\pi$. We can see here that all the isolines fit the slot. Results at $t = \pi$ in Fig. 3 are here to point the fact that all three shapes really pass through the refined part of the mesh.

Table 1 L_1 error for sinus translation

Mesh size	$1.25.10^{-1}$	$6.25.10^{-2}$	$3.125.10^{-2}$	$1.5625.10^{-2}$
Error L_1	$5.699.10^{-1}$	$1.448.10^{-1}$	$3.363.10^{-2}$	$7.884.10^{-3}$
Order	–	1.98	2.10	2.09

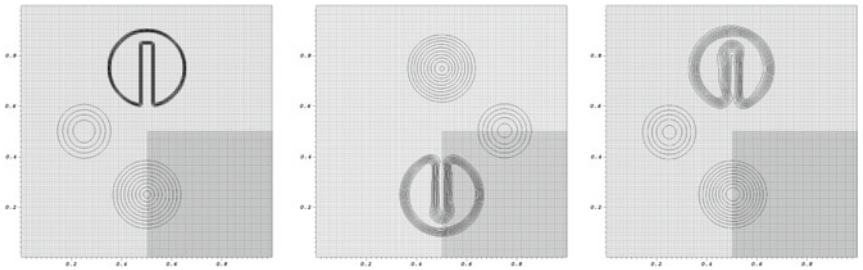


Fig. 3 Isovalues from 0.1 to 0.9 for SBR. From the *left* to the *right* Isovalues at $t = 0, \pi, 2\pi$

5 Conclusion

We presented in this paper a DDFV approach for WENO scheme working on any structured and unstructred grids. We exhibited the expected order 2 of the scheme on smooth test case. The experiment on the SBR test seems also very promising. This approach will have many applications in moving domains on adaptative meshes.

References

1. Abgrall, R.: On essentially non-oscillatory schemes on unstructured meshes: analysis and implementation. J. Comput. Phys. **114**(1), 45–58 (1994)
2. Clain, S., Diot, S., Loubère, R.: A high-order finite volume method for hyperbolic systems: Multi-dimensional Optimal Order Detection (MOOD). J. Comput. Phys. pp. 0–0 (2011)
3. Domelevo, Komla, Omnes, Pascal: A finite volume method for the laplace equation on almost arbitrary two-dimensional grids. ESAIM: Math. Model. Numer. Anal. (Modlisation Mathmatique et Analyse Numrique) **39**(6), 1203–1249 (2005)
4. Eymard, R., Gallout, T., Herbin, R.: Finite volume methods. In: Solution of Equation in n (Part 3), Techniques of Scientific Computing (Part 3), *Handbook of Numerical Analysis*, vol. 7, pp. 713–1018. Elsevier (2000). http://dx.doi.org/10.1016/S1570-8659(00)07005-8. http://www.sciencedirect.com/science/article/pii/S1570865900070058
5. Friedrich, O.: Weighted essentially non-oscillatory schemes for the interpolation of mean values on unstructured grids. J. Comput. Phys. **144**(1), 194–212 (1998)
6. Harten, A., Engquist, B., Osher, S., Chakravarthy, S.R.: Uniformly high order accurate essentially non-oscillatory schemes, iii. J. Comput. Phys. **131**(1), 3–47 (1997)
7. Harten, A., Osher, S.: Uniformly high-order accurate nonoscillatory schemes. I. SIAM J. Numer. Anal. **24**(2), 279–309 (1987). doi:10.1137/0724022
8. Harten, A., Osher, S., Engquist, B., Chakravarthy, S.R.: Some results on uniformly high-order accurate essentially nonoscillatory schemes. Appl. Numer. Math. **2**(3), 347–377 (1986)
9. Hermeline, F.: A finite volume method for the approximation of diffusion operators on distorted meshes. J. Comput. Phys. **160**(2), 481–499 (2000)
10. Hu, C., Shu, C.W.: Weighted essentially non-oscillatory schemes on triangular meshes. J. Comput. Phys. **150**(1), 97–127 (1999)
11. Krell, S.: Schémas volumes finis en mécanique des fluides complexes. Ph.D. thesis, Aix-Marseille Université (2010)
12. LeVeque, R.J.: High-resolution conservative algorithms for advection in incompressible flow. SIAM J. Numer. Anal. **33**(2), 627–665 (1996)

13. Shu, C.W., Osher, S.: Efficient implementation of essentially non-oscillatory shock-capturing schemes. J. Comput. Phys. **77**(2), 439–471 (1988)
14. Zalesak, S.T.: Fully multidimensional flux-corrected transport algorithms for fluids. J. Comput. Phys. **31**(3), 335–362 (1979)

New Types of Jacobian-Free Approximate Riemann Solvers for Hyperbolic Systems

Manuel J. Castro, José M. Gallardo and Antonio Marquina

Abstract We present recent advances in PVM (Polynomial Viscosity Matrix) methods based on internal approximations to the absolute value function. These solvers only require a bound on the maximum wave speed, so no spectral decomposition is needed. Moreover, they can be written in Jacobian-free form, in which only evaluations of the physical flux are used. This is particularly interesting when considering systems with complex Jacobians, as the relativistic magnetohydrodynamics (RMHD) equations. The proposed solvers have also been extended to the case of approximate DOT (Dumbser-Osher-Toro) methods, which can be regarded as simple and efficient approximations to the classical Osher-Solomon method. Some numerical experiments involving the RMHD equations are presented. The obtained results are in good agreement with those found in the literature and show that our schemes are robust and accurate. Finally, notice that although this work focuses on RMHD, the proposed schemes can be directly applied to general hyperbolic systems.

Keywords Hyperbolic systems · Incomplete riemann solvers · Osher-solomon method · Relativistic magnetohydrodynamics

MSC (2010): 65M08 · 35L65

M.J. Castro · J.M. Gallardo (✉)
Facultad de Ciencias, Department Análisis Matemático,
Estadística e I.O., y Matemática Aplicada, Universidad de Málaga,
Campus de Teatinos s/n, 29080 Málaga, Spain
e-mail: jmgallardo@uma.es

M.J. Castro
e-mail: mjcastro@uma.es

A. Marquina
Department Matemáticas, Universidad de Valencia, Avda. Dr. Moliner 50,
46100 Burjassot, Valencia, Spain
e-mail: marquina@uv.es

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_3

1 Preliminaries

Consider a hyperbolic system of conservation laws

$$\partial_t w + \partial_x F(w) = 0, \tag{1}$$

where $w(x,t)$ takes values on an open convex set $\mathscr{O} \subset \mathbb{R}^N$ and $F\colon \mathscr{O} \to \mathbb{R}^N$ is a smooth flux function. We are interested in the numerical solution of (1) using finite volume methods of the form

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}(F_{i+1/2} - F_{i-1/2}), \tag{2}$$

where w_i^n denotes the approximation to the average of the exact solution on the cell $I_i = [x_{i-1/2}, x_{i+1/2}]$ at time $t^n = n\Delta t$ (unless necessary, dependence on time will be dropped). The numerical flux is assumed to have the form

$$F_{i+1/2} = \frac{F(w_i) + F(w_{i+1})}{2} - \frac{1}{2} Q_{i+1/2}(w_{i+1} - w_i), \tag{3}$$

where $Q_{i+1/2}$ denotes the numerical *viscosity matrix*, which determines the numerical diffusion of the scheme.

It is worth noticing that Roe's method [10] can be written in the form (3) with viscosity matrix $Q_{i+1/2} = |A_{i+1/2}|$, where $A_{i+1/2}$ is a Roe matrix for the system. This remark has originated several numerical methods in the literature (see, e.g., [5, 6, 11] and the references therein), for which the corresponding viscosity matrix consists of some approximation to the absolute value matrix $|A_{i+1/2}|$.

2 Some Comments on PVM Riemann Solvers

Polynomial Viscosity Matrix (PVM) Riemann solvers were introduced in [2]. They are based on the idea of approximating the absolute value of the Roe matrix $A_{i+1/2}$ by means of a suitable polynomial evaluation of such matrix. If $P(x)$ is some polynomial approximation of $|x|$ in the interval $[-1, 1]$, and $\lambda_{i+1/2,\max}$ is the eigenvalue of $A_{i+1/2}$ with maximum modulus (or an upper bound of it), the numerical flux of the PVM method associated to $P(x)$ is given by (3) with viscosity matrix

$$Q_{i+1/2} = |\lambda_{i+1/2,\max}| P(|\lambda_{i+1/2,\max}|^{-1} A_{i+1/2}),$$

which provides an approximation to $|A_{i+1/2}|$, the viscosity matrix of Roe's method. It is important to notice that no spectral decomposition of the matrix $A_{i+1/2}$ is needed to build a PVM method, but only a bound on its spectral radius. This feature makes

PVM methods greatly efficient and applicable to systems in which the eigenstructure is not known or difficult to obtain.

A number of well-known schemes in the literature can be viewed as particular cases of PVM methods: Lax-Friedrichs, Rusanov, HLL, FORCE, Roe, etc. In the cases in which a Roe matrix is not available or is difficult to compute, $A_{i+1/2}$ can be taken as the Jacobian matrix of the system evaluated at some average state.

The stability of a PVM scheme strongly depends on the properties of the basis polynomial $P(x)$. In particular, it must verify the *stability condition*

$$|x| \leq P(x) \leq 1, \quad \forall\, x \in [-1, 1]. \tag{4}$$

Of course, a standard CFL restriction has also to be imposed.

The technique for constructing PVM methods has been further extended in [3] to the case of rational functions, which has originated new families of very precise incomplete Riemann solvers. Moreover, in [4] the authors have introduced the so-called *approximate DOT* (Dumbser-Osher-Toro) solvers, which combine the technique of PVM methods with the universal Osher-type solvers proposed in [7]. These methods can be viewed as simple and efficient approximations to the classical Osher-Solomon method [9], sharing most of its interesting features and being applicable to general hyperbolic systems, unlike the original Osher-Solomon method.

With respect to the choice of the basis function, in [3, 4] Chebyshev polynomials (which provide optimal uniform approximations to $|x|$) were considered. An advantage of these methods is that they can be implemented in a recursive way using only vector operations. On the other hand, as Chebyshev approximations cross the origin, PVM-Chebyshev methods need an entropy fix to handle sonic flow correctly. Furthermore, Chebyshev functions do not satisfy the stability condition (4) strictly, which may cause the scheme to be unstable under certain conditions (see [3]). For these reasons, it would be interesting to consider another family of polynomial approximations to $|x|$ fixing the above mentioned problems. Such a family of internal approximations can be iteratively constructed as follows:

$$p_0(x) \equiv 1, \quad p_{n+1}(x) = \frac{1}{2}\big(2p_n(x) - p_n(x)^2 + x^2\big), \; n = 0, 1, 2, \ldots \tag{5}$$

For instance, the polynomial used in the numerical tests in Sect. 4 is given by

$$p_3(x) = -\tfrac{1}{128}x^8 + \tfrac{3}{32}x^6 - \tfrac{23}{64}x^4 + \tfrac{31}{32}x^2 + \tfrac{39}{128}.$$

3 Jacobian-Free Implementation

In this section we build Jacobian-free PVM solvers associated to the internal approximation $p_n(x)$ introduced in the previous section. First of all, it should be noted that the recursive form (5) is not suitable for that purpose due to the term $p_n(x)^2$. For this

reason, the explicit form of $p_n(x)$ combined with Horner's method will be considered instead. On the other hand, notice that it will not be necessary to compute the viscosity matrix $Q_{i+1/2}$ explicitly, but only the vector $Q_{i+1/2}(w_{i+1} - w_i)$ appearing in the numerical flux (3).

To illustrate the procedure, consider the polynomial

$$p_2(x) = \alpha_0 x^4 + \alpha_1 x^2 + \alpha_2 = x^2(\alpha_0 x^2 + \alpha_1) + \alpha_2,$$

where $\alpha_0 = -1/8$, $\alpha_1 = 3/4$ and $\alpha_2 = 3/8$. Let $A \equiv A(w)$ be the Jacobian matrix of F evaluated at an intermediate state w, and let v be an arbitrary state; for simplicity, assume that $\lambda_{\max} = 1$. Then, as stated in Sect. 2, the following approximation holds:

$$|A|v \approx p_2(A)v = (A^2(\alpha_0 A^2 + \alpha_1 I) + \alpha_2 I)v.$$

The above expression can be computed using Horner's algorithm:

- Define $v_0 = v$ and compute $\widetilde{v}_0 = A^2 v_0$.
- Calculate $v_1 = \alpha_0 \widetilde{v}_0 + \alpha_1 v_0$ and $\widetilde{v}_1 = A^2 v_1$.
- Compute $v_2 = \widetilde{v}_1 + \alpha_2 v_0$. Then, $|A(w)|v \approx p_2(A)v = v_2$.

The product $A(w)v$ can be approximated using the finite difference formulation

$$A(w)v \approx \frac{F(w + \varepsilon v) - F(w)}{\varepsilon},$$

which leads to

$$A(w)^2 v \approx \frac{F\big(w + F(w + \varepsilon v) - F(w)\big) - F(w)}{\varepsilon} \equiv \Phi_\varepsilon(w; v).$$

In practice, the value ε has to be chosen small relative to the norm of w; for instance, in Sect. 4 we have taken $\varepsilon = 10^{-6}\|w\|_{L^2}$, which provides good results. Finally, the vector $|A(w)|v$ can be approximated using the following steps, in which only vector operations and evaluations of the physical flux F are involved:

- Define $v_0 = v$ and compute $\widetilde{v}_0 = \Phi_\varepsilon(w; v_0)$.
- Calculate $v_1 = \alpha_0 \widetilde{v}_0 + \alpha_1 v_0$ and $\widetilde{v}_1 = \Phi_\varepsilon(w; v_1)$.
- Compute $v_2 = \widetilde{v}_1 + \alpha_2 v_0$. Then, $|A(w)|v \approx v_2$.

4 Numerical Results

The equations of relativistic ideal magnetohydrodynamics (RMHD) have been chosen to analyze the behavior of the proposed schemes, mainly due to the complex form of the Jacobian of the system. It will be assumed throughout this section that the speed of light is normalized to $c = 1$. The notations are standard: see, e.g., [8].

4.1 One-Dimensional Test Problems

In this section we have chosen two one-dimensional tests that constitute standard references in RMHD. The initial conditions for these Riemann problems can be found in [1]. The adiabatic coefficient in Test 1 is $\gamma = 2$, while for Test 2 it is $\gamma = 5/3$. The tests have been computed using 800 cells, a Courant number of 0.8, and a final time of $t_f = 0.4$. To save space, only the results for the density component will be shown; similar comments and results apply for the other variables.

The numerical experiments have been performed with the Jacobian-free versions of the following methods:

- PVM-Cheb-12 and PVM-int-8: PVM methods based, respectively, on the Chebyshev approximation of degree 12 and the internal approximation of degree 8. The intermediate matrix $A_{i+1/2}$ has been taken as the Jacobian of the flux evaluated at the mean state $\frac{1}{2}(w_i + w_{i+1})$.
- DOT-Cheb-12 and DOT-int-8: approximate DOT solvers using the same polynomials as above and a Gauss-Legendre quadrature with $q = 3$ points.

The results have also been compared with the classical Harten-Lax-van Leer (HLL) method. In this case, the minimum and maximum speeds of propagation have been taken as -1 and 1 respectively, so HLL reduces to Rusanov's method.

Finally, with respect to the higher order schemes, a third-order PVM method has been considered in space, combined with a third-order TVD Runge-Kutta method for time stepping.

4.1.1 Blast Wave with Strong Initial Pressure Difference

This problem, first proposed in [1], consists in a blast wave problem with a very strong initial pressure. The maximal Lorenz factor is about 3.37, which means that the flow is strongly relativistic. The solution consists of two left-propagating fast and slow rarefaction waves, a contact discontinuity, and two right-going slow and fast shocks. In this case, the relativistic length-contraction effect induces a compression of the waves travelling to the right. Thus, the contact discontinuity and the right-going shocks remain under-resolved: see [1] for more details about this pathology. As it is shown in Fig. 1, our results are in good agreement with those in [1, 12].

4.1.2 Relativistic Shock Reflection Problem

In this section we consider the relativistic MHD analog of the Noh test problem proposed in [1]. Initially, there are two streams approaching each other with a Lorenz factor of 22.366, which makes this problem a extremely strong relativistic one. The solution, shown in Fig. 2, has two very strong fast shocks propagating outwards symmetrically in opposite directions. Moreover, two slow shocks travelling in opposite directions are formed, which are also properly computed.

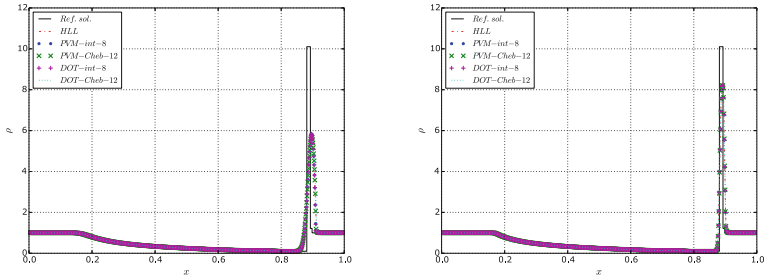


Fig. 1 Density component of the solution in test Sect. "Blast Wave with Strong Initial Pressure Difference". *Left* first order. *Right* third order

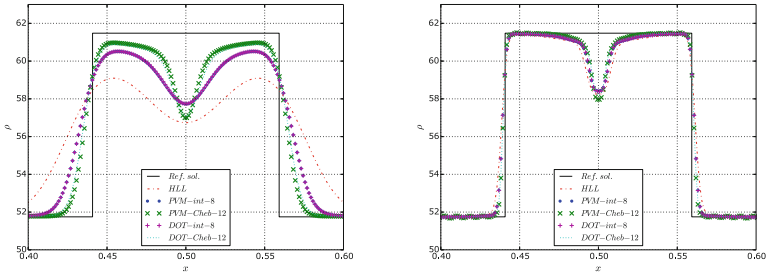


Fig. 2 Density component of the solution in test Sect. "Relativistic Shock Reflection Problem". Zoom of the central zone. *Left* first order. *Right* third order

The post-shock oscillations at the fast shocks are minimal and can be greatly damped by reducing the Courant number in the computation. At the point of symmetry $x = 0.5$ there is a spurious density undershoot due to the numerical pathology known as *wall heating*, that is produced by an undesired accumulation of entropy in a few zones around the point of symmetry. The numerical error around the undershoot is about 4.44%, which is quite acceptable (see [1, 12]).

4.2 Two-Dimensional Test Problems

Due to the challenging nature of the two-dimensional tests considered here, we have considered second-order TVD versions of the schemes, which seem robust enough to resolve the complex features of the solutions accurately. In general, the results obtained with the PVM-int-8 and PVM-Cheb-12 schemes are very similar. For this reason, in order to save space only the results obtained with the PVM-int-8 scheme will be shown. On the other hand, DOT-type schemes are not considered, as they produce similar results as PVM-type schemes but at a higher computational cost.

4.2.1 Relativistic Orszag-Tang Problem

This test constitutes the relativistic version of the Orszag-Tang vortex problem, which is a well-known model for testing the transition to supersonic MHD turbulence. The initial conditions can be found in [13]. The problem has been solved up to time $t = 4$ in the computational domain $[0, 2\pi] \times [0, 2\pi]$ using a 512×512 grid, with CFL=0.5 and adiabatic index $\gamma = 4/3$. Periodic boundary conditions have been considered.

Figure 3 shows the results obtained with the second-order PVM-int-8 scheme for the density component at times $t = 4$ and $t = 7$. At a qualitative level, our results are in good agreement with those presented in [13].

4.2.2 Cylindrical Blast Wave

This test concerns the evolution of a blast wave in a plasma with an initial uniform magnetic field. It is a canonical problem for testing the evolution of strong multidimensional shocks.

The blast wave is initiated using a cylinder of overpressured and overdense gas placed at the center of the domain, where the plasma is at rest and subject to a constant magnetic field in the x-direction. We have taken the same initial conditions as in [13]. The problem has been solved in the computational domain $[-6, 6] \times [-6, 6]$ using a 250×250 grid, until time $t = 4$ with CFL number 0.5. The adiabatic index is $\gamma = 4/3$ and transmissive boundary conditions have been applied.

Figure 4 shows the results for the case in which the initial magnetic field in the x-direction is taken as $B_x = 0.1$, obtained with the second-order PVM-int-8 scheme.

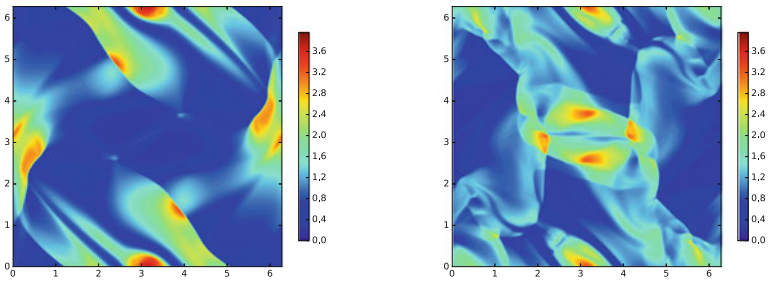


Fig. 3 Orszag-Tang vortex. Density at times $t = 4$ and $t = 7$

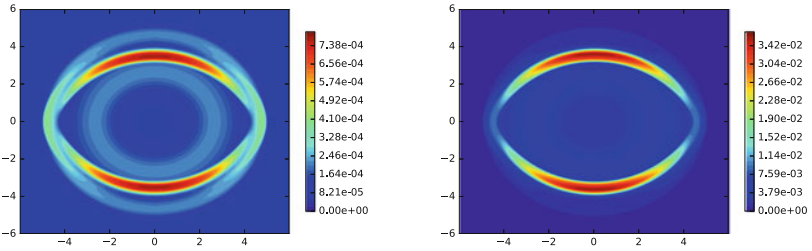


Fig. 4 Cylindrical blast wave with $B_x = 0.1$. Solution at time $t = 4$. *Left* density. *Right* pressure

As it can be seen, the main waves are properly captured: an almost circular external fast shock and an oblate reverse shock. The results agree with those found in the literature (see, e.g., [12, 13]).

Acknowledgements This research has been partially supported by the Spanish Government Research projects MTM2015-70490-C2-1R and MTM2014-56218-C2-2-P. The numerical computations have been performed at the Laboratory of Numerical Methods of the University of Málaga.

References

1. Balsara, D.: Total variation diminishing scheme for relativistic magnetohydrodynamics. APjS **132**, 83–101 (2001)
2. Castro, M.J., Fernández-Nieto, E.D.: A class of computationally fast first order finite volume solvers: PVM methods. SIAM J. Sci. Comput. **34**, A2173–A2196 (2012)
3. Castro, M.J., Gallardo, J.M., Marquina, A.: A class of incomplete Riemann solvers based on uniform rational approximations to the absolute value function. J. Sci. Comput. **60**, 363–389 (2014)
4. Castro, M.J., Gallardo, J.M., Marquina, A.: Approximate Osher-Solomon schemes for hyperbolic systems. Appl. Math. Comput. **272**, 347–368 (2016)
5. Cordier, F., Degond, P., Kumbaro, A.: Phase appearance or disappearance in two-phase flows. J. Sci. Comput. **58**, 115–148 (2014)
6. Degond, P., Peyrard, P.F., Russo, G., Villedieu, P.: Polynomial upwind schemes for hyperbolic systems. C. R. Acad. Sci. Paris Sér. I **328**, 479–483 (1999)
7. Dumbser, M., Toro, E.F.: On universal Osher-type schemes for general nonlinear hyperbolic conservation laws. Commun. Comput. Phys. **10**, 635–671 (2011)
8. Martí, J.M., Müller, E.: Numerical hydrodynamics in special relativity. Living Rev. Relativ. **6** (2003). http://www.livingreviews.org/lrr-2003-7
9. Osher, S., Solomon, F.: Upwind difference schemes for hyperbolic conservation laws. Math. Comput. **38**, 339–374 (1982)
10. Roe, P.L.: Approximate Riemann solvers, parameter vectors, and difference schemes. J. Comput. Phys. **43**, 357–372 (1981)
11. Torrilhon, M.: Krylov-Riemann solver for large hyperbolic systems of conservation laws. SIAM J. Sci. Comput. **34**, A2072–A2091 (2012)

12. Zanna, L.d., Bucciantini, N., Londrillo, P.: An efficient shock-capturing central-type scheme for multidimensional relativistic flows II. Magnetohydrodynamics. A & A **400**, 397–413 (2003)
13. Zanotti, O., Fambri, F., Dumbser, M.: Solving the relativistic magnetohydrodynamics equations with ADER discontinuous Galerkin methods, a posteriori subcell limiting and adaptive mesh refinement. Mon. Not. R. Astron. Soc. **452**, 3010–3029 (2015)

A Fractional Step Method to Simulate Mixed Flows in Pipes with a Compressible Two-Layer Model

Charles Demay, Christian Bourdarias, Benoît de Laage de Meux, Stéphane Gerbi and Jean-Marc Hérard

Abstract The so-called mixed flows in pipes include two-phase stratified regimes as well as single-phase pressurized regimes with transitions. It is proposed to handle those configurations numerically with the compressible two-layer model developed in [7]. Thus, a fractional step method is proposed to deal explicitly with the *slow* propagation phenomena and implicitly with the *fast* ones. It results in a large time-step scheme accurate in both regimes. Numerical experiments are performed including convergence results and academical test cases.

Keywords Two-layer model · Implicit-explicit scheme · Mixed flow

1 Introduction

We focus on air-water flows in pipes and particularly on the so-called mixed flows. The latter include stratified regimes driven by *slow* surface waves as well as pressurized regimes (pipe full of water or air) driven by *fast* acoustic waves. This type of flow occurs in piping systems of several industrial areas such as nuclear and hydraulic power plants or sewage pipelines.

C. Demay (✉) · B. de Laage de Meux · J.-M. Hérard
EDF Lab, 6 Quai Watier, 78400 Chatou, France
e-mail: charles.demay@edf.fr

C. Bourdarias · S. Gerbi
LAMA, UMR 5127-CNRS, Université Savoie Mont Blanc, 73376 Le Bourget-du-lac, France
e-mail: christian.bourdarias@univ-smb.fr

B. de Laage de Meux
e-mail: benoit.de-laage-de-meux@edf.fr

S. Gerbi
e-mail: stephane.gerbi@univ-smb.fr

J.-M. Hérard
I2M, UMR 7373-CNRS, Université Aix-Marseille, 13453 Marseille, France
e-mail: jean-marc.herard@edf.fr

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_4

Numerous modelling and numerical issues are tackled when dealing with mixed flows due to the different nature of each regime. Using a 1D approach, a model with an associated numerical scheme is proposed in [1] without computing the air phase. With the aim of accounting for air-water interactions, a compressible two-layer model is developed in [7]. It results in an hyperbolic two-phase two-pressure model which presents strong similarities with the isentropic form of two-fluid models introduced in [3]. In that framework, classical explicit schemes bring large numerical diffusivity in the *slow* stratified regime.

Thus, a fractional step method is derived herein to split the *slow* dynamics from the *fast* dynamics and adapt the numerical treatment. This approach is used in [2, 5] for the Baer-Nunziato model and more recently in [6] for the model under consideration. Furthermore, an implicit-explicit time discretization is also proposed in the sequel to end up with a large time-step scheme and get accuracy in the stratified regime. Contrary to the work presented in [6], the overall approach is driven by the fast pressure relaxation and the shallow-water structure of the system such that interesting results are obtained even for low speed flows.

2 The Compressible Two-Layer Model

The considered model deals with stratified gas-liquid flows in pipes. It results from a depth-averaging of the isentropic Euler set of equations for each phase where the classical hydrostatic assumption is made for the liquid, see [7] for details. Considering a two-layer air-water flow through a pipe of height H, it reads:

$$\left\{\begin{array}{l} \partial_t h_1 + U_I \partial_x h_1 = \lambda_p (P_I - P_2(\rho_2)), \\ \partial_t m_k + \partial_x m_k u_k = 0, \\ \partial_t m_k u_k + \partial_x m_k u_k^2 + \partial_x h_k P_k(\rho_k) - P_I \partial_x h_k = (-1)^k \lambda_u (u_1 - u_2), \end{array}\right. \qquad (\mathscr{S})$$

where $k = 1$ for water, $k = 2$ for air, $m_k = h_k \rho_k$ and $h_1 + h_2 = H$. Here, h_k, ρ_k, $P_k(\rho_k)$ and u_k denote respectively the height, the mean density, the mean pressure and the mean velocity of phase k. The interfacial dynamics is represented by the transport equation on h_1 while the other two equations account for mass and momentum conservation in each phase. The interfacial pressure is denoted by P_I and closed by the hydrostatic constraint, while the interfacial velocity is denoted by U_I and closed following an entropy inequality, one obtains (see [7]):

$$(U_I, P_I) = (u_2, P_1(\rho_1) - \rho_1 g \frac{h_1}{2}), \qquad (1)$$

where g is the gravity field magnitude. As the phases are compressible, state equations are required for gas and liquid pressures. For instance, perfect gas law may be used for air and linear law for water. The celerity of acoustic waves is defined by

$c_k = \sqrt{P_k^{'}(\rho_k)}$. In practice, one uses $\lambda_p = \frac{3h_1h_2}{4\pi\mu_1 H}$ and $\lambda_u = \frac{h_1h_2}{2H^2} f_i \rho_2 |u_2 - u_1|$, where μ_1 is the dynamic viscosity of water and f_i is a friction factor, see [6] for details.

Properties of $(\mathscr{S})$

(i) Smooth solutions of $(\mathscr{S})$ *comply with an entropy inequality.*
(ii) The convective part of $(\mathscr{S})$ *is hyperbolic under the condition* $|u_1 - u_2| \neq c_1$. *Its eigenvalues are unconditionally real and given by* $\lambda_1 = u_2$, $\lambda_{2,3} = u_1 \pm c_1$, $\lambda_{4,5} = u_2 \pm c_2$. *The field associated with the 1-wave is linearly degenerate while the other fields are genuinely nonlinear.*
(iii) Unique jump conditions hold within each isolated field.
(iv) The positivity of h_k *and* ρ_k *is verified.*

The details and proofs are provided in [7]. Two additional properties of $(\mathscr{S})$ are used in the proposed fractional step method. Firstly, using (1), the momentum equation for water can be written under a *Saint-Venant*-like form (see [8]):

$$\partial_t m_1 u_1 + \partial_x m_1 u_1^2 + \partial_x \rho_1 g \frac{h_1^2}{2} + h_1 \partial_x P_I = \lambda_u (u_2 - u_1). \tag{2}$$

Secondly, the pressure relaxation in the first equation of $(\mathscr{S})$ writes classically:

$$P_I \underset{t\to\infty}{\to} P_2, \tag{3}$$

and this relaxation is very fast in our framework as $\lambda_p \gg 1$. In addition, regarding the pressurized regime, $(\mathscr{S})$ degenerates towards a single-phase Euler system when one phase vanishes, as soon as the source terms also vanish.

3 Fractional Step Method Adapted to Mixed Flows

In order to handle both regimes included in mixed flows, the proposed fractional step method splits $(\mathscr{S})$ into three sub-systems. The *material* component of $(\mathscr{S})$ is treated in $(\mathscr{S}_m)$ including the pressure relaxation source term and using the Saint-Venant structure (2) for the water phase:

$$\begin{cases} \partial_t h_1 + u_2 \partial_x h_1 = \lambda_p (P_I - P_2), \\ \partial_t m_k + \partial_x m_k u_k = 0, \ k = 1, 2, \\ \partial_t m_1 u_1 + \partial_x m_1 u_1^2 + \partial_x \rho_1 g \dfrac{h_1^2}{2} = 0, \\ \partial_t m_2 u_2 + \partial_x m_2 u_2^2 = 0. \end{cases} \tag{$\mathscr{S}_m$}$$

$(\mathscr{S}_a)$ refers to the *acoustic* component of $(\mathscr{S})$ including the pressure gradients:

$$\begin{cases} \partial_t h_k = 0,\ \partial_t m_k = 0,\ k = 1, 2, \\ \partial_t m_1 u_1 + h_1 \partial_x P_I = 0, \\ \partial_t m_2 u_2 + h_2 \partial_x P_2 + (P_2 - P_I)\partial_x h_2 = 0, \end{cases} \qquad (\mathscr{S}_a)$$

where $P_I = P_1(\rho_1) - \rho_1 g \frac{h_1}{2}$. Finally, $(\mathscr{S}_u)$ deals with the velocity relaxation source terms:

$$\partial_t h_k = 0,\ \partial_t m_k = 0,\ \partial_t m_k u_k = (-1)^k \lambda_u (u_1 - u_2),\ k = 1, 2. \qquad (\mathscr{S}_u)$$

A key feature is that the fast relaxation (3) solved in $(\mathscr{S}_m)$ is explicitly seen by $(\mathscr{S}_a)$.

Proposition 1 (Hyperbolicity of $(\mathscr{S}_m)$) *The convective part of $(\mathscr{S}_m)$ is weakly hyperbolic. Its eigenvalues are given by* $\{u_2; u_1 \pm \sqrt{g \frac{h_1}{2}}\}$.

$(\mathscr{S}_a)$ is not hyperbolic as its spectrum reduces to zero. This singularity is handled in the sequel using a relaxation approach.

In the discrete setting, the time step is denoted Δt and the space step Δx. The space is partitioned into cells $C_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}[$ where $x_{i+\frac{1}{2}} = (i + \frac{1}{2})\Delta x$ are the cell interfaces. At times $t^n = n\Delta t$, the solution is approximated on each cell C_i by $\mathbf{W}_i^n = \Big((h_1)_i^n, (h_1\rho_1)_i^n, (h_2\rho_2)_i^n, (h_1\rho_1 u_1)_i^n, (h_2\rho_2 u_2)_i^n\Big)^T$.

Step 1: Explicit scheme for $(\mathscr{S}_m)$. In this step, W_i is updated from W_i^n to W_i^*. A classical explicit finite-volume scheme with Rusanov fluxes is used on the convective part while the pressure relaxation source term is treated implicitly. It writes:

$$\mathbf{W}_i^* = \mathbf{W}_i^n - \frac{\Delta t}{\Delta x}\Big(\mathbf{F}(\mathbf{W}_{i+\frac{1}{2}}^n) - \mathbf{F}(\mathbf{W}_{i-\frac{1}{2}}^n)\Big) - \frac{\Delta t}{2\Delta x}\mathbf{B}(\mathbf{W}_i^n)\Big(\mathbf{W}_{i+1}^n - \mathbf{W}_{i-1}^n\Big) + \mathbf{S}(\mathbf{W}_i^*), \qquad (4)$$

where $\mathbf{F}(\mathbf{W}) = (0, m_1u_1, m_2u_2, m_1u_1^2 + m_1 g\frac{h_1}{2}, m_2u_2^2)^T$, $\mathbf{B}(\mathbf{W}) = (u_2, 0, 0, 0, 0)^T$ and $\mathbf{S}(\mathbf{W}) = (\lambda_p(P_I - P_2), 0, 0, 0, 0)^T$. The fluxes are defined by:

$$\begin{cases} \mathbf{F}(\mathbf{W}_{i+\frac{1}{2}}^n) = \dfrac{1}{2}\Big(\mathbf{F}(\mathbf{W}_i^n) + \mathbf{F}(\mathbf{W}_{i+1}^n) - r_{i+\frac{1}{2}}(\mathbf{W}_{i+1}^n - \mathbf{W}_i^n)\Big), \\ r_{i+\frac{1}{2}} = \max\limits_{j\in\{i;i+1\}}\Big(|u_{2,j}^n|; |\Big(u_1 \pm \sqrt{g\frac{h_1}{2}}\Big)_j^n|\Big). \end{cases} \qquad (5)$$

In order to solve implicitly the source term, the mass terms $m_{k,i}^n$ are updated first and the first equation in $(\mathscr{S}_m)$ is solved under the form $f(h_{1,i}^*) = 0$ where:

$$f(y) = y - h_{1,i}^n + \Delta t \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u_2^n \frac{\partial h_1^n}{\partial x} dx - \Delta t \lambda_{p,i}^n \Big(P_I\Big(\frac{m_{1,i}^*}{y}\Big) - P_2\Big(\frac{m_{2,i}^*}{H - y}\Big)\Big). \qquad (6)$$

One may easily demonstrate that f is strictly increasing on $[0; H]$ with the limits $f \underset{0^+}{\to} -\infty$ and $f \underset{H^-}{\to} +\infty$, such that $f(x) = 0$ admits a unique solution $h^*_{1,i}$ on $[0; H]$.

Proposition 2 (Positivity of heights and densities) *The proposed scheme for* $(\mathscr{S}_m)$ *ensures the positivity of heights and densities under the classical CFL condition:*

$$\frac{\Delta t}{\Delta x} \max_i \left(\frac{r_{i+\frac{1}{2}} + r_{i-\frac{1}{2}}}{2} \right) < 1, \tag{7}$$

which only implies material velocities.

Step 2: Implicit relaxation approach for $(\mathscr{S}_a)$. In this step, only u_k is updated from u_k^* to u_k^{**}. The lack of hyperbolicity is handled with a relaxation approach, see [4, 5], introducing the system $(\mathscr{S}_a^r)$ which relaxes towards $(\mathscr{S}_a)$ in the limit $\varepsilon \to 0$:

$$\begin{cases} \partial_t h_k = 0, \ \partial_t m_k = 0, \ k = 1, 2, \\ \partial_t m_1 u_1 + h_1 \partial_x \Pi_I = 0, \\ \partial_t m_2 u_2 + h_2 \partial_x \Pi_2 + (\Pi_2 - \Pi_I) \partial_x h_2 = 0, \\ \partial_t m_k \Pi_k + a_k^2 h_k \partial_x u_k + a_k^2 (u_k - u_2) \partial_x h_k = \dfrac{1}{\varepsilon} m_k (\Pi_k - P_k), \ k = 1, 2, \end{cases} \tag{$\mathscr{S}_a^r$}$$

where $\Pi_I = \Pi_1 - \rho_1 g \frac{h_1}{2}$ and Π_k relaxing toward P_k as $\varepsilon \to 0$. The PDE verified by Π_k is derived from the PDE verified by P_k in $(\mathscr{S})$. In addition, a_k are positive numerical parameters used to ensure the stability of the relaxation approximation in the regime of small ε, their definition is provided later according to the flow regime.

Proposition 3 (Hyperbolicity of $(\mathscr{S}_a^r)$) *When* $a_k > 0$*, the convective part of* $(\mathscr{S}_a^r)$ *is strictly hyperbolic. Its eigenvalues are given by* $\{0; \pm\frac{a_1}{\rho_1}; \pm\frac{a_2}{\rho_2}\}$.

In order to keep a numerical diffusivity based on the *material* CFL condition (7), an implicit-explicit time discretization is proposed for the convective part of $(\mathscr{S}_a^r)$:

$$\begin{cases} h_k^{**} = h_k^*, m_k^{**} = m_k^*, \ k = 1, 2, \\ (m_1^{**} u_1^{**} - m_1^* u_1^*)/\Delta t + h_1^{**} \partial_x \Pi_I^{**} = 0, \\ (m_2^{**} u_2^{**} - m_2^* u_2^*)/\Delta t + h_2^{**} \partial_x \Pi_2^{**} + (\Pi_2^* - \Pi_I^*) \partial_x h_2^* = 0, \\ (m_k^{**} \Pi_k^{**} - m_k^* \Pi_k^*)/\Delta t + a_k^{2^*} h_k^{**} \partial_x u_k^{**} + a_k^{2^*} (u_k^* - u_2^*) \partial_x h_k^* = 0, \ k = 1, 2. \end{cases} \tag{8}$$

Classical combinations on (8) lead to the following semi-discrete equations on u_k:

$$\begin{cases} \dfrac{u_1^{**}-u_1^*}{\Delta t} - \dfrac{\Delta t}{\rho_1^*}\partial_x\left(\dfrac{a_1^{2^*}}{\rho_1^*}\partial_x u_1^{**}\right) = -\dfrac{1}{\rho_1^*}\partial_x P_I^* + \dfrac{\Delta t}{\rho_1^*}\partial_x\left(\dfrac{a_1^{2^*}(u_1^*-u_2^*)}{m_1^*}\partial_x h_1^*\right), \\ \dfrac{u_2^{**}-u_2^*}{\Delta t} - \dfrac{\Delta t}{\rho_2^*}\partial_x\left(\dfrac{a_2^{2^*}}{\rho_2^*}\partial_x u_2^{**}\right) = -\dfrac{1}{\rho_2^*}\partial_x P_2^* - \dfrac{(P_2^*-P_I^*)}{m_2^*}\partial_x h_2^*. \end{cases} \tag{9}$$

In (9), instantaneous relaxation ($\varepsilon \to 0$) is assumed between Π_k and P_k such that $\Pi_k^* = P_k^*$. Thus, the proposed implicit relaxation approach acts as a stabilization process involving a diffusion term weighted by a_k.

Definition 1 Under the light of (9), a_k is defined according to the flow regime:

- In the stratified regime ($h_1 < H$): the pressure gradient $h_1\partial_x P_I$ in ($\mathscr{S}_a$) is seen as a source term. It accounts for variable interfacial pressure which can be interpreted as air phase pressure due to the relaxation (3) solved in the first step. Thus, a_1 is set to zero.
- In the pressurized regime ($h_1 = H$): the stabilization process is applied and a_1 must follow the so-called Whitham condition: $a_1^2 > \max_{\rho_1}(\rho_1^2 c_1^2)$, see [4, 5].
- In all the regimes, a_2 follows the Whitham condition $a_2^2 > \max_{\rho_2}(\rho_2^2 c_2^2)$.

After integrating (9) on a cell C_i and using centered schemes for gradients, one obtains an implicit system which may be written in matrix form:

$$A_k^* \mathbb{U}_k^{**} = \mathbb{S}_k^*, \tag{10}$$

where A_k^* is a non-singular tridiagonal matrix (M-matrix structure) and $\mathbb{S}_k^*$ corresponds to the integrated source term. Calculations are not detailed here. In practice, the diffusion coefficient $(a_k^2/\rho_k)^*_{i+\frac{1}{2}}$ is computed using an harmonic average and a threshold on h_1 is introduced to identify the flow regime.

Step 3: Implicit scheme for ($\mathscr{S}_u$). In this step, only u_k is updated from u_k^{**} to u_k^{n+1}. The velocity relaxation source term is treated implicitly (except the λ_u coefficient) such that the following non-singular 2x2 system is obtained:

$$\begin{pmatrix} m_{1,i}^{**} + \Delta t\lambda_{u,i}^{**} & -\Delta t\lambda_{u,i}^{**} \\ -\Delta t\lambda_{u,i}^{**} & m_{2,i}^{**} + \Delta t\lambda_{u,i}^{**} \end{pmatrix} \begin{pmatrix} u_{1,i}^{n+1} \\ u_{2,i}^{n+1} \end{pmatrix} = \begin{pmatrix} (m_1u_1)_i^{**} \\ (m_2u_2)_i^{**} \end{pmatrix}. \tag{11}$$

This step concludes the overall scheme which ensures the positivity of heights and densities under the *material* CFL condition (7).

4 Numerical Results

In this section, the proposed scheme is denoted $\mathrm{SP_r}$ and compared with a classical Rusanov scheme applied on ($\mathscr{S}$) under an *acoustic* CFL condition.

Riemann problem for the convective part. One considers an analytical solution which contains two shocks for each phase traveling with the *fast* acoustic waves and a contact discontinuity (*slow* wave) where h_1 jumps. Without the pressure relaxation (3), note that a_1 follows the Whitham condition. Fields are displayed on Fig. 1 at $T = 23.10^{-5}s$ with 500 cells. A mesh refinement is also performed to check the numerical convergence of the method.

As expected, the $\mathrm{SP_r}$ scheme is accurate on the slow wave. Regarding the fast waves, it is more diffusive than Rusanov on phase 1 (the fastest) while better results are obtained on phase 2. Indeed, the optimal regime for the Rusanov scheme is on phase 1 with *acoustic* time steps. Stability and convergence towards relevant shock solutions are obtained with the expected convergence rate $\frac{1}{2}$ due to the contact discontinuity.

Dambreak. The source terms are activated and one considers the dambreak problem where the initial condition is a discontinuity on h_1 with constant density and zero speed. Regarding the water layer, the (incompressible) Saint-Venant system admits an analytical solution, see [8]. As the compressibility of water as well as the additional air layer should have a minor influence here, one expects to obtain the same kind of solution for phase 1. Therefore, a 1 m long pipe is considered with

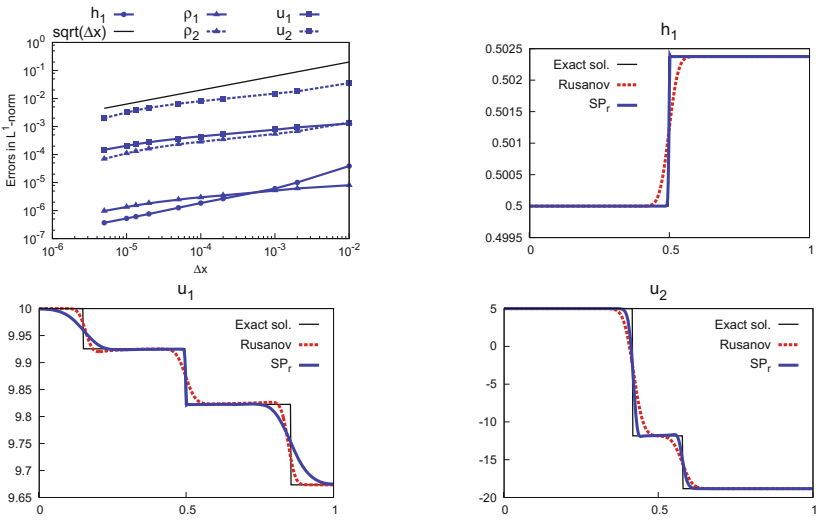


Fig. 1 Errors in L^1-norm and fields at $T = 23.10^{-5}s$ with 500 cells for the Riemann problem

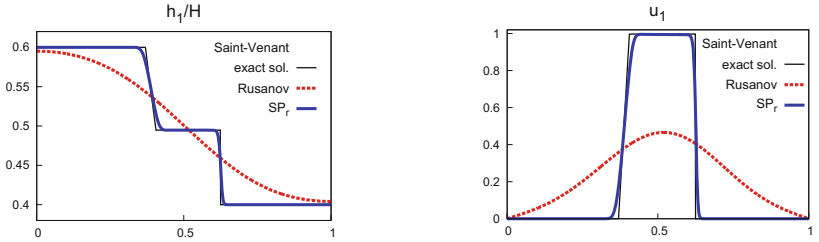


Fig. 2 Fields at $T = 24.10^{-3}s$ with 1000 cells for the dambreak problem

$(h_1/H)_L = 0.6$ and $(h_1/H)_R = 0.4$ as initial conditions. The velocity and height fields for phase 1 are plotted on Fig. 2 at $T = 24.10^{-3}s$ using 1000 cells.

Contrary to the results obtained with the large time-step scheme proposed in [6], the $\mathrm{SP_r}$ scheme displays accurate fields regarding the Saint-Venant solution. The Rusanov scheme is highly diffusive and regarding CPU time, it needs 3 minutes while $\mathrm{SP_r}$ takes 6 seconds. Those results emphasize the fact that a classical explicit scheme applied on $(\mathscr{S})$ is not adapted to low speed configurations.

Mixed flow. One considers a closed sloping pipe with constant height and zero speed as initial conditions. The pipe is 5 m long with $H = 1\mathrm{m}$, $h_1 = 0.8\mathrm{m}$, $\theta = 30$ degrees. A mesh of 250 cells is used and the threshold is set to $0.99H$. The flow becomes pressurized at the bottom (only water) and dried at the top (only air), see Fig. 3 for a snapshot of the water height and Fig. 4 for the pressure field.

Interesting qualitative results are obtained which demonstrates the ability of the $\mathrm{SP_r}$ scheme to handle mixed flows. Regarding the pressure field, one observes oscillations at the transition point between the regimes which are classical when dealing with mixed flows, see [1]. In the pressurized region, the pressure gradient slope is given by the expected equilibrium $\frac{\partial P_1}{\partial x} = -\rho_1 g \sin(\theta)$.

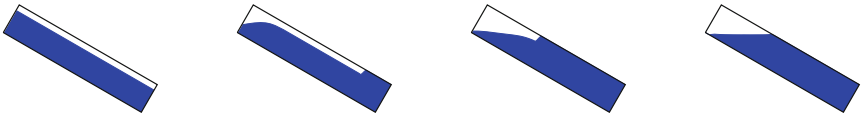


Fig. 3 Pipe filling snapshots for water height with 250 cells

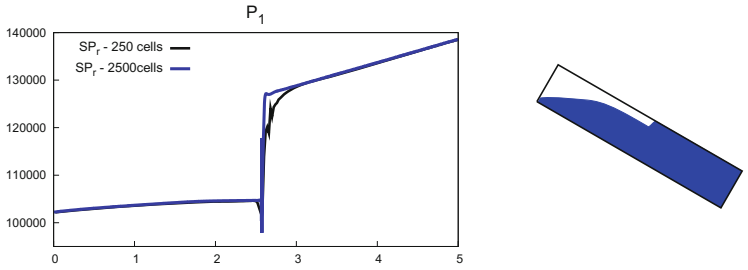


Fig. 4 Pressure field and water height at $T = 0.5s$

Acknowledgements C. Demay received a financial support by ANRT through an EDF-CIFRE contract 2014/749. Numerical facilities were provided by EDF.

References

1. Bourdarias, C., Gerbi, S.: A finite volume scheme for a model coupling free surface and pressurised flows in pipes. J. Comput. Appl. Math. **209**, 1–47 (2007)
2. Chalons, C., Coquel, F., Kokh, S., Spillane, N.: Large time-step numerical scheme for the seven-equation model of compressible two-phase flows. Springer Proc. Math. Stat. **4**, 225–233 (2011)
3. Coquel, F., Gallouët, T., Hérard, J.M., Seguin, N.: Closure laws for a two-fluid two-pressure model. C. R. Acad. Sci. Paris **334**(I), 927–932 (2002)
4. Coquel, F., Godlewski, E., Seguin, N.: Relaxation of fluid systems. Math. Models Methods Appl. Sci. **22**(8) (2012)
5. Coquel, F., Hérard, J.M., Saleh, K.: A splitting method for the isentropic Baer-Nunziato two-phase flow model. ESAIM: Proc. **38**(3), 241–256 (2012)
6. Demay, C., Bourdarias, C., de Laage de Meux, B., Gerbi, S., Hérard, J.M.: Numerical simulation of a compressible two-layer model: a first attempt with an implicit-explicit splitting scheme. *Submitted* (2016). https://hal.archives-ouvertes.fr/hal-01421889
7. Demay, C., Hérard, J.M.: A compressible two-layer model for transient gas-liquid flows in pipes. Contin. Mech. Thermodyn. **29**(2), 385–410 (2017)
8. Gerbeau, J.F., Perthame, B.: Derivation of viscous Saint-Venant system for laminar shallow water; numerical validation. Discret. Contin. Dyn. Syst. Ser. B **1**, 89–102 (2001)

A Second Order Cell-Centered Scheme for Lagrangian Hydrodynamics

Théo Corot

Abstract We describe a high-order cell-centered Godunov type scheme for Lagrangian hydrodynamics on general unstructured meshes using nodal fluxes. The nodal solver only depends on the angular repartition of the physical variables around the node. A second order extension of the scheme, using a linear reconstruction and a Runge–Kutta method is described.

Keywords Lagrangian hydrodynamics · Godunov scheme · High-order finite volume method

1 Introduction

Lagrangian methods, which have the mesh moving with the fluid, are commonly used to simulate multi-material fluid flows. Indeed these methods have the advantage of capturing interfaces sharply. They are widely used in computational fluid dynamics. In this work we are insteresed in solving the two-dimensional compressible gas dynamics equations in the Lagrangian framework [2]. The physical variables considered are the density ρ, the velocity $\mathbf{u}$, the total energy E and the pressure p. The equations can be written in an integral form

$$\begin{cases} \frac{d}{dt}\int_{\Omega_j(t)} \rho dV = 0 \\ \frac{d}{dt}\int_{\Omega_j(t)} \rho \mathbf{u} dV + \int_{S(t)} p\mathbf{n} dS = \mathbf{0} \\ \frac{d}{dt}\int_{\Omega_j(t)} \rho E dV + \int_{S(t)} p\,(\mathbf{u},\mathbf{n})\, dS = 0 \\ \frac{d}{dt}\int_{\Omega_j(t)} dV - \int_{S(t)} (\mathbf{u},\mathbf{n})\, dS = 0 \end{cases} \tag{1}$$

T. Corot (✉)
Department IMATH, Conservatoire National des Arts Et Métiers, 2 Rue Conté, 75003 Paris, France
e-mail: theo.corot@cnam.fr

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_5

where Ω_j is, in pratice, a cell moving with the fluid. The first three equations of (1) describe mass, momentum and total energy conservation. The last one is the geometric conservation law.

Several methods have been developed to solve this system of conservation laws. For example staggered schemes [2] or free-Lagrange methods [11]. There are also complete codes developed using Lagrangian methods, like [15] for instance.

We choose to use a Godunov type scheme with node based fluxes. Indeed node velocities are needed to move the mesh. The development of this type of schemes has been enabled by Després and Mazeran [10] with GLACE scheme. Then it has been pursued by Maire, who highlighted a strong sensitivity of GLACE to the cell aspect in [13] which led to EUCCLHYD scheme. Later Burton and others developed a new scheme using similar ideas [3]. These schemes have been extended to high orders, unstructured grids, multi-dimension and arbitrary Lagrange-Euler [4, 5, 9, 12, 14]. However, we remarked in [8] that in a particular case of one dimensional Riemann problem, node velocities computed with GLACE and EUCCLHYD nodal solvers can have an incorrect direction. To overcome this problem we proposed an alternative scheme using a continous approach around the nodes. This approach led to the construction of a nodal solver which only depends on the angular repartition of the physical variables around the node and not on edge lengths. In this paper we recall this last scheme and describe a second order extension.

2 Presentation of the First Order Scheme

Here we present the first order scheme developed in [8] which will be extended to the second order in the next section.

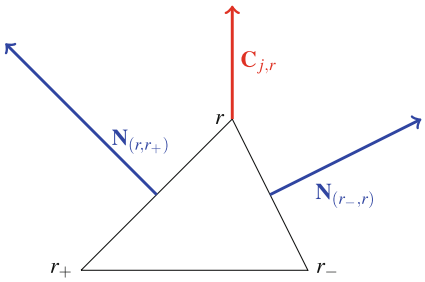
2.1 Mesh and Definition

Let us first provide some basic definitions concerning the mesh and our finite volume scheme. We note $\tau = \frac{1}{\rho}$ the specific volume, c the sound speed and $z = \rho c$ the acoustic impedance. We use general unstructured mesh, denote r the nodes and j the cells. We define V_j the volume of the cell j and M_j the constant mass of this cell. We note $\mathbf{x}_r$ the coordinates of the node r and $\mathbf{x}_j = \frac{1}{V_j}\int_{\Omega_j} \mathbf{x} dV$ the centroid of the cell j. We write $r \sim j$ if and only if the node r is a vertex of the cell j.

Since we use nodal fluxes, we need to define some normal at each node of a cell j,

$$\mathbf{C}_{j,r} = \frac{\mathbf{N}_{(r,r_-)} + \mathbf{N}_{(r,r_+)}}{2}, \tag{2}$$

Fig. 1 Cell j



where $\mathbf{N}_{(r,r_-)}$ and $\mathbf{N}_{(r,r_+)}$ are the normal vectors of the two edges of the cell j having r as a vertex (see Fig. 1) such that $|\ \mathbf{N}_{(r,r_-)}\ |$ (resp $|\ \mathbf{N}_{(r,r_+)}\ |$) is the length of the edge (r, r_-) (resp (r, r_+)). Our scheme can be written with a semi-discrete formulation

$$\begin{cases} M_j \frac{d}{dt}\tau_j - \sum\limits_{r\sim j} \left(\mathbf{u}_r, \mathbf{C}_{j,r}\right) = 0 \\ M_j \frac{d}{dt}\mathbf{u}_j + \sum\limits_{r\sim j} p_r \mathbf{C}_{j,r} = \mathbf{0} \\ M_j \frac{d}{dt}E_j + \sum\limits_{r\sim j} p_r \left(\mathbf{u}_r, \mathbf{C}_{j,r}\right) = 0 \end{cases} . \tag{3}$$

Now we need to describe how to compute $\mathbf{u}_r$ and p_r at each node. In order to describe the nodal solver, we define, as in Fig. 2, for each cell having r as a vertex, $\phi^1_{j,r}$ and $\phi^2_{j,r}$ the angles made by each of the two edges of the cell j having r as a vertex with a horizontal line. We note $\theta_{j,r} = \phi^2_{j,r} - \phi^1_{j,r}$. Furthermore we define the variation of a physical variable w (pressure, velocity or acoustic impedance) around a node with respect to an angle θ as the function $\theta \to w(\theta)$. These functions are piecewise constant. Moreover we define the nodal vector (see Fig. 2)

$$\mathbf{n}_\theta = -\begin{pmatrix} cos(\theta) \\ sin(\theta) \end{pmatrix}. \tag{4}$$

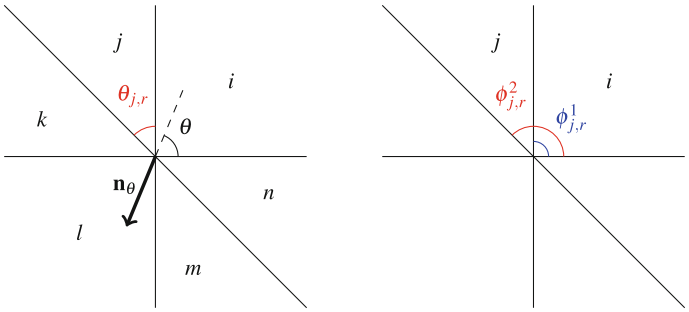


Fig. 2 Node r

2.2 The Nodal Solver

To compute nodal pressure p_r and velocity $\mathbf{u}_r$ we need to define an intermediate unknown pressure around each node $\theta \rightarrow \tilde{p}_r(\theta)$. Then our nodal solver can be written

$$\begin{cases} p(\theta) + z(\theta)\left(\mathbf{u}(\theta), \mathbf{n}_\theta\right) = \tilde{p}_r(\theta) + z(\theta)\left(\mathbf{u}_r, \mathbf{n}_\theta\right) \\ \int_0^{2\pi} \tilde{p}_r(\theta)\,\mathbf{n}_\theta d\theta = \mathbf{0} \\ \left(\int_0^{2\pi} \frac{1}{z(\theta)} d\theta\right) p_r = \int_0^{2\pi} \frac{1}{z(\theta)} \tilde{p}_r(\theta)\, d\theta \end{cases} . \tag{5}$$

Note that in one dimension, $p + zu$ and $p - zu$, are the acoustic Riemann invariants. Then the first equation of (5) is the multidimensional extension of these invariants. The second expresses that the sum of forces around the node vanishes. The last one defines p_r as a mean weighted value of $\tilde{p}_r$, which permit us to recover one unique pressure and, consequently, a conservative scheme. Weights are chosen in order to be as close as possible to the acoustic solver in the case of one dimensional Riemann problem [8]. The system (5) can be written

$$\begin{cases} A_r \mathbf{u}_r = \mathbf{b}_r \\ \Gamma_r p_r = \sum\limits_{j \sim r} \left[\Gamma_{j,r} p_j + \left(\mathbf{u}_j, \mathbf{n}_{j,r}\right)\right] \end{cases} \tag{6}$$

where we define

$$\begin{cases} A_r = \sum\limits_{j \sim r} A_{j,r} \\ \mathbf{b}_r = \sum\limits_{j \sim r} \left[A_{j,r}\mathbf{u}_j + p_j \mathbf{n}_{j,r}\right] \\ A_{j,r} = z_j \int_{\phi^1_{j,r}}^{\phi^2_{j,r}} \mathbf{n}_\theta \otimes \mathbf{n}_\theta d\theta \\ \mathbf{n}_{j,r} = \int_{\phi^1_{j,r}}^{\phi^2_{j,r}} \mathbf{n}_\theta d\theta \\ \Gamma_r = \sum\limits_{j \sim r} \Gamma_{j,r} \\ \Gamma_{j,r} = \frac{\theta_{j,r}}{z_j} \end{cases} . \tag{7}$$

Remark 1 This nodal solver only depends on the angular repartition of physical variables around the node and has no dependence on edge lengths. Moreover, it has been proven in [8] that, this solver always recovers the right direction of the velocity when one considers a one dimensional problem around a node. Finally the first order scheme using this nodal solver is conservative and verifies a weak consistency property.

3 Second Order Extension

In this section we describe a second order extension of our scheme. This extension is made thanks to a piecewise linear reconstruction of pressure and velocity in each cell and a Runge–Kutta two-step integration method to discretize the time derivative.

3.1 Piecewise Linear Reconstruction

Let w be a fluid variable (pressure or velocity components). The idea is to assume that w has a linear variation in the cell j

$$w_j(\mathbf{x}) = w_j + \left(\nabla w_j, \mathbf{x} - \mathbf{x}_j\right) \tag{8}$$

with $\mathbf{x}_j = \frac{1}{V_j}\int_{\Omega_j} \mathbf{x} dV$. Consequently we need to compute the gradient ∇w_j. Many ways are possible, such as using ENO [6], WENO [7] or a least squares method [12] to reconstruct the slope. In this paper we chose to construct it using a least squares method. The gradient is constructed by imposing that

$$\forall k \in Neig(j)\ w_j(\mathbf{x}_k) = w_k \tag{9}$$

where $Neig(j)$ is the set of all neighboring cells of j. In practice this problem is over-determined and the gradient is computed using a least squares method. Consequently it is computed as the solution of the minimization problem

$$\nabla w_j = argmin \sum_{k \in Neig(j)} \left[w_k - w_j - \left(\nabla w_j, \mathbf{x}_k - \mathbf{x}_j\right)\right]^2. \tag{10}$$

The solution of (10) is easy to compute, searching for zeros of the first order derivative one obtains

$$\nabla w_j = A_j^{-1} \sum_{k \in Neig(j)} \left(w_k - w_j\right)\left(\mathbf{x}_k - \mathbf{x}_j\right) \tag{11}$$

with

$$A_j = \sum_{k \in Neig(j)} \left(\mathbf{x}_k - \mathbf{x}_j\right) \otimes \left(\mathbf{x}_k - \mathbf{x}_j\right). \tag{12}$$

Once the gradient computed, one has to limit its value in order to preserve monotonicity. $\mathbf{x} \rightarrow w(\mathbf{x})$ becomes

$$w_j(\mathbf{x}) = w_j + \left(\phi_j \nabla w_j, \mathbf{x} - \mathbf{x}_j\right) \tag{13}$$

where $\phi_j \in [0, 1]$ and

$$\phi_j = \min_{r \sim j} \phi_{j,r} \tag{14}$$

with

$$\phi_{j,r} = \begin{cases} \mu\left(\frac{w_j^{max} - w_j}{w_j(\mathbf{x}_r) - w_j}\right) & \textbf{if } w_j(\mathbf{x}_r) - w_j > 0 \\ \mu\left(\frac{w_j^{min} - w_j}{w_j(\mathbf{x}_r) - w_j}\right) & \textbf{if } w_j(\mathbf{x}_r) - w_j < 0 \\ 1 & \text{else} \end{cases} . \tag{15}$$

Here w_j^{max} (resp w_j^{min}) denotes the maximum (resp minimum) of w in the neighbouring cells of j. The function μ characterizes the limiter. In this article we chose to use $\mu(x) = min(1, x)$ to obtain the Barth Jepersen limiter [1]. Once this linear reconstruction is done, in order to obtain a second order scheme, we use the values extrapolated at each node $w_j(\mathbf{x}_r)$ instead of the mean values w_j in the nodal solver.

3.2 Time Discretization

We chose to discretize the semi-discrete equations (3) using a second order Runge–Kutta method. Let us assume that we know the physical variables at a time t^n and we want to compute these quantities at a time $t^{n+1} = t^n + \delta t^n$. The second order Runge–Kutta scheme can be described using a predictor and a corrector step.

3.2.1 Predictor Step

- Compute nodal velocities $\mathbf{u}_r^n$ and pressures p_r^n using our nodal solver and extrapolated values at each node deduced with the linear reconstruction of Sect. 3.1.
- Compute nodal normals $\mathbf{C}_{j,r}^n$ and update the momentum and the total energy with

$$\begin{cases} \mathbf{u}_j^{n+\frac{1}{2}} = \mathbf{u}_j^n - \frac{\delta t^n}{2M_j} \sum_{r \sim j} p_r^n \mathbf{C}_{j,r}^n \\ E_j^{n+\frac{1}{2}} = E_j^n - \frac{\delta t^n}{2M_j} \sum_{r \sim j} \left(p_r^n \mathbf{u}_r^n, \mathbf{C}_{j,r}^n\right) \end{cases} . \tag{16}$$

- Update node positions

$$\mathbf{x}_r^{n+\frac{1}{2}} = \mathbf{x}_r^n + \frac{\delta t^n}{2} \mathbf{u}_r^n \tag{17}$$

and deduce cell densities $\rho_j^{n+\frac{1}{2}}$ using mass conservation

$$\rho_j^{n+\frac{1}{2}} = \frac{V_j^n}{V_j^{n+\frac{1}{2}}} \rho_j^n. \tag{18}$$

3.2.2 Corrector Step

- Knowing the physical variables and the geometry at the end of the predictor step, compute nodal velocities $\mathbf{u}_r^{n+\frac{1}{2}}$ and pressures $p_r^{n+\frac{1}{2}}$ with our nodal solver using extrapolated values given by the linear reconstruction.
- Compute nodal normals $\mathbf{C}_{j,r}^{n+\frac{1}{2}}$ and update the momentum and the total energy with

$$\begin{cases} \mathbf{u}_j^{n+1} = \mathbf{u}_j^n - \frac{\delta t^n}{M_j} \sum\limits_{r \sim j} p_r^{n+\frac{1}{2}} \mathbf{C}_{j,r}^{n+\frac{1}{2}} \\ E_j^{n+1} = E_j^n - \frac{\delta t^n}{M_j} \sum\limits_{r \sim j} \left(p_r^{n+\frac{1}{2}} \mathbf{u}_r^{n+\frac{1}{2}}, \mathbf{C}_{j,r}^{n+\frac{1}{2}} \right) \end{cases}. \tag{19}$$

- Update node positions

$$\mathbf{x}_r^{n+1} = \mathbf{x}_r^n + \delta t^n \mathbf{u}_r^{n+\frac{1}{2}} \tag{20}$$

and deduce the cell densities ρ_j^{n+1} using mass conservation.

4 Numerical Validation

Let us consider the classical one dimensional Sod shock tube in order to validate the second order extension. This case is a simple one dimensional Riemann problem involving one perfect gas with an adiabatic constant $\gamma = 1.4$. The shock tube is a $[0, 1] \times [0, 1]$ box where we initially have

$$(\rho, u, p)(t = 0, x) = \begin{cases} (1, 0, 1) & \text{if } x < 0.5 \\ (0.125, 0, 0.1) & \text{if } x > 0.5 \end{cases}. \tag{21}$$

All the boundaries are sliding walls. We compare the convergence order of the first and second order scheme. To do it we compute the test on four meshes : M_1 is a 5000×3 cartesian grid, M_2 a 10000×3, M_3 a 15000×3 and M_4 a 20000×3.

We note u_{err}, p_{err} and ρ_{err} L^1 errors made on the pressure, velocity and density. Let us also denote u_{err}^r, p_{err}^r and ρ_{err}^r L^1 errors made on a subdomain containing only the rarefaction wave. These errors are written in Tables 1 and 2.

The second order scheme gives higher convergence speed (Table 1) than the classical one obtained in the Eulerian framework, especially for the density. This can be explained by the fact that contrary to the Eulerian methods, Lagrangian schemes permit to follow the contact discontinuity exactly. Moreover one can see in Table 1

Table 1 Sod shock tube. Convergence of the numerical solution toward the exact solution

Mesh	Order	u_{err}	p_{err}	ρ_{err}	Order	u_{err}	p_{err}	ρ_{err}
M_1	1	4.7E-3	2.85E-3	2.86E-3	2	5.92E-4	2.79E-4	3.2E-4
M_2	1	2.65E-3	1.62E-3	1.64E-3	2	2.96E-4	1.41E-4	1.68E-4
M_3	1	1.89E-3	1.15E-3	1.18E-3	2	2E-4	9.51E-5	1.19E-4
M_4	1	1.45E-3	9.02E-4	9.26E-4	2	1.52E-4	7.26E-5	9.31E-5
Convergence order		≈ 0.84	≈ 0.84	≈ 0.82		≈ 0.96	≈ 0.96	≈ 0.86

Table 2 Sod shock tube. Convergence of the rarefaction wave toward the exact solution

Mesh	Order	u^r_{err}	p^r_{err}	ρ^r_{err}	Order	u^r_{err}	p^r_{err}	ρ^r_{err}
M_1	1	1.39E-3	9.15E-4	7.75E-4	2	1.26E-4	7.68E-5	6.84E-5
M_2	1	7.95E-4	5.21E-4	4.44E-4	2	6.32E-5	3.91E-5	3.47E-5
M_3	1	5.69E-4	3.73E-4	3.19E-4	2	4.22E-5	2.64E-5	2.33E-5
M_4	1	4.48E-4	2.93E-4	2.51E-4	2	3.17E-5	2E-5	1.76E-5
Convergence order		≈ 0.83	≈ 0.83	≈ 0.83		≈ 0.99	≈ 0.97	≈ 0.98

that our second order scheme has a convergence order close to 1 for the pressure and the velocity while it is slightly lower for the density. However when one looks at the convergence speed only on the rarefaction (Table 2) all convergence orders are close to 1 for the second order scheme.

References

1. Barth, T., Jespersen, D.: The design and application of upwind schemes on unstructured meshes. In: 27th Aerospace Sciences Meeting, pp. 366 (1989)
2. Benson, D.J.: Computational methods in lagrangian and eulerian hydrocodes. Comput. methods Appl. mech. Eng. **99**(2), 235–394 (1992)
3. Burton, D., Carney, T., Morgan, N., Sambasivan, S., Shashkov, M.: A cell-centered lagrangian godunov-like method for solid dynamics. Comput. & Fluids **83**, 33–47 (2013)
4. Burton, D.E., Morgan, N.R., Carney, T.C., Kenamond, M.A.: Reduction of dissipation in lagrange cell-centered hydrodynamics (cch) through corner gradient reconstruction (cgr). J. Comput. Phys. **299**, 229–280 (2015)
5. Carré, G., Del Pino, S., Després, B., Labourasse, E.: A cell-centered lagrangian hydrodynamics scheme on general unstructured meshes in arbitrary dimension. Journal of Computational Physics **228**(14), 5160–5183 (2009)
6. Cheng, J., Shu, C.W.: Positivity-preserving lagrangian scheme for multi-material compressible flow. J. Comput. Phys. **257**, 143–168 (2014)
7. Cheng, J., Shu, C.W.: Second order symmetry-preserving conservative lagrangian scheme for compressible euler equations in two-dimensional cylindrical coordinates. J. Comput. Phys. **272**, 245–265 (2014)
8. Corot, T., Mercier, B.: A new nodal solver for the two dimensional lagrangian hydrodynamics. Submitted to Journal of Computational Physics (2016)

9. Després, B.: Weak consistency of the cell-centered lagrangian glace scheme on general meshes in any dimension. Comput. Methods Appl. Mech. Eng. **199**(41), 2669–2679 (2010)
10. Després, B., Mazeran, C.: Lagrangian gas dynamics in two dimensions and lagrangian systems. Arch. Ration. Mech. Anal. **178**(3), 327–372 (2005)
11. Howell, B., Ball, G.: A free-lagrange augmented godunov method for the simulation of elastic-plastic solids. J. Comput. Phys. **175**(1), 128–167 (2002)
12. Maire, P.H.: A high-order cell-centered lagrangian scheme for two-dimensional compressible fluid flows on unstructured meshes. J. Comput. Phys. **228**(7), 2391–2425 (2009)
13. Maire, P.H., Abgrall, R., Breil, J., Ovadia, J.: A cell-centered lagrangian scheme for two-dimensional compressible flow problems. SIAM J. Sci. Comput. **29**(4), 1781–1824 (2007)
14. Maire, P.H., De Buhan, M., Diaz, A., Dobrzynski, C., Kluth, G., Lagoutière, F.: A cell-centered arbitrary lagrangian eulerian (ale) method for multi-material compressible flows. In: ESAIM: proceedings, vol. 24, pp. 1–13. EDP Sciences (2008)
15. Rathkopf, J.A., Miller, D.S., Owen, J., Stuart, L., Zika, M., Eltgroth, P., Madsen, N., McCandless, K., Nowak, P., Nemanic, M., et al.: Kull: Llnls asci inertial confinement fusion simulation code. In: Physor 2000, ANS Topical Meeting on Advances in Reactor Physics and Mathematics and Computation into the Next Millennium (2000)

An Implicit Integral Formulation for the Modeling of Inviscid Fluid Flows in Domains Containing Obstacles

Clément Colas, Martin Ferrand, Jean-Marc Hérard, Erwan Le Coupanec and Xavier Martin

Abstract We focus here on an integral approach to compute compressible inviscid fluid flows in physical domains cluttered up with many small obstacles. This approach is based on a multidimensional porous integral formulation of Euler system of equations. Its discretization uses a first order semi-implicit finite volume scheme with pressure-correction algorithm preserving the positivity of both density and pressure. Numerical tests are completed by simulating a 2D channel flow containing two aligned tubes. The results are compared to reference solutions obtained with a pure fluid approach on a fine mesh.

Keywords Finite volumes · Integral formulation · Porous media · Compressible flow

C. Colas (✉) · M. Ferrand · J.-M. Hérard · E. Le Coupanec
EDF R&D, MFEE, 6 quai Watier, 78400 Chatou, France
e-mail: clement.colas@edf.fr

M. Ferrand
e-mail: martin.ferrand@edf.fr

J.-M. Hérard
e-mail: jean-marc.herard@edf.fr

E. Le Coupanec
e-mail: erwan.lecoupanec@edf.fr

C. Colas · J.-M. Hérard
Aix-Marseille Université, I2M, UMR CNRS 7373, 39 rue Joliot Curie,
13453 Marseille, France

X. Martin
IFP Énergies nouvelles, 1 et 4 avenue du Bois Prau, 92852 Rueil-Malmaison, France
e-mail: xavier.martin@ifpen.fr

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_6

1 Introduction

In this paper we introduce a way to investigate fluid flows in thermohydraulic circuits components in nuclear reactors where three so-called "system", "component" and "local" representation scales coexist. The first one is a 0D/1D description and aims at providing a real time simulation of full circuits. The third one is the CFD scale and allows a fine description on restricted physical domains. The intermediate scale relies on a homogenized representation of some components [7, 9]; it consists in taking into account a fluid and solid volume in cells. Our purpose is to build a formulation embedding the "local" and "component" scales and ensuring the continuity between these two scales. A possible approach has been introduced in [6] using **explicit** schemes. The basic idea consists in an integral formulation of PDEs in a domain where a fluid flows around many small obstacles. Herein an **implicit** finite volume scheme is considered, using the open-source code *Code_Saturne* [4]. The compressible Euler equations (1) governing inviscid fluid flows are considered, and the unknowns ρ, $\boldsymbol{u}$, P respectively denote the density, the velocity and the pressure of the fluid, while the momentum is $\boldsymbol{Q} = \rho\boldsymbol{u}$. The volumetric total energy E is such that $E = \rho\left(\frac{\boldsymbol{u}^2}{2} + \epsilon(P,\rho)\right)$. The internal energy $\epsilon(P,\rho)$ is prescribed by the EOS (Equation Of State), $\boldsymbol{f}$ is a mass volumetric external force and Φ_v a volumetric heat transfer source term. Thus the set of governing equations is:

$$\begin{cases} \partial_t \rho \ + \nabla \cdot \boldsymbol{Q} = 0 \\ \partial_t \boldsymbol{Q} + \boldsymbol{\nabla} \cdot (\boldsymbol{u} \otimes \boldsymbol{Q}) + \boldsymbol{\nabla} P = \rho \boldsymbol{f} \\ \partial_t E \ + \nabla \cdot (\boldsymbol{u}(E+P)) = \rho \boldsymbol{f} \cdot \boldsymbol{u} + \rho \Phi_v \end{cases} \tag{1}$$

The speed of acoustic waves noted c is such that: $c^2 = \left(\frac{P}{\rho^2} - \frac{\partial\epsilon(P,\rho)}{\partial\rho}\right) / \left(\frac{\partial\epsilon(P,\rho)}{\partial P}\right)$. The total enthalpy is: $H = \frac{E+P}{\rho}$, and $\boldsymbol{W}$ is the conservative variable: $\boldsymbol{W} = (\rho, \boldsymbol{Q}, E)^t$.

2 Integral Formulation

The integral form of conservation laws described in [6, 8] is considered. Set of equations (1) is integrated on control volumes Ω_i which may contain many solid obstacles. All Ω_i cells form a mesh of the computational domain $\Omega \subset \mathbb{R}^d$ ($d = 1, 2$ or 3), such that: $\overline{\Omega} = \cup_i \overline{\Omega}_i$. The obstacles may be completely or partially included in Ω_i. Part of a control volume boundary may coincide with the surface of an obstacle. Figure 1 is a sketch of the admissible situtations. In the sequel, the subscript ij refers to interfaces between neighbouring control volumes Ω_i and Ω_j, and the superscript ϕ refers to fluid volumes and interfaces ij where the fluid may cross the interface, noted Γ_{ij}^{ϕ} of measure $S_{ij}^{\phi} = meas\left(\Gamma_{ij}^{\phi}\right)$. Besides, the superscript w refers to solid interfaces where a wall boundary Γ_i^w of measure S_i^w is located inside the control volume Ω_i

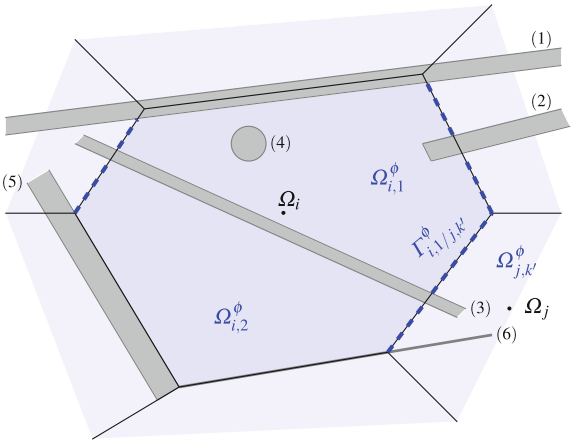


Fig. 1 A (*blue*) control volume Ω_i includes (*gray*) obstacles numbered from 1 to 5. Obstacles may: overlap part of the boundary of cell i (*1*); partially occupy one fluid cell (or subcell) (*2*); and split it into two fluid sub-cells $\Omega^{\phi}_{i,1}$ and $\Omega^{\phi}_{i,2}$ (*3*); be totally included in cell i or one of its subcells (*4*); be aligned with part of the boundary of cell i (*5*). The *dashed blue* surface corresponds to the fluid-fluid part of the boundary $\Gamma^{\phi}_{i,k}$ between sub-cells $\Omega_{i,k}$ and their neighbouring sub-cells occupied by the fluid

or on its boundary. The mass flux is null through surfaces S^w_i. The volume occupied by the fluid within the control volume Ω_i is denoted by Ω^{ϕ}_i. Nonetheless, a control volume Ω_i may contain several fluid sub-domains $\Omega^{\phi}_{i,k}$ ($k \in [\![1, N(i)]\!]$ with $N(i)$ the number of sub-elements), that are not connected to each other. We introduce within each fluid sub-cell $\Omega^{\phi}_{i,k}$ a mean value of the fluid state variable $\boldsymbol{W}(\mathbf{x}, t)$ noted $\boldsymbol{W}_{i,k}(t)$. The mean fluid state variable in cell Ω_i, $\boldsymbol{W}_i(t)$, is introduced as follows:

$$meas\left(\Omega^{\phi}_i\right)\boldsymbol{W}_i(t) = \sum_{k\in\{1,\ldots,N(i)\}} \int_{\Omega^{\phi}_{i,k}} \boldsymbol{W}(\mathbf{x}, t)dx$$

By additivity, using $\Omega^{\phi}_i = \bigcup_{k\in\{1,\ldots,N(i)\}} \Omega^{\phi}_{i,k}$, where $\Omega^{\phi}_{i,k}$ are all mutually disjoint:

$$meas\left(\Omega^{\phi}_i\right) = \sum_{k\in\{1,\ldots,N(i)\}} meas\left(\Omega^{\phi}_{i,k}\right)$$

The conservation laws (1) can be rewritten as follows:

$$\partial_t \boldsymbol{W} + \boldsymbol{\nabla}\cdot\boldsymbol{F}(\boldsymbol{W}) = \boldsymbol{D}(\boldsymbol{W}) \tag{2}$$

where $\boldsymbol{F}(\boldsymbol{W}) = (\rho\boldsymbol{u}, \rho\boldsymbol{u}\otimes\boldsymbol{u} + P\mathbf{I}, \boldsymbol{u}(E+P))^t$ is the convective flux and $\boldsymbol{D}(\boldsymbol{W}) = (0, \rho\boldsymbol{f}, \rho(\boldsymbol{f}\cdot\boldsymbol{u} + \Phi_v))^t$ represents the source term. Equation (2) is integrated over a bounded time interval $[t_0, t_1] \subset \mathbb{R}^+$and space with respect to the $\Omega^{\phi}_{i,k}$ sub-cell, the

divergence theorem allows to rewrite:

$$\int_{\Omega_{i,k}^{\phi}} (\boldsymbol{W}(\mathbf{x}, t_1) - \boldsymbol{W}(\mathbf{x}, t_0))\, dx + \int_{t_0}^{t_1} \int_{\Gamma_{i,k}} \boldsymbol{F}(\boldsymbol{W}(\mathbf{x}, t)) \cdot \boldsymbol{n} d\Gamma dt = \int_{t_0}^{t_1} \int_{\Omega_{i,k}^{\phi}} \boldsymbol{D}(\boldsymbol{W}(\mathbf{x}, t))\, dx\, dt \tag{3}$$

Here, $\Gamma_{i,k} = \partial\Omega_{i,k}^{\phi}$ denotes the whole boundary of the fluid sub-cell $\Omega_{i,k}^{\phi}$ with $\boldsymbol{n}$ the outward normal vector. Fluid $\Gamma_{i,k}^{\phi}$ and wall $\Gamma_{i,k}^{w}$ boundaries of each sub-cell $\Omega_{i,k}^{\phi}$ are distinguished, such that: $\Gamma_{i,k} = \Gamma_{i,k}^{\phi} \cup \Gamma_{i,k}^{w}$ and $\Gamma_{i,k}^{\phi} \cap \Gamma_{i,k}^{w} = \emptyset$. Summing up over the $N(i)$ fluid sub-cells of the control volume Ω_i, we get the integral formulation:

$$\begin{aligned} meas\left(\Omega_i^{\phi}\right)(\boldsymbol{W}_i(t_1) - \boldsymbol{W}_i(t_0)) + \sum_{k\in\{1,\ldots,N(i)\}} \int_{t_0}^{t_1} \int_{\Gamma_{i,k}^{\phi}\cup\Gamma_{i,k}^{w}} \boldsymbol{F}(\boldsymbol{W}(\mathbf{x}, t)) \cdot \boldsymbol{n} d\Gamma dt \\ = \sum_{k\in\{1,\ldots,N(i)\}} \int_{t_0}^{t_1} \int_{\Omega_{i,k}^{\phi}} \boldsymbol{D}(\boldsymbol{W}(\mathbf{x}, t))\, dx\, dt \end{aligned} \tag{4}$$

3 Time Scheme

The time discretization of the dynamic equation (4) is based on an implicit first order scheme. It is assumed that all numerical fluxes may be evaluated by means of a standard finite volume method, considering one mean value $\mathbf{W_i^n}$ per cell Ω_i at each time t^n. $\mathbf{W_i^n}$ is an approximation of $\boldsymbol{W}_i(t^n)$, and the time step at the nth iteration is: $\Delta t^n = t^{n+1} - t^n$. The numerical algorithm uses a fractional step method, with prediction and correction of the pressure [1, 5]. Each time stepping is divided in three steps: a mass balance step which is used to update the density and to predict the pressure, a momentum balance step where the velocity is updated and an energy balance step that allows to update the total energy and to correct the pressure. The superscript $(\cdot)^{n+1,-}$ states that the variable is implicit for the current step (or known from the last step).

1. *Mass balance*

Pressure and density are implicit, while velocity and entropy are considered frozen at time t^n. The following scheme is set:

$$meas\left(\Omega_i^{\phi}\right) \frac{1}{(c^2)_i^n} \left(P_i^{n+1,-} - P_i^n\right) + \Delta t^n \int_{\Gamma_i^{\phi}} \boldsymbol{Q}^* \cdot \boldsymbol{n} d\Gamma = 0 \tag{5}$$

where: $\left(c^2\right)^n = c^2(P^n, \rho^n)$, and the approximation $\delta P = \left(c^2\right)^n \delta\rho$ is considered, with $\delta P = P^{n+1,-} - P^n$. The approximation of the implicit mass flux $\boldsymbol{Q}^*$ is:

$$\boldsymbol{Q}^* = \boldsymbol{Q}^n - \Delta t^n \nabla P^{n+1,-} \tag{6}$$

and a two-point flux approximation is used: $\int_{\Gamma_i^\phi} \nabla\phi \cdot \boldsymbol{n} d\Gamma = \sum_{j\in V(i)} (\phi_j - \phi_i) S_{ij}^\phi / h_{ij}$, on admissible meshes.

2. *Momentum balance*

In this step, velocity (and momentum) is implicit, whereas density and pressure are known from Eq. (5), and total energy is frozen. Integration of the momentum equation gives:

$$meas\left(\Omega_i^\phi\right)\left(\boldsymbol{Q}_i^{n+1,-} - \boldsymbol{Q}_i^n\right) + \Delta t^n \int_{\Gamma_i^\phi} \left((\boldsymbol{Q}^* \cdot \boldsymbol{n})\boldsymbol{u}\right)^{n+1,-} d\Gamma + \Delta t^n \int_{\Gamma_i} P^{n+1,-} \boldsymbol{n} d\Gamma - \Delta t^n meas\left(\Omega_i^\phi\right) \rho_i^{n+1,-} \boldsymbol{f}_i^{n+1,-} = 0 \tag{7}$$

where: $P_w^{n+1,-}$ is equal to $P_i^{n+1,-}$ for all wall interfaces of cell i. This second step provides the velocity $\boldsymbol{u}^{n+1,-}$ and thus $\boldsymbol{Q}^{n+1,-} = \rho^{n+1,-}\boldsymbol{u}^{n+1,-}$, using:

$$\left(\left(\boldsymbol{Q}^* \cdot \boldsymbol{n}\right)\phi\right)_{ij} = \left(\boldsymbol{Q}^* \cdot \boldsymbol{n}\right)_{ij} \phi_{ij}^{upwind} \tag{8}$$

with: $\phi_{ij}^{upwind} = \beta_{ij}\phi_i + \left(1 - \beta_{ij}\right)\phi_j$, and: $\beta_{ij} = max\left(0, sgn\left(\boldsymbol{Q}^* \cdot \boldsymbol{n}\right)_{ij}\right)$.

3. *Energy balance*

Total energy is implicit while pressure, density and velocity are explicit from the previous steps. Using upwind scheme (8), the total energy $E^{n+1,-}$ is updated:

$$meas\left(\Omega_i^\phi\right)\left(E_i^{n+1,-} - E_i^n\right) + \Delta t^n \int_{\Gamma_i^\phi} \left((\boldsymbol{Q}^* \cdot \boldsymbol{n})\frac{E+P}{\rho}\right)^{n+1,-} d\Gamma - \Delta t^n meas\left(\Omega_i^\phi\right)\left(\rho \boldsymbol{f} \cdot \boldsymbol{u} + \rho\Phi_v\right)_i^{n+1,-} = 0 \tag{9}$$

Property 1 (Positivity of the density and the pressure): *If the initial conditions are such that:* $\rho_i^n > 0$ *and* $P_i^n > 0$ *and the EOS is such that:* $\hat{\gamma} = \rho c^2 / P > 1$. *The density* $\rho_i^{n+1,-}$ *and the pressure* $P_i^{n+1,-}$ *remain positive for all* i, *if the time step* Δt^n *complies with the CFL-like condition* (10)*:*

$$meas\left(\Omega_i^\phi\right) \geq \Delta t^n \sum_{j\in V(i)} \beta_{ij} \left(\frac{\rho_i c_i^2}{P_i}\right)^n \left(\boldsymbol{u}^n \cdot \boldsymbol{n}\right)_{ij} S_{ij}^\phi \tag{10}$$

Sketch of proof Equation (5) gives an invertible linear system: $\mathbf{A}\,\mathbf{P}^{n+1,-} = \mathbf{B}$ with $(\mathbf{A}^{-1})_{ij} \geq 0$, and also $B_i \geq 0$ if and only if condition (10) holds, thus implying $P_i^{n+1,-} \geq 0$. To be conservative, we have set: $\rho_i^{n+1,-} = (c^{-2})_i^n P_i^{n+1,-} + \rho_i^n(\hat{\gamma}_i^n - 1)/\hat{\gamma}_i^n$ which completes the proof: $\rho_i^{n+1,-} \geq 0$, since $\hat{\gamma}_i^n > 1$.

Eventually, the variables are updated: $\rho^{n+1} = \rho^{n+1,-}$, $\boldsymbol{u}^{n+1} = \boldsymbol{u}^{n+1,-}$, $E^{n+1} = E^{n+1,-}$, and $P^{n+1} = P(\rho^{n+1}, \epsilon^{n+1})$, where: $\rho^{n+1}\epsilon^{n+1} = E^{n+1} - 0.5\,\rho^{n+1}\left(\boldsymbol{u}^{n+1}\right)^2$.

4 Numerical Results

Mesh refinement impact

Numerical approximations obtained thanks to this new approach are compared with approximate solutions of the fluid model when the mesh perfectly matches obstacles inside the computational domain (i.e. without any porous control volume). The integral approach is applied on porous meshes so that fluid cells are partially obstructed by obstacles. The numerical example consists in computing the steady flow of a compressible inviscid fluid in a channel aligned with the x-direction. At mid-length, the channel is cluttered by two identical, steady and impermeable tubes aligned with it. A sketch of the test case is displayed on Fig. 2. The two-dimensional computational domain is $\Omega = [0, L] \times [0, h]$. It contains a discontinuous transition interface between a totally fluid area and an obstructed area at $x = \frac{L}{2}$. We consider admissible meshes, with faces aligned with the obstacles. At the inlet and outlet sections of the domain, boundary conditions from the resolution of half Riemann problems are enforced [3] and a steady state is computed. Slip wall boundary conditions are imposed at the top ($y = h$) and bottom ($y = 0$) of the computational domain. The time step is controlled by the CFL-like condition (10). Several numerical approximations of the steady state are obtained using coarse and fine meshes. Six meshes are perfectly adapted to the domain, thus including either totally fluid cells or fully solid cells. They are respectively composed of 24×5, 48×10, 96×20, 192×40, 384×80 and 768×160 regular cells. The four other meshes include porous cells, they are respectively composed of 24×6, 48×12, 96×24 and 192×48 regular cells. We assume that a steady state is reached when the dimensionless time residuals on pressure and velocity in L^2 norm become small enough ($\approx 10^{-7}$, see Fig. 3). The time to steadiness is mainly governed by the velocity time residual. We note P^w the pressure on the intern upstream vertical faces, and S_w the vertical wall surface of these intern upstream faces, such that $S_w = S_{in} - S_{out}$ (see Fig. 2). We define the flux vector $\boldsymbol{\varphi} = [QS, QSH, (QU + P)\, S]^t$ and the head losses vector $\boldsymbol{\Delta} = [0, 0, P^w S_w]^t$, with Q the momentum, S the fluid cross section, H the total enthalpy, U the bulk velocity in the x-direction and P the pressure. When the perfect steady state is reached, the conservation laws provide: $\boldsymbol{\varphi_{in}} = \boldsymbol{\varphi_{out}} + \boldsymbol{\Delta}$. The relative deviation between inlet and outlet boundary values for all the variables is defined as:

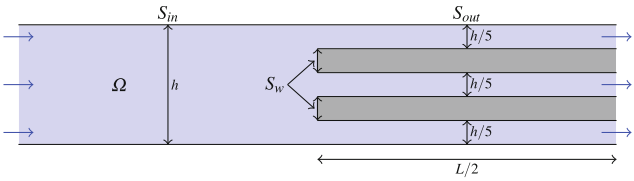


Fig. 2 The Ω domain has a $L \times h$ size, containing two internal obstacles (in *gray*). They lie in the downstream middle of Ω and are spaced of $\frac{h}{5}$ such that $S_{out} + S_w = S_{in}$. The fluid flows from the *left inlet* towards the *right outlet*

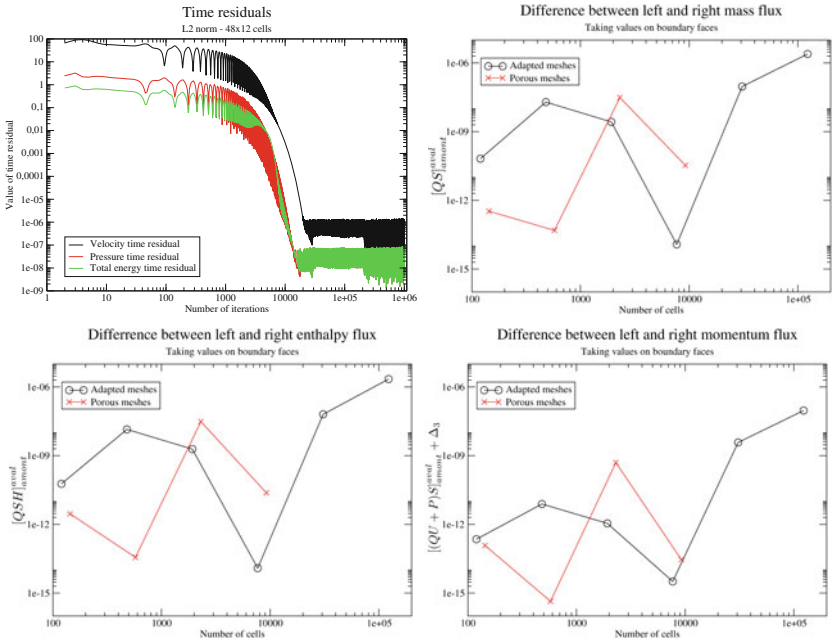


Fig. 3 Time residuals and value of $e(\varphi)$ for the adapted and porous meshes. The adapted meshes correspond to the *black* plots and the porous meshes to the *red* plots

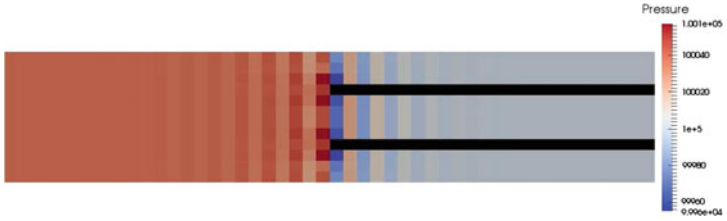


Fig. 4 Pressure field (Pa) for 48 × 12 porous mesh (*black* cells are solid)

$$e(\varphi) = \frac{|\varphi_{in} - (\varphi_{out} + \boldsymbol{\Delta})|}{|\varphi_{in}| + |\varphi_{out}| + |\boldsymbol{\Delta}|}.$$

For each mesh, $e(\varphi)$ is plotted for φ components on Fig. 3. This deviation is small ($\leq 10^{-6}$). When the cells number increases, $e(\varphi)$ may slightly increase since unsteady terms (vortices) appear on refined meshes (Fig. 4).

Mesh adaptation w.r.t. obstacles position: sensibility analysis

The coarsest mesh, composed of 24 × 5 cells, is considered for two sensibility tests. The differences in results between the adapted fluid mesh and any configuration where the bottom of one of the obstacles is slightly shifted off its grid edge are compared. The first configuration, called M_1, corresponds to the mesh adapted to

Table 1 Comparison of M_1 and M_2 configurations, x and y are the cell center coordinates

Variables	x	y	$E^{M_1M_2}$
Density	4.6875	0.1	2.3842×10^{-6}
Pressure	2.60417	0.1	1.6911×10^{-8}

Table 2 Comparison of M_1 and M_3 configurations, x and y are the cell center coordinates

Variables	x	y	$E^{M_1M_3}$
Density	2.8125	0.1	1.5046×10^{-3}
Pressure	2.8125	0.1	2.2050×10^{-3}

Table 3 Comparison of M_2 and M_3 configurations, x and y are the cell center coordinates

Variables	x	y	$E^{M_2M_3}$
Density	2.8125	0.1	1.5046×10^{-3}
Pressure	2.8125	0.1	2.2099×10^{-3}

the tubes position. In the second configuration, M_2, the lower tube width is slightly reduced (10^{-5} h) so that weakly porous cells exist. In the last situation, M_3, the width of the same tube is reduced again at the top (10^{-5} h), and its upstream wall is also slightly shifted in the downstream direction (10^{-5} h). The relative deviation, $E^{M_kM_l}$, between two simulations of different M_k and M_l configurations ($k, l = 1, 2$ or 3) on all N^{ϕ}_{cells} fluid cells for each discrete variable $\varphi_i = (\rho_i, P_i)$, $i \in \{1, \ldots, N^{\phi}_{cells}\}$ (see Tables 1, 2 and 3) is defined as follows:

$$E^{M_kM_l} = \max_{i \in \{1,\ldots,N^{\phi}_{cells}\}} \left| \varphi_i^{M_k} - \varphi_i^{M_l} \right| / \left| \varphi_i^{M_l} \right|.$$

Here the domain measures are: $L = 5$ and $h = 1$. The deviations are rather weak ($\leq 10^{-3}$). The porous formulation is robust w.r.t. standard computations. We note that the gaps are concentrated in the same area, near the upstream faces ($x = 2.5$). They are higher between M_3 and the other configurations. Current work aims at extending the integral formulation to incompressible viscous fluid flows governed by the Navier-Stokes equations. Viscous effects are taken into account thanks to a wall function which vanishes when the mesh is refined [2].

Acknowledgements The first author receives a financial support by ANRT through an EDF-CIFRE contract 2016/0728.

References

1. Archambeau, F., Hérard, J.M., Laviéville, J.: Comparative study of pressure-correction and Godunov-type schemes on unsteady compressible cases. Comput. Fluids **38**, 1495–1509 (2009)
2. Colas, C., Ferrand, M., Hérard, J.M., Le Coupanec, E.: Approche intégrale pour la modélisation des écoulements en milieux encombrés - Prise en compte des effets visqueux. Note interne 6125-3013-2016-17220-FR, EDF R&D (2016)
3. Dubois, F.: Boundary conditions and the Osher scheme for the Euler equations of gas dynamics. Internal Report CMAP 170, Ecole Polytechnique, Palaiseau, France (1987)
4. EDF R&D. http://code-saturne.org/cms/sites/default/files/docs/4.2/theory.pdf: Code_Saturne 4.2.0 Theory Guide (2015)
5. Ferrand, M., Hérard, J.M., Le Coupanec, E., Martin, X.: Schémas implicites dans une formulation intégrale pour la prise en compte d'obstacles immergés dans un fluide compressible. Note interne H-I83-2015-05276-FR, EDF R&D (2015)
6. Hérard, J.M., Martin, X.: An integral approach to compute compressible fluid flows in domains containing obstacles. Int. J. Finite Vol. **12**(1), 1–39 (2015)
7. Le Coq, G., Aubry, S., Cahouet, J., Lequesne, P., Nicolas, G., Pastorini, S.: The THYC computer code. A finite volume approach for 3 dimensional two-phase flows in tube bundles. Bulletin de la Direction des études et recherches - Électricité de France, pp. 61–76 (1989)
8. Martin, X.: Modélisation d'écoulements fluides en milieu encombré d'obstacles. Ph.D. thesis. Aix-Marseille Université (2015). https://tel.archives-ouvertes.fr/tel-01235089/
9. Toumi, I., Bergeron, A., Gallo, D., Royer, E., Caruge, D.: FLICA-4: A three-dimensional two-phase flow computer code with advanced numerical methods for nuclear applications. Nucl. Eng. Des. **200**, 139–155 (2000)

A High-Order Discontinuous Galerkin Lagrange Projection Scheme for the Barotropic Euler Equations

Christophe Chalons and Maxime Stauffert

Abstract This work considers the barotropic Euler equations and proposes a high-order conservative scheme based on a Lagrange-Projection decomposition. The high-order in space and time are achieved using Discontinuous Galerkin (DG) and Runge-Kutta (RK) strategies. The use of a Lagrange-Projection decomposition enables the use of time steps that are not constrained by the sound speed thanks to an implicit treatment of the acoustic waves (Lagrange step), while the transport waves (Projection step) are treated explicitly. We compare our DG discretization with the recent one (Renac in Numer Math 1-27, 2016, [7]) and state that it satisfies important non linear stability properties. The behaviour of our scheme is illustrated by several test cases.

Keywords Barotropic Euler equations · High-order discontinuous Galerkin schemes · Lagrange-projection decomposition · Implicit explicit · Large time steps

MSC (2010): 35L40 · 35Q35 · 65M12 · 65M60

1 Introduction

We are interested in the gas dynamics equations in Eulerian coordinates

$$\begin{cases} \partial_t \rho + \partial_x(\rho u) = 0, \\ \partial_t(\rho u) + \partial_x\left(\rho u^2 + p\right) = 0, \end{cases} \tag{1}$$

C. Chalons · M. Stauffert (✉)
Laboratoire de Mathématiques de Versailles, UVSQ, CNRS,
Université Paris-Saclay, 45 avenue des États-Unis, 78035 Versailles, France
e-mail: maxime.stauffert@uvsq.fr

C. Chalons
e-mail: christophe.chalons@uvsq.fr

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_7

where $\rho > 0$ is the density, u the velocity and $p = p(\rho)$ is the pressure of the fluid such that $p'(\rho) > 0$. In the numerical experiments, we will choose $p(\rho) = g\rho^2/2$ where $g > 0$ is the gravity constant so that the model can also be understood as the Shallow-Water equations with flat topography (in this case, ρ stands for the water depth). The unknowns depend on the space and time variables x and t, with $x \in \mathbb{R}$ and $t \in [0, \infty)$. At time $t = 0$, the model is supplemented with a given initial data $\rho(x, t = 0) = \rho_0(x)$ and $u(x, t = 0) = u_0(x)$.

The aim of this paper is to propose a high-order discretization based on a Lagrange-Projection decomposition of the governing equations and using a Discontinuous Galerkin (DG) [4, 9] strategy for the space variable.

The Lagrange-Projection (or equivalently Lagrange-Remap) decomposition is interesting since it allows to naturally decouple the acoustic and transport terms of the model. It proved to be useful and very efficient when considering subsonic or low-Mach number flows. In this case, the CFL restriction of Godunov-type schemes is driven by the acoustic waves and can be very restrictive. The Lagrange-Projection strategy allows for a very natural implicit-explicit scheme with a CFL restriction based on the (slow) transport waves and not on the (fast) acoustic waves. We refer for instance the reader to [1, 2, 5], to the recent contribution [3], and to the references therein. Note that the later contribution considers the Shallow-Water equations with non flat topography and that the corresponding (implicit-explicit) Lagrange-Projection scheme is well-balanced but only first-order accurate. It is the purpose of this contribution to extend the first-order Lagrange-Projection schemes of the above references to high-order of accuracy in both space and time. The proposed approach is quite close to the one recently developed in [7], but as we will see, the corresponding Projection step turns out to be different.

2 Lagrange-Projection Decomposition and Finite-Volume Scheme

In this section, we briefly present the Lagrange-Projection decomposition considered in this paper and the corresponding first-order finite volume scheme.

Operator splitting decomposition and relaxation approximation. Using the chain rule for the space derivatives of (1), the Lagrange-Projection decomposition consists in first solving

$$\begin{cases} \partial_t \rho + \rho \partial_x u = 0, \\ \partial_t (\rho u) + \rho u \partial_x u + \partial_x p = 0, \end{cases} \tag{2}$$

which gives in Lagrangian coordinates $\tau \partial_x = \partial_m$, with $\tau = 1/\rho$,

$$\begin{cases} \partial_t \tau - \partial_m u = 0, \\ \partial_t u + \partial_m p = 0, \end{cases} \tag{3}$$

and then the transport system

$$\begin{cases} \partial_t \rho + u \partial_x \rho = 0, \\ \partial_t (\rho u) + u \partial_x (\rho u) = 0. \end{cases} \tag{4}$$

The numerical approximation of (3) and (4) will be given in the next sections but let us notice from now on that the Lagrangian system (3) will be treated considering the following relaxation approximation [6, 8],

$$\begin{cases} \partial_t \tau - \partial_m u = 0, \\ \partial_t u + \partial_m \Pi = 0, \\ \partial_t \Pi + a^2 \partial_m u = \lambda \left(p - \Pi \right). \end{cases} \tag{5}$$

Here, the new variable Π represents a linearization of the real pressure p, the constant parameter a is a linearization of the Lagrangian sound speed ρc such that the sub-characteristic condition $a > \rho c$, $c = \sqrt{p'(\rho)}$, is satisfied, and the relaxation parameter λ allows to recover $\Pi = p$ and the original system (3) in the asymptotic regime $\lambda \to \infty$. As usual, the relaxation system will be solved using a splitting strategy which consists in first setting $\Pi = p$ at initial time (which is formally equivalent to considering $\lambda \to \infty$ in (5)), and then solving the relaxation system (5) with $\lambda = 0$.
First-order numerical scheme. The first-order finite volume scheme associated with the above decomposition and relaxation approximation is classical and given for instance in [2]. Nevertheless, it will be recovered in the DG extension proposed in the next section by setting the degree of all polynomials p to 0. Space and time will be discretized using a space step Δx and a time step Δt. We will consider a set of cells $\kappa_j = [x_{j-1/2}, x_{j+1/2})$ and instants $t^n = n\Delta t$, where $x_{j+1/2} = j\Delta x$ and $x_j = (x_{j-1/2} + x_{j+1/2})/2$ are respectively the cell interfaces and cell centers, for $j \in \mathbb{Z}$ and $n \in \mathbb{N}$.

3 Discontinuous Galerkin Discretization

We begin this section by introducing the notations of the DG discretization. Recall that the DG approach considers that the approximate solution at each time t^n is defined on each cell κ_j by a polynomial in space of order less or equal than p for a given integer $p \geq 1$ ($p = 0$ corresponds to the usual first-order and piecewise constant finite volume scheme). With this in mind, we consider the $(p + 1)$ Lagrange polynomials $\{\ell_i\}_{i=0,\dots,p}$ associated with the Gauss-Lobatto quadrature points in $[-1, 1]$. More precisely, denoting $-1 = s_0 < s_1 < \cdots < s_p = 1$ the $p + 1$ Gauss-Lobatto quadrature points, ℓ_i is defined by the relations $\ell_i(s_k) = \delta_{i,k}$ for $k = 0, \dots, p$, where δ is the Kronecker symbol. Then, in each cell κ_j, we define the shifted Lagrange polynomials $\Phi_{i,j}$ by $\Phi_{i,j}(x) = \ell_i\left(\frac{2}{\Delta x}(x - x_j)\right)$ and we take $\{\Phi_{i,j}\}_{i=0,\dots,p}$ as a basis for polynomials of order less or equal than p on κ_j. If we denote by $X_{\Delta x}$ the DG approximation of X, we thus have $X_{\Delta x}(x, t) = \sum_{k=0}^{p} X_{k,j}(t)\, \Phi_{k,j}(x), \quad \forall x \in \kappa_j$, where the coef-

ficients $X_{k,j}$ depend on time and correspond to the value of the numerical solution at the shifted Gauss-Lobatto quadrature points $x_{k,j} = x_j + \frac{\Delta x}{2} s_k$.

Before entering the details of the numerical approximation, let us briefly recall that the Gauss-Lobatto quadrature formula for $f : \kappa_j \times \mathbb{R}^+ \to \mathbb{R}$ writes

$$\int_{\kappa_j} f(x,t)\,\mathrm{d}x \approx \frac{\Delta x}{2} \sum_{k=0}^{p} \omega_k f(x_{k,j}, t),$$

where ω_k are the weights of the Gauss-Lobatto quadrature. It is well-known that this formula is exact as soon as f is a polynomial of order less or equal than $(2p-1)$ with respect to x on κ_j. Just note that the integral $\int_{\kappa_j} \Phi_{i,j}(x)\Phi_{k,j}(x)\,\mathrm{d}x$ will be therefore approximated by $\frac{\Delta x}{2}\omega_i \delta_{i,k}$ in the following. At last, note that the piecewise constant case $p = 0$ can be also considered in this framework provided that we set $s_0 = 0$, $\Phi_{0,j} = 1$ and $\omega_0 = 2$.

Time discretization ($t^n \to t^{n+1}$). We begin with the acoustic step (5) with $\lambda = 0$. Multiplying the three equations by $\Phi_{i,j}$, integrating over κ_j, and considering the piecewise polynomial approximations $X_{\Delta x}$ for $X = \tau, u, \Pi$ easily leads to

$$\begin{cases} \dfrac{\Delta x}{2}\omega_i \partial_t \tau_{i,j}(t) - \displaystyle\int_{\kappa_j} \Phi_{i,j}(x)\partial_m u(x,t)\,\mathrm{d}x = 0, \\ \dfrac{\Delta x}{2}\omega_i \partial_t u_{i,j}(t) + \displaystyle\int_{\kappa_j} \Phi_{i,j}(x)\partial_m \Pi(x,t)\,\mathrm{d}x = 0, \\ \dfrac{\Delta x}{2}\omega_i \partial_t \Pi_{i,j}(t) + a^2 \displaystyle\int_{\kappa_j} \Phi_{i,j}(x)\partial_m u(x,t)\,\mathrm{d}x = 0, \end{cases}$$

that we discretize in time by

$$\begin{cases} \tau_{i,j}^{n+1^-} = \tau_{i,j}^n + \dfrac{2\Delta t}{\omega_i \Delta x} \displaystyle\int_{\kappa_j} \Phi_{i,j}(x)\partial_m u(x,t^\alpha)\,\mathrm{d}x, \\ u_{i,j}^{n+1^-} = u_{i,j}^n - \dfrac{2\Delta t}{\omega_i \Delta x} \displaystyle\int_{\kappa_j} \Phi_{i,j}(x)\partial_m \Pi(x,t^\alpha)\,\mathrm{d}x, \\ \Pi_{i,j}^{n+1^-} = \Pi_{i,j}^n - a^2 \dfrac{2\Delta t}{\omega_i \Delta x} \displaystyle\int_{\kappa_j} \Phi_{i,j}(x)\partial_m u(x,t^\alpha)\,\mathrm{d}x, \end{cases} \tag{6}$$

where the superscript $n+1^-$ formally represents the fictitious time t^{n+1^-}, and $\alpha = n$ or $\alpha = n+1^-$ if the time discretization is taken to be explicit or implicit.

As far as the transport step is concerned, the same process of reasoning leads to

$$\begin{cases} \rho_{i,j}^{n+1} = \rho_{i,j}^{n+1^-} - \dfrac{2\Delta t}{\omega_i \Delta x} \displaystyle\int_{\kappa_j} \Phi_{i,j}(x) u(x,t^\alpha)\partial_x \rho(x,t^{n+1^-})\,\mathrm{d}x, \\ (\rho u)_{i,j}^{n+1} = (\rho u)_{i,j}^{n+1^-} - \dfrac{2\Delta t}{\omega_i \Delta x} \displaystyle\int_{\kappa_j} \Phi_{i,j}(x) u(x,t^\alpha)\partial_x (\rho u)(x,t^{n+1^-})\,\mathrm{d}x. \end{cases} \tag{7}$$

Note that this transport step is always treated explicitly in time.

Volume integrals and flux calculations. Considering the acoustic step, we aim at approximating the integrals $\int_{\kappa_j} \Phi_{i,j}(x)\partial_m X(x,t^\alpha)\,\mathrm{d}x$ with $X = u, \Pi$. Observe that

$$\int_{\kappa_j} \Phi_{i,j}(x)\partial_m X(x,t^\alpha)\,\mathrm{d}x \approx \frac{\Delta x}{2}\omega_i \tau_{i,j}^n \partial_x X(x_{i,j},t^\alpha)\,\mathrm{d}x = \tau_{i,j}^n \int_{\kappa_j} \Phi_{i,j}(x)\partial_x X(x,t^\alpha)\,\mathrm{d}x,$$

the last equality is indeed exact since X and Φ are polynomials of order less or equal than p, so that $\Phi_{i,j}\partial_x X(\cdot,t)$ is of order less or equal than $(2p-1)$. The objective is now to use one integration by part to move the derivative from X to Φ, and to use the numerical fluxes to evaluate the interfacial terms, which gives

$$\int_{\kappa_j} \Phi_{i,j}(x)\partial_x X(x,t^\alpha)\,\mathrm{d}x \approx \delta_{i,p} X_{j+1/2}^{*,\alpha} - \delta_{i,0} X_{j-1/2}^{*,\alpha} - \frac{\Delta x}{2}\sum_{k=0}^{p} \omega_k X_{k,j}^\alpha \partial_x \Phi_{i,j}(x_{k,j}).$$

Again, we refer the reader to [2] for the expressions of the star quantities in the above formula and the following ones, which are nothing but the numerical fluxes of the first-order finite volume scheme. At last, from (6) we obtain the acoustic step

$$\begin{cases} \tau_{i,j}^{n+1^-} = \tau_{i,j}^n + \frac{2\Delta t}{\omega_i \Delta x}\tau_{i,j}^n \left[\delta_{i,p} u_{j+1/2}^{*,\alpha} - \delta_{i,0} u_{j-1/2}^{*,\alpha} - \frac{\Delta x}{2}\sum_{k=0}^{p}\omega_k u_{k,j}^\alpha \partial_x \Phi_{i,j}(x_{k,j})\right] \\ \qquad\qquad = L_{i,j}^\alpha \tau_{i,j}^n, \\ u_{i,j}^{n+1^-} = u_{i,j}^n - \frac{2\Delta t}{\omega_i \Delta x}\tau_{i,j}^n \left[\delta_{i,p} \Pi_{j+1/2}^{*,\alpha} - \delta_{i,0}\Pi_{j-1/2}^{*,\alpha} - \frac{\Delta x}{2}\sum_{k=0}^{p}\omega_k \Pi_{k,j}^\alpha \partial_x \Phi_{i,j}(x_{k,j})\right], \\ \Pi_{i,j}^{n+1^-} = \Pi_{i,j}^n - a^2\frac{2\Delta t}{\omega_i \Delta x}\tau_{i,j}^n \left[\delta_{i,p} u_{j+1/2}^{*,\alpha} - \delta_{i,0} u_{j-1/2}^{*,\alpha} - \frac{\Delta x}{2}\sum_{k=0}^{p}\omega_k u_{k,j}^\alpha \partial_x \Phi_{i,j}(x_{k,j})\right], \end{cases} \tag{8}$$

with $L_{i,j}^\alpha = 1 + \frac{2\Delta t}{\omega_i \Delta x}\left[\delta_{i,p} u_{j+1/2}^{*,\alpha} - \delta_{i,0} u_{j-1/2}^{*,\alpha} - \frac{\Delta x}{2}\sum_{k=0}^{p}\omega_k u_{k,j}^\alpha \partial_x \Phi_{i,j}(x_{k,j})\right]$.

Regarding the transport step, we want to evaluate the integrals

$$\int_{\kappa_j} \Phi_{i,j}(x) u(x,t^\alpha)\partial_x X(x,t^{n+1^-})\,\mathrm{d}x$$

with $X = \rho, \rho u$. The same process as before leads to

$$\int_{\kappa_j} \Phi_{i,j}(x) u(x,t^\alpha)\partial_x X(x,t^{n+1^-})\,\mathrm{d}x = \delta_{i,p} X_{j+1/2}^{*,n+1^-} u_{j+1/2}^{*,\alpha} - \delta_{i,0} X_{j-1/2}^{*,n+1^-} u_{j-1/2}^{*,\alpha} - \int_{\kappa_j} (Xu)\partial_x \Phi_{i,j}\,\mathrm{d}x - X_{i,j}^{n+1^-}\int_{\kappa_j} \Phi_{i,j}(x)\partial_x u(x,t^\alpha)\,\mathrm{d}x,$$

where we take

$\int_{\kappa_j} \Phi_{i,j}\partial_x u(x,t^\alpha)\,\mathrm{d}x = \delta_{i,p} u_{j+1/2}^{*,\alpha} - \delta_{i,0} u_{j-1/2}^{*,\alpha} - \frac{\Delta x}{2}\sum_{k=0}^{p}\omega_k u_{k,j}^\alpha \partial_x \Phi_{i,j}(x_{k,j})$

and $\int_{\kappa_j} (Xu)\partial_x \Phi_{i,j}\,\mathrm{d}x \approx \frac{\Delta x}{2}\sum_{k=0}^{p}\omega_k X_{k,j}^{n+1^-} u_{k,j}^\alpha \partial_x \Phi_{i,j}(x_{k,j})$.

Conservativity property and mean values. Easy calculations not reported here show that the whole Lagrange-Projection scheme can be written as follows

$$\rho_{i,j}^{n+1} = \rho_{i,j}^{n} - \frac{2\Delta t}{\omega_i \Delta x}\left[\delta_{i,p}\rho_{j+1/2}^{*,n+1^-}u_{j+1/2}^{*,\alpha} - \delta_{i,0}\rho_{j-1/2}^{*,n+1^-}u_{j-1/2}^{*,\alpha} - \frac{\Delta x}{2}\sum_{k=0}^{p}\omega_k \rho_{k,j}^{n+1^-}u_{k,j}^{\alpha}\partial_x \Phi_{i,j}(x_{k,j})\right],$$

$$(\rho u)_{i,j}^{n+1} = (\rho u)_{i,j}^{n} - \frac{2\Delta t}{\omega_i \Delta x}\left[\delta_{i,p}\Pi_{j+1/2}^{*,\alpha} - \delta_{i,0}\Pi_{j-1/2}^{*,\alpha} - \frac{\Delta x}{2}\sum_{k=0}^{p}\omega_k \Pi_{k,j}^{n+1^-}\partial_x \Phi_{i,j}(x_{k,j})\right]$$

$$-\frac{2\Delta t}{\omega_i \Delta x}\left[\delta_{i,p}(\rho u)_{j+1/2}^{*,n+1^-}u_{j+1/2}^{*,\alpha} - \delta_{i,0}(\rho u)_{j-1/2}^{*,n+1^-}u_{j-1/2}^{*,\alpha} - \frac{\Delta x}{2}\sum_{k=0}^{p}\omega_k (\rho u)_{k,j}^{n+1^-}u_{k,j}^{\alpha}\partial_x \Phi_{i,j}(x_{k,j})\right]$$

while the mean values $\overline{X}_j^{n+1} = \frac{1}{\Delta x}\int_{\kappa_j} X(x, t^{n+1})\,\mathrm{d}x = \sum_{i=0}^{p}\frac{\omega_i}{2}X_{i,j}^{n+1}$ with $X = \rho, \rho u$ obey the conservative formulas

$$\begin{cases} \overline{\rho}_j^{n+1} = \overline{\rho}_j^{n} - \dfrac{\Delta t}{\Delta x}\left[\rho_{j+1/2}^{*,n+1^-}u_{j+1/2}^{*,\alpha} - \rho_{j-1/2}^{*,n+1^-}u_{j-1/2}^{*,\alpha}\right], \\ \overline{(\rho u)}_j^{n+1} = \overline{(\rho u)}_j^{n} - \dfrac{\Delta t}{\Delta x}\left[\Pi_{j+1/2}^{*,\alpha} + (\rho u)_{j+1/2}^{*,n+1^-}u_{j+1/2}^{*,\alpha}\right. \\ \qquad\qquad\qquad\qquad \left. -\Pi_{j-1/2}^{*,\alpha} - (\rho u)_{j-1/2}^{*,n+1^-}u_{j-1/2}^{*,\alpha}\right]. \end{cases} \tag{9}$$

Additional nonlinear stability properties can be proved for both the implicit and explicit schemes ($\alpha = n$ and $\alpha = n + 1^-$). In particular, we have been able to prove the positivity of the nodal densities $\rho_{i,j}^{n+1^-}$ at time t^{n+1^-} and of the mean densities $\overline{\rho}_j^{n+1}$ at time t^{n+1}, but also the validity of a discrete entropy inequality for the mean energy following the same lines as in [7].

Comparison with the double integration by part used in [7]. The present scheme turns out to be very close to the one recently proposed in [7], and it shares the same stability properties. However, the overall process in [7] is based on double integrations by part leading to the use of both numerical and exact fluxes at the interfaces, instead of only numerical fluxes in our approach. Interestingly, we observed that both schemes are strictly equivalent if one considers the mean values, but the nodal values turn out to be different because of the transport step. These little differences are due to the use of quadrature formulas to integrate the polynomials $Xu\partial_x \Phi_{i,j}$. In this case, the numerical integrations are not exact since polynomials $Xu\partial_x \Phi_{i,j}$ are of order $3p - 1 > 2p - 1$.

Positivity and generalized slope limiters. We have already stated the positivity of the nodal values $\rho_{i,j}^{n+1^-}$ at the end of the acoustic step and of the mean values $\overline{\rho}_j^{n+1}$ at the end of the transport step. Similarly to [7], we suggest to use a positivity limiter to ensure that $\rho_{i,j}^{n+1} > 0$. More precisely, we replace $\rho_{i,j}^{n+1}$ by $\theta_j \rho_{i,j}^{n+1} + \left(1 - \theta_j\right)\overline{\rho}_j^{n+1}$, where the coefficients θ_j are taken to be $\theta_j = \min\left(1, \frac{\overline{\rho}_j^{n+1} - \varepsilon}{\overline{\rho}_j^{n+1} - \min_i \rho_{i,j}^{n+1}}\right)$. This formula ensures that if ρ is less than the threshold ε, the nodal values of the corresponding cell are corrected, using the positive mean value, towards values greater than ε. In general we set the parameter ε to $1.0\mathrm{e}^{-10}$. Note that in the forthcoming numerical experiments, the positivity limiter is not active. In order to avoid non physical oscil-

lations, we also use the generalized slope limiters introduced in [4]. More precisely, considering the *minmod* function $m(a, b, c) = s \cdot \min(|a|, |b|, |c|)$ if $s = \text{sign}(a) = \text{sign}(b) = \text{sign}(c)$ and 0 otherwise, the increments $\Delta_+\overline{X}_j^{n+1} = \overline{X}_{j+1}^{n+1} - \overline{X}_j^{n+1}$, $\Delta_-\overline{X}_j^{n+1} = \overline{X}_j^{n+1} - \overline{X}_{j-1}^{n+1}$, and the values

$$X_{j+1/2}^{-,n+1} = \overline{X}_j^{n+1} + m\left(X_{p,j}^{n+1} - \overline{X}_j^{n+1}, \Delta_+\overline{X}_j^{n+1}, \Delta_-\overline{X}_j^{n+1}\right),$$
$$X_{j-1/2}^{+,n+1} = \overline{X}_j^{n+1} - m\left(\overline{X}_j^{n+1} - X_{0,j}^{n+1}, \Delta_+\overline{X}_j^{n+1}, \Delta_-\overline{X}_j^{n+1}\right),$$

the new states at time t^{n+1} are defined by

$$\begin{cases} X_{i,j}^{n+1} & \text{if } X_{j+1/2}^{-,n+1} = X_{p,j}^{n+1} \text{ and } X_{j-1/2}^{+,n+1} = X_{0,j}^{n+1}, \\ \overline{X}_j^{n+1} + \frac{2}{\Delta x}\left(x_{i,j} - x_j\right) \cdot m\left(\partial_x X^{n+1}(x_j), \Delta_+\overline{X}_j^{n+1}, \Delta_-\overline{X}_j^{n+1}\right) & \text{otherwise.} \end{cases}$$

4 Numerical Results

The aim of this section is to compare our explicit-explicit EXEX_p and implicit-explicit IMEX_p Lagrange-Projection schemes, where p refers to the polynomial order of the DG approach. The time integrations are performed using Strong Stability Preserving Runge-Kutta methods described in [4]. Recall that $p(\rho) = g\rho^2/2$ so that the parameter a is chosen locally at each interface according to $a_{j+1/2} = \kappa \max\left(\rho_j^n\sqrt{g\rho_j^n}, \rho_{j+1}^n\sqrt{g\rho_{j+1}^n}\right)$ with $\kappa = 1.01$ and $g = 9.81$. We set $\Delta t = \min(\Delta t_{\text{Lag}}, \Delta t_{\text{Tra}})$ for the EXEX_p schemes and $\Delta t = \Delta t_{\text{Tra}}$ for the IMEX_p schemes where $\Delta t_{\text{Lag}} = \frac{\Delta x}{2p+1}\min_j\left(2a_{j+1/2}\min(\tau_{p,j}, \tau_{0,j+1})\right)$ is the DG time-step restriction associated with the Lagrangian step, while the Transport step CFL restriction reads $\Delta t_{\text{Tra}} = \Delta x \min_{i,j}\frac{2}{\omega_i}\left(\int_{\kappa_j} u^\alpha \partial_x \Phi_{i,j}\,\mathrm{d}x - \delta_p u_{j+1/2}^{*,\alpha,-} + \delta_0 u_{j-1/2}^{*,\alpha,+}\right)$.

Manufactured smooth solution. This preliminary test case is taken from [7] and allows us to test the experimental order of accuracy (EOA) of the schemes, especially on the Transport step. The space domain is [0, 1], the boundary conditions are periodic and the initial conditions are $\rho_0(x) = 1 + 0.2\sin(2\pi x)$ and $u_0(x) = 1$. We solve (1) with a source term such that the exact solution is $\rho(x, t) = 1 + 0.2\sin(2\pi(x - t))$ and $u(x, t) = 1$, which just means that we impose $u_{i,j}^{n+1^-} = 1$ and $\Pi_{i,j}^{n+1^-} = \Pi_{i,j}^n$, so that the Acoustic step is trivial. Note that we use in this special case the EXEX_p schemes. The EOA are reported in Table 1.

Dam break problem. In this test case, we take $\rho_0(x) = 20$ if $x \in [0, 750[$, $\rho_0(x) = 10$ if $x \in\ ]750, 1500]$, and $u_0 = 0$ everywhere. The solutions given by the EXEX_p and IMEX_p schemes with $p = 0$, 1 and 2 are shown on Fig. 1 using a 100-cell mesh, and compared with the classical first-order HLL scheme over a 100-cell mesh and a reference 1000-cell refined mesh. Note that the slope limiters allow to reduce spurious oscillations, but there is still a little undershoot for the EXEX_1 scheme.

Table 1 EOA for the manufactured smooth solution at time $T = 0.5$

Δx	$p = 0$		$p = 1$		$p = 2$	
	$\|\rho_{\Delta x} - \rho\|_1$	order	$\|\rho_{\Delta x} - \rho\|_1$	order	$\|\rho_{\Delta x} - \rho\|_1$	order
1/512	9.3986E-03	0.9432	1.0196E-05	1.9996	1.3457E-08	2.9907
1/1024	4.7945E-03	0.9710	2.5493E-06	1.9998	1.6849E-09	2.9977
1/2048	2.4217E-03	0.9854	6.3736E-07	1.9999	2.1070E-10	2.9994

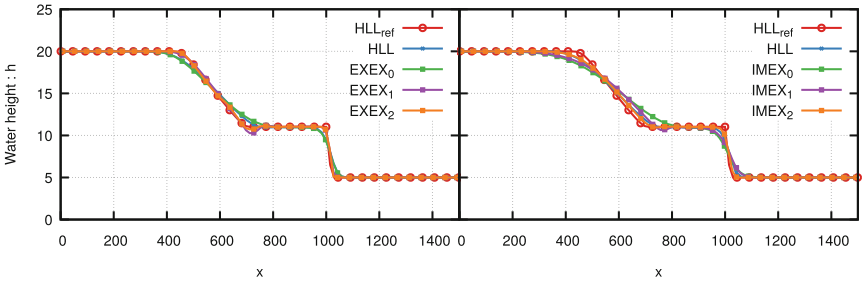


Fig. 1 Dam Break problem, water height at time $T = 10$, EXEX_p (*left*), IMEX_p (right)

Acknowledgements The authors are very grateful to P. Kestener, S. Kokh and F. Renac for stimulating discussions, and the "Maison de la Simulation" for providing excellent working conditions to the second author.

References

1. Chalons, C., Girardin, M., Kokh, S.: Large time step and asymptotic preserving numerical schemes for the gas dynamics equations with source terms. SIAM J. Sci. Comput. **35**(6), A2874–A2902 (2013)
2. Chalons, C., Girardin, M., Kokh, S.: An all-regime Lagrange-Projection like scheme for the gas dynamics equations on unstructured meshes. Commun. Comput. Phys. **20**(01), 188–233 (2016)
3. Chalons, C., Kestener, P., Kokh, S., Stauffert, M.: A large time-step and well-balanced Lagrange-Projection type scheme for the shallow-water equations. Communic. Math. Sci. **15**(3), 765–788 (2017)
4. Cockburn, B., Shu, C.W.: Runge-Kutta discontinuous Galerkin methods for convection-dominated problems. J. Sci. Comput. **16**(3), 173–261 (2001)
5. Coquel, F., Nguyen, Q., Postel, M., Tran, Q.: Entropy-satisfying relaxation method with large time-steps for Euler IBVPs. Math. Comput. **79**, 1493–1533 (2010)
6. Jin, S., Xin, Z.P.: The relaxation schemes for systems of conservation laws in arbitrary space dimension. Comm. Pure Appl. Math. **48**(3), 235–276 (1995)
7. Renac, F.: A robust high-order Lagrange-Projection like scheme with large time steps for the isentropic Euler equations. Numer. Math., 1–27 (2016)
8. Suliciu, I.: On the thermodynamics of fluids with relaxation and phase transitions. Fluids with relaxation. Internat. J. Engrg. Sci **36**, 921–947 (1998)
9. Xing, Y., Zhang, X., Shu, C.W.: Positivity-preserving high order well-balanced discontinuous Galerkin methods for the shallow water equations. Adv. Water Resour. **33**(12), 1476–1493 (2010)

Sensitivity Analysis for the Euler Equations in Lagrangian Coordinates

Christophe Chalons, Régis Duvigneau and Camilla Fiorini

Abstract Sensitivity analysis (SA) is the study of how changes in the inputs of a model affect the outputs. SA has many applications, among which uncertainty quantification, quick evaluation of close solutions, and optimization. Standard SA techniques for PDEs, such as the continuous sensitivity equation method, call for the differentiation of the state variable. However, if the governing equations are hyperbolic PDEs, the state can be discontinuous and this generates Dirac delta functions in the sensitivity. The aim of this work is to define and approximate numerically a system of sensitivity equations which is valid also when the state is discontinuous: to do that, one can define a correction term to be added to the sensitivity equations starting from the Rankine-Hugoniot conditions, which govern the state across a shock. We show how this procedure can be applied to the Euler barotropic system with different finite volumes methods.

Keywords Sensitivity analysis · $p-$system · Dirac delta function · Finite volume

MSC (2010) : 35L60 · 65M08 · 49Q12

C. Chalons · C. Fiorini (✉)
Laboratoire de Mathématiques de Versailles, UVSQ, CNRS, Université Paris-Saclay,
45 Avenue des États-Unis, 78035 Versailles, France
e-mail: camilla.fiorini@uvsq.fr

C. Chalons
e-mail: christophe.chalons@uvsq.fr

R. Duvigneau
Université Côte D'Azur, INRIA, CNRS, LJAD, INRIA Sophia-Antipolis Méditerranée Center,
ACUMES Project-Team, 2004 Route des Lucioles - B.P. 93, 06902 Sophia Antipolis, France
e-mail: regis.duvigneau@inria.fr

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_8

1 Introduction

Sensitivity analysis (SA) concerns the quantification of changes in Partial Differential Equations (PDEs) solution due to perturbations in the model input. It has been a topic of active research for the last years, due to its many applications, for instance in uncertainty quantification, quick evaluation of close solutions [4], and optimization [2], to name but a few. Note that SA approaches differ from adjoint methods, which are restricted to the evaluation of functional derivatives. Standard SA methods work only under certain hypotheses of regularity of the solution $\mathbf{U}$ [1]. These assumptions are not verified in the case of hyperbolic systems of the general form

$$\begin{cases} \partial_t \mathbf{U} + \partial_x \mathbf{F}(\mathbf{U}) = 0, \quad x \in \mathbb{R}, \quad t > 0, \\ \mathbf{U}(x, 0) = \mathbf{U}_0(x), \end{cases}$$

due to possible discontinuities, which can occur even when the initial condition is smooth. If the state $\mathbf{U}$ is discontinuous, Dirac delta functions will appear in the sensitivity $\mathbf{U}_a = \partial_a \mathbf{U}$. Here and throughout this work, a denotes the input parameter of the model which may vary and induce a non trivial sensitivity $\mathbf{U}_a$ of the state solution $\mathbf{U}$.

In this work, we consider the Euler equations in Lagrangian coordinates in a barotropic case, i.e. the p-system:

$$\begin{cases} \partial_t \tau - \partial_x u = 0, \\ \partial_t u + \partial_x p(\tau) = 0, \end{cases} \tag{1}$$

where $\tau > 0$ is the covolume (i.e. $\tau = \mathbf{F}\frac{1}{\rho}$, where ρ is the density of the fluid), u is the velocity and the pressure $p(\tau)$ is a function of τ such that $p'(\tau) < 0$ and $p''(\tau) > 0$. Under these assumptions, (1) is strictly hyperbolic with eigenvalues $\lambda_\pm = \pm c$ where $c = \sqrt{-p'(\tau)}$ is the Lagrangian sound speed. In this work we will consider $p(\tau) = \tau^{-\gamma}$, where $\gamma = 1.4$ is the heat capacity ratio.
By differentiating the system (1) with respect to the parameter of interest a and considering *smooth* solutions of (1), we obtain the sensitivity equations:

$$\begin{cases} \partial_t \tau_a - \partial_x u_a = 0, \\ \partial_t u_a + \partial_x (p'(\tau)\tau_a) = 0. \end{cases} \tag{2}$$

One can define the following vectors:

$$\mathbf{U} = \begin{bmatrix} \tau \\ u \end{bmatrix}, \qquad \mathbf{F}(\mathbf{U}) = \begin{bmatrix} -u \\ p(\tau) \end{bmatrix}, \qquad \mathbf{U}_a = \begin{bmatrix} \tau_a \\ u_a \end{bmatrix}, \qquad \mathbf{F}_a(\mathbf{U}, \mathbf{U}_a) = \begin{bmatrix} -u_a \\ p'(\tau)\tau_a \end{bmatrix},$$

and rewrite the systems (1) and (2) in a vectorial form:

$$\begin{cases} \partial_t \mathbf{U} + \partial_x \mathbf{F}(\mathbf{U}) = 0, \\ \partial_t \mathbf{U}_a + \partial_x \mathbf{F}_a(\mathbf{U}, \mathbf{U}_a) = 0. \end{cases} \tag{3}$$

At this stage, it is easy but important to remark that the global system (3) admits the same real eigenvalues $\lambda_\pm$ (both with multiplicity 2) as the original system (1) but it is only weakly hyperbolic as soon as $\tau_a \neq 0$. We recall that weak hyperbolicity means that the Jacobian matrix of the system admits real eigenvalues but is not $\mathbb{R}$-diagonalizable. As a consequence and without any modification of (3), discontinuous weak solutions of the state variable $\mathbf{U}$ will generally induce Dirac delta functions in the sensitivity variable $\mathbf{U}_a$, in addition to the usual discontinuity, so that the solutions of (3) have to be understood in the sense of measures. However, and as already stated above, we are not interested in considering Dirac delta functions. Instead, we would like to introduce a modification in the sensitivity equations in order to make the system (3) valid in the usual sense of weak solutions also for discontinuous state variables (as done in [5]). This is the aim of the next section.

2 Source Term

In order to remove the Dirac delta functions that are naturally present in the solutions of (3), we suggest to add to (2) a source term $\mathbf{S}$, which is of the following form:

$$\mathbf{S} = \sum_{k=1}^{N_s} \delta_k \boldsymbol{\rho}_k, \tag{4}$$

where N_s is to be associated to the number of discontinuities in the state solution $\mathbf{U}$, $\boldsymbol{\rho}_k$ is the amplitude of the k-th correction (to be computed), and δ_k is the Dirac delta function $\delta_k = \delta(x - x_{s,k})$, where $x_{s,k}$ is the position of the k-th discontinuity. Let us consider then a control volume $(x_1, x_2) \times (t_1, t_2)$ containing a single discontinuity indexed by k and propagating at speed σ_k. By integrating the equations (2) with the additional source term (4) over the control volume, when the size of the control volume tends to zero one has:

$$\boldsymbol{\rho}_k(t) = (\mathbf{U}_a^- - \mathbf{U}_a^+)\sigma_k + \mathbf{F}_a^+ - \mathbf{F}_a^-, \tag{5}$$

where the plus (respectively minus) indicates the value of the sensitivity $\mathbf{U}_a$ and of the flux $\mathbf{F}_a$ to the right (respectively left) of the discontinuity. In other words, (5) gives a natural meaning of $\boldsymbol{\rho}_k$ in terms of a defect measure of the Rankine-Hugoniot relations applied to (2). It is now a matter of defining $\boldsymbol{\rho}_k$ in such a way that the new model including the source term is also valid for discontinuities of the state variable (recall that (2) was obtained by differentiating with respect to a the *smooth* solutions

of (1)). Considering the Rankine-Hugoniot conditions across a discontinuity of the state variable, namely $(\mathbf{U}^- - \mathbf{U}^+)\sigma_k = \mathbf{F}^- - \mathbf{F}^+$, we differentiate with respect to the parameter a to obtain $(\mathbf{U}_a^- - \mathbf{U}_a^+)\sigma_k + (\mathbf{U}^- - \mathbf{U}^+)\sigma_{k,a} = \mathbf{F}_a^- - \mathbf{F}_a^+$, with $\sigma_{k,a} = \partial_a \sigma_k$. Comparing the latter with (5), one is thus led to set

$$\boldsymbol{\rho}_k(t) = \sigma_{k,a}(\mathbf{U}^+ - \mathbf{U}^-). \tag{6}$$

3 Numerical Schemes

We introduce a constant space step Δx and a varying time step Δt^n. We define the mesh interfaces $x_{j+1/2} = j\Delta x$, the cells $C_j = [x_{j-1/2}, x_{j+1/2}]$, the cell centres x_j and the intermediate times $t^{n+1} = t^n + \Delta t^n$, where Δt^n is chosen according to the usual CFL condition.

The Godunov method.

In this paragraph, we present the usual Godunov method based on the exact resolution of the Riemann problem including the source term, and associated with the initial data $\mathbf{U}(x,0) = \mathbf{U}_L \mathbb{1}_{(x<x_c)} + \mathbf{U}_R \mathbb{1}_{(x>x_c)}$ and $\mathbf{U}_a(x,0) = \mathbf{U}_{a,L} \mathbb{1}_{(x<x_c)} + \mathbf{U}_{a,R} \mathbb{1}_{(x>x_c)}$. The details are not reported here but one has been able to prove that the analytical solution is known and that its structure is resumed in Figs. 1 and 2. In particular, the solution for the state consists of two waves, which can be either shocks or rarefaction waves, and whose speed can be computed exactly. On the other hand, the sensitivity has the same two-wave structure, however both of the waves are discontinuities. This simplification for the sensitivity is due to the fact that we are considering a reduced Euler system, under barotropic conditions (cf. [6]).

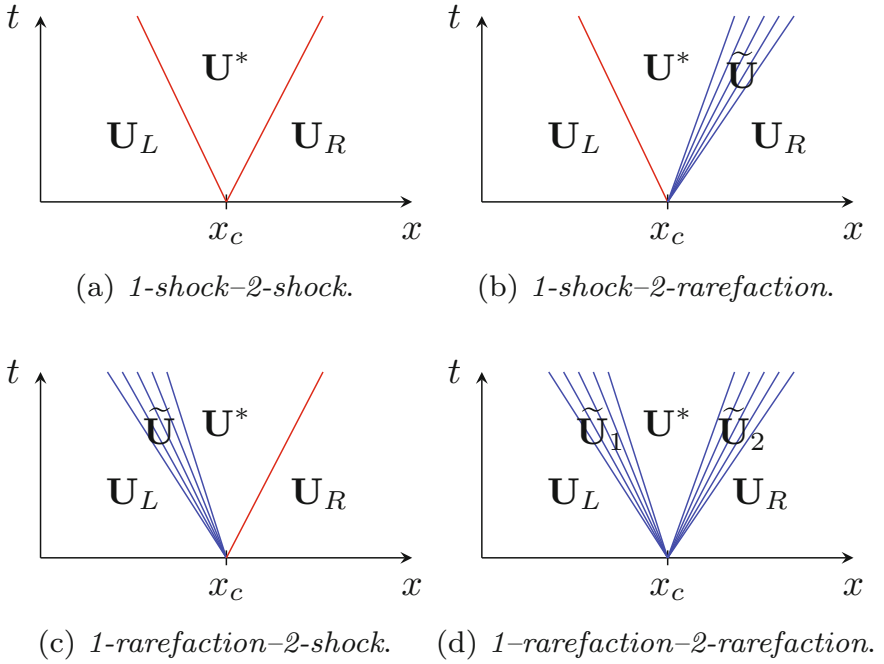
Since the state equations (1) are conservative, the Godunov method can be written with the classic update formula

$$\mathbf{U}_j^{n+1} = \mathbf{U}_j^n - \frac{\Delta t}{\Delta x}(\mathbf{F}(\mathbf{U}_{j+1/2}^*) - \mathbf{F}(\mathbf{U}_{j-1/2}^*)), \tag{7}$$

where $\mathbf{U}_{j-1/2}^*$ denotes the exact intermediate state variable in the Riemann solution associated to left and right states in the $j-1$-th and j-th cells.

Due to the presence of the source term in the sensitivity equations, a conservative update formula like (7) cannot be obtained. However, recall that the structure of the sensitivity at each interface is very simple and made of two shocks. As a consequence, one can easily perform the average on the cells, provided that the slopes κ_1 and κ_2, i.e. the slope of the red lines and the blue solid line in Fig. 2, are known at each interface $j - 1/2$. More precisely, we easily get

$$\mathbf{U}_{a,j}^{n+1} = \mathbf{U}_{a,j}^n + \frac{\Delta t}{\Delta x}\left(\kappa_{2,j-1/2}(\mathbf{U}_{a,j-1/2}^* - \mathbf{U}_{a,j}^n) - \kappa_{1,j+1/2}(\mathbf{U}_{a,j+1/2}^* - \mathbf{U}_{a,j}^n)\right), \tag{8}$$



(a) *1-shock–2-shock.* (b) *1-shock–2-rarefaction.*

(c) *1-rarefaction–2-shock.* (d) *1–rarefaction–2-rarefaction.*

Fig. 1 Configurations for the state variable $\mathbf{U}$

where the intermediate states $\mathbf{U}^*_{a,j-1/2}$ and $\mathbf{U}^*_{a,j+1/2}$ are known analytically and depend on $\mathbf{U}^*_{j-1/2}$ and $\mathbf{U}^*_{j+1/2}$ respectively. We remark that $\kappa_{\ell,j\pm1/2}$ depends on the solution structure: in case of shock it is given by the Rankine-Hugoniot conditions, in case of rarefaction wave it is the eigenvalue λ evaluated in the intermediate state $\mathbf{U}^*_{j\pm1/2}$. Let us observe that (8) already encompasses the source term to remove the Dirac delta function, therefore we do not need to discretise it. However, this correction is taken into account even in a numerical rarefaction profile (we recall that two points in a rarefaction wave are not necessarily linked by a rarefaction), and this is the cause of the failure of Godunov's method. In the next section, we present a Roe's method which uses shock detectors in order to overcome this problem.

A Roe-type method.

The proposed approximate Riemann solver of Roe-type consists of three constant states (say $\mathbf{U}_L$, $\mathbf{U}^*$ and $\mathbf{U}_R$ for the state, and $\mathbf{U}_{a,L}$, $\mathbf{U}^*_a$ and $\mathbf{U}_{a,R}$ for the sensitivity) separated by two shock waves propagating at velocities $\lambda^{ROE}_{L,j-1/2} = -\sqrt{-(p(\tau^n_{j-1}) - p(\tau^n_j))/(\tau^n_{j-1} - \tau^n_j)}$ and $\lambda^{ROE}_{R,j-1/2} = -\lambda^{ROE}_{L,j-1/2}$ if $\tau^n_{j-1} \neq \tau^n_j$ (and of course $\mp\sqrt{-p'(\tau^n_j)}$ otherwise). In the following, we will use the notation

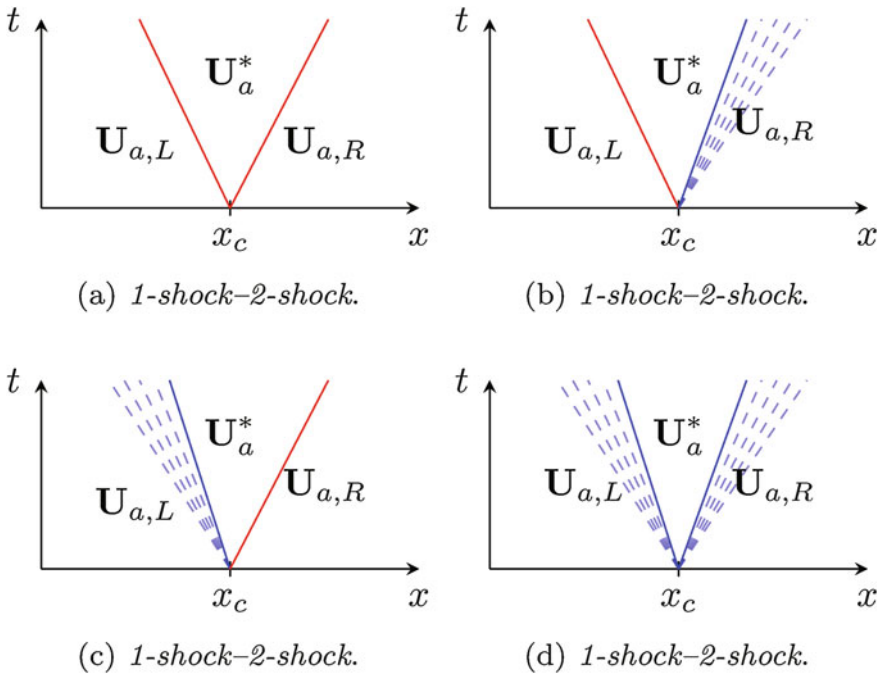


Fig. 2 Corresponding configurations for the sensitivity $\mathbf{U}_a$

$\lambda_{j-1/2}^{ROE} = \lambda_{R,j-1/2}^{ROE} = -\lambda_{L,j-1/2}^{ROE}$. The fact that the velocities at each interface are equal and opposite in sign allows us to write at the interface $j-1/2$ the Harten, Lax and van Leer consistency relations for the state in the following way:

$$\mathbf{U}_{j-1/2}^* = \frac{1}{2}(\mathbf{U}_{j-1}^n + \mathbf{U}_j^n) - \frac{\mathbf{F}(\mathbf{U}_j^n) - \mathbf{F}(\mathbf{U}_{j-1}^n)}{2\lambda_{j-1/2}^{ROE}}. \tag{9}$$

Knowing $\mathbf{U}_{j-1/2}^*$ and the velocity $\lambda_{j-1/2}^{ROE}$ at each interface one can average the solution value on the cells, obtaining the following update formula for the state:

$$\mathbf{U}_j^{n+1} = \mathbf{U}_j^n - \frac{\Delta t}{\Delta x}(\Phi_{j+1/2}^n - \Phi_{j-1/2}^n), \tag{10}$$

where Φ^n is the numerical flux and is defined as follows:

$$\Phi_{j-1/2}^n = \frac{\mathbf{F}(\mathbf{U}_j^n) + \mathbf{F}(\mathbf{U}_{j-1}^n)}{2} - \frac{\lambda_{j-1/2}^{ROE}}{2}(\mathbf{U}_j^n - \mathbf{U}_{j-1}^n).$$

We now consider the sensitivity equation, with the following source term, defined according to Eqs. (4)–(6) and given by

$$\Delta x \mathbf{S}_{j-1/2} = -\partial_a \lambda^{ROE}_{j-1/2}(\mathbf{U}^*_{j-1/2} - \mathbf{U}^n_{j-1})d_{1,j-1} + \partial_a \lambda^{ROE}_{j-1/2}(\mathbf{U}^n_j - \mathbf{U}^*_{j-1/2})d_{2,j},$$

where $d_{\ell,j}$ is a shock detector to be defined, and $d_{\ell,j} = 1$ if there is an $\ell-$shock in the $j-$th cell, it is zero elsewhere. The shock detector used in this work is based on the fact that the velocity u is always decreasing across a shock, and the covolume τ is decreasing across a $1-$shock, whilst it is increasing across a $2-$shock.

Finally, the update formula for the sensitivity is the following:

$$\mathbf{U}^{n+1}_{a,j} = \mathbf{U}^n_{a,j} - \frac{\Delta t}{\Delta x}(\Phi^n_{a,j+1/2} - \Phi^n_{a,j-1/2}) + \frac{\Delta t}{2}(\mathbf{S}_{j-1/2} + \mathbf{S}_{j+1/2}), \tag{11}$$

where Φ^n_a is defined as follows:

$$\Phi^n_{a,j-1/2} = \frac{\mathbf{F}_a(\mathbf{U}^n_j, \mathbf{U}^n_{a,j}) + \mathbf{F}_a(\mathbf{U}^n_{j-1}, \mathbf{U}^n_{a,j-1})}{2} - \frac{\lambda^{ROE}_{j-1/2}}{2}(\mathbf{U}^n_{a,j} - \mathbf{U}^n_{a,j-1}).$$

We remark that in (11), we add the source term $\mathbf{S}_{j-1/2}$ to both cells $j-$th and $(j-1)-$th: indeed, it is defined starting from an integral balance done on both cells.

Finally, we extended this scheme to the second order: we used a standard two-step Runge-Kutta method in time, whereas in space we used a MUSCL scheme adapted to take into account a second order discretization of the source term.

4 Numerical Results

In this section we present some numerical results. The spatial domain is $(0, 1)$, $x_c = 0.5$, and final time $T = 0.03$.

The test case considered is a 1-shock–2-rarefaction, with the following initial conditions: $\mathbf{U}_L = (0.7, 0)^T$, $\mathbf{U}_R = (0.2, 0)^T$, $\mathbf{U}_{a,L} = (0, 1)^T$, $\mathbf{U}_{a,R} = (0, 0)^T$, and the parameter of interest is $a = u_L$. Figure 3 shows the state τ and its sensitivity τ_a (u and u_a have a similar behaviour) at the final time T. As one can see, all the meth-

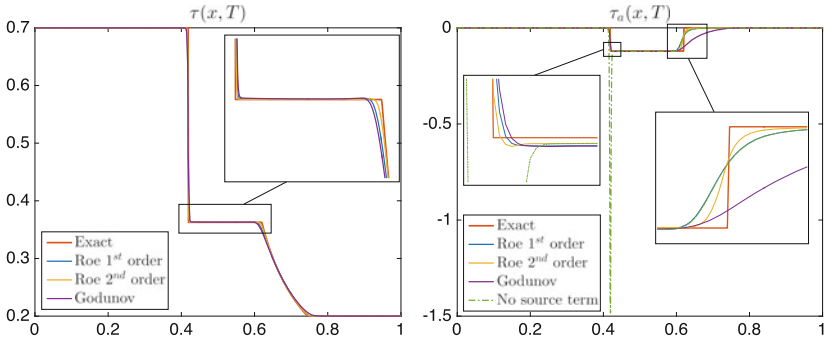


Fig. 3 Test case: 1-shock - 2-rarefaction, $\Delta x = 10^{-3}$

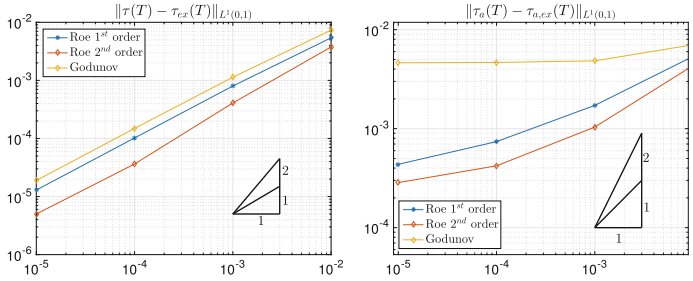


Fig. 4 Convergence results for test case 1-shock - 2-rarefaction

ods approximate well the state solution and the modified sensitivity formulation succeeds in removing the Dirac delta function located at point $x \approx 0.4$ and noticeable without source term (scheme labelled "No source term" in Fig. 3), however we remark that there are two main problems in the sensitivity solution: the shock corresponding to the state rarefaction is not captured and the value in the star zone is not correct. The first problem can be solved by refining the mesh or by using higher-order schemes (one can observe that the second order Roe method captures the discontinuity better); whilst the second problem, in our opinion, is due to the numerical diffusion in the shock. Figure 4 shows the convergence of the schemes: for the state we have the expected convergence; concerning the sensitivity, for coarser meshes the error is decreasing because its main part is in the rarefaction zone, however when this contribution becomes comparable with the error in the star zone, a plateau is reached.

5 Conclusion and Discussion

The numerical results show that the modified sensitivity system here proposed is well defined and it allows us to achieve the main goal of this work, i.e. to have a sensitivity without Dirac delta function. However, the proposed modified formulation

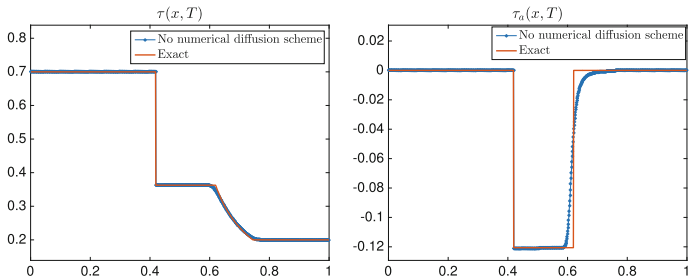


Fig. 5 Scheme with no numerical diffusion in the shock

yields an incorrect value of the sensitivity in the star zone. Interestingly, numerical results show that even the exact Godunov method does not provide a correct solution, neither does a higher order scheme. In our opinion, this problem is due to the numerical diffusion in the shock. To illustrate this, we briefly show in Fig. 5 the results obtained with a modified Godunov method based on sampling techniques, introduced in [3], which does not have any numerical diffusion and seems to provide a correct sensitivity in the star region.

References

1. Bardos, C., Pironneau, O.: A formalism for the differentiation of conservation laws. Compte rendu de l'Académie des Sciences **335**(10), 839–845 (2002)
2. Borggaard, J., Burns, J.: A PDE sensitivity equation method for optimal aerodynamic design. Journal of Computational Physics **136**(2), 366–384 (1997)
3. Chalons, C., Goatin, P.: Godunov scheme and sampling technique for computing phase transitions in traffic flow modeling. Interfaces and Free Boundaries **10**(2), 197–221 (2008)
4. Duvigneau, R., Pelletier, D.: A sensitivity equation method for fast evaluation of nearby flows and uncertainty analysis for shape parameters. Int. J. of CFD **20**(7), 497–512 (2006). August
5. Guinot, V.: Upwind finite volume solution of sensitivity equations for hyperbolic systems of conservation laws with discontinuous solutions. Computers & Fluids **38**(9), 1697–1709 (2009)
6. Guinot, V., Leménager, M., Cappelaere, B.: Sensitivity equations for hyperbolic conservation law-based flow models. Advances in water resources **30**(9), 1943–1961 (2007)

Semi-implicit Level Set Method with Inflow-Based Gradient in a Polyhedron Mesh

Jooyoung Hahn, Karol Mikula, Peter Frolkovič and Branislav Basara

Abstract In this paper, a semi-implicit method is proposed to solve a propagation in a normal direction with a cell-centered finite volume method. An inflow-based gradient is used to discretize the magnitude of the gradient and it brings the second order upwind difference in an evenly spaced one dimensional domain. In three dimensional domain, we numerically verify that the proposed scheme is second order. The implementation is straightforwardly combined with a conventional finite volume code and 1-ring face neighborhood for parallel computation. An experimental order of convergence and a comparison of wall clock time between semi-implicit and explicit method are illustrated by numerical examples.

Keywords Semi-implicit method · Level set method · Polyhedron mesh

MSC (2010): 65M08 · 65N08 · 35F25 · 35F30

The original version of the book was revised: Missed out corrections have been updated. The erratum to the book is available at https://doi/org/10.1007/978-3-319-57394-6_58

J. Hahn (✉) · B. Basara
AVL LIST GmbH, Hans-List-Platz 1, 8020 Graz, Austria
e-mail: jooyoung.hahn@avl.com

B. Basara
e-mail: branislav.basara@avl.com

K. Mikula · P. Frolkovič
Department of Mathematics and Descriptive Geometry, Slovak University of Technology, Radlinskeho 11, 810 05 Bratislava, Slovakia
e-mail: karol.mikula@stuba.sk

P. Frolkovič
e-mail: peter.frolkovic@stuba.sk

C. Cancés and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_9

1 Introduction

We solve a partial differential equation for a propagation in a normal direction [11]:

$$\partial_t \phi(\mathbf{x}, t) + F(\mathbf{x})|\nabla \phi(\mathbf{x}, t)| = G(\mathbf{x}), \quad (\mathbf{x}, t) \in \Omega \times [0, T], \tag{1}$$

where $\Omega \subset \mathbb{R}^3$ is a computational domain, T is the final time, the speed function F and the force term G are fixed, and the initial condition is given on Ω. Equation (1) has been extensively used to solve evolving interfacial problems in many applications such as image processing, computer vision, combustion, fluid dynamics, etc.; more details are given in [10, 14] and the references therein. In contrast to a standard structured mesh in image processing, real world three dimensional (3D) problems from physics or engineering are usually defined on a complicated geometry, for example, the combustion problems in 3D engines. Moreover, in industrial applications, a polyhedron mesh has been used extensively because of its shape flexibility [12]. In this paper, in order to extend topological advantages of the level set method into industrial problems with complicated geometry, we propose a numerical algorithm to solve the governing equation (1) on polyhedron meshes. We impose a linear extension at boundary, that is, a "ghost" value is linearly extrapolated from the boundary value. It can be properly discretized in a cell-centered finite volume method.

Inspired by the methods in [3, 5–8, 16], we propose a semi-implicit method to solve (1). It is very crucial to design a method to reduce a time step restriction in a polyhedron mesh. When the geometrical shape of a computational domain is complex, it is inevitable to have nonuniform size of cell volumes and it gives a severe restriction of time stepping in an explicit method because of the CFL condition for very small cells. The main difference between the proposed method and the methods in [2, 7–9] is an approximation of the gradient. Instead of using a cell-centered gradient to achieve the second order scheme, we approximate a gradient by an inverse distance average of face gradients only from inflow sides, named by the inflow-based gradient. In an evenly discretized 1D domain, the inflow-based gradient brings the correct second order upwind discretization of magnitude of the gradient in (1). Moreover, it allows us to use the simplest structure of decomposed domains for parallel computation which is the 1-ring face neighborhood structure.

In the rest of the paper, the inflow-based gradient is introduced in Sect. 2 and then a semi-implicit method is proposed. In Sect. 3, the experimental order of convergence (EOC) is investigated and the wall clock time of semi-implicit and explicit method is compared.

2 Semi-implicit Method with Inflow-Based Gradient

In order to explain the proposed method for a 3D mesh, some notations are introduced. The sets of indices to uniquely indicate cells, faces, and vertices are denoted by $\mathscr{C}$, $\mathscr{F}$, and $\mathscr{V}$, respectively. A whole computational domain $\bar{\Omega} \subset \mathbb{R}^3$ is discretized by open cells Ω_p such that $\bar{\Omega} = \bigcup_{p\in\mathscr{C}} \bar{\Omega}_p$ with the volume $|\Omega_p| \neq 0$. We define two sets of neighbor information of Ω_p, $p \in \mathscr{C}$; the first is the neighbor cells whose face is shared by Ω_p, $\mathscr{N}_p = \{q \in \mathscr{C} : \text{there exists a face } e_f \in \partial\Omega_q \cap \partial\Omega_p,\ f \in \mathscr{F}\}$ and the second is the attached faces to Ω_p and they are indicated by two sets: $\mathscr{F}_p = \{f \in \mathscr{F} : e_f \in \partial\Omega_p\}$ and $\mathscr{B}_p = \{f \in \mathscr{F}_p : e_f \in \partial\Omega_p \cap \partial\Omega\}$. Note that a nonplanar face of a polyhedron cell should be tessellated into triangles to make its sub-face as a plane. From a given face center of a nonplanar face, a triangle is defined by two consecutive vertices of the face and the face center. By an abuse of notation, $\mathscr{F}$ includes all tessellated faces.

2.1 Inflow-Based Gradient Finite Volume Method

With simple coefficient $G = 0$ in the governing equation (1) and using Gauss's theorem, we have a standard spatial discretization at $p \in \mathscr{C}$ in cell-centered finite volume method:

$$\int_{\Omega_p} \partial_t \phi + \sum_{f\in\mathscr{F}_p} \left(\phi_{pf} - \phi_p\right) a_{pf} = 0, \quad a_{pf} = \int_{e_f} F \frac{\nabla\phi}{|\nabla\phi|} \cdot \mathbf{n} \simeq F_f \frac{\nabla\phi_f}{|\nabla\phi_f|_\sigma} \cdot \mathbf{n}_{pf}, \tag{2}$$

where $\mathbf{n}$ is a unit outward normal vector at a face e_f, $f \in \mathscr{F}_p$, F_f is a value of F at the face center, $\mathbf{n}_{pf} = |e_f|\mathbf{n}$, and $|\nabla\phi_f|_\sigma = (\sigma^2 + |\nabla\phi_f|^2)^{\frac{1}{2}}$ with a small $\sigma > 0$. The spatial discretization is explained by two steps; the first is to define an inflow-based gradient and the second is to compute a face value ϕ_{pf}.

An inflow-based gradient computed by face gradients is defined at a cell center. A face gradient is obtained by a minimization from the values close to a face such as cell centers and vertices. A vertex value is linearly interpolated from cell-centered gradients. A cell-centered gradient with a linear extrapolation at boundary is obtained by a minimizer of a functional $f(\mathbf{y}) = \sum_{\mathbf{x}\in\mathscr{S}_p} w_p(\mathbf{x}) \left|\mathbf{y} \cdot (\mathbf{x} - \mathbf{x}_p) - (\phi(\mathbf{x}) - \phi_p)\right|^2$, where a weight function is $w_p(\mathbf{x}) = |\mathbf{x} - \mathbf{x}_p|^{-2}$ and a set of points $\mathscr{S}_p$ at the cell $p \in \mathscr{C}$ is either $\{\mathbf{x}_q \mid q \in \mathscr{N}_p\}$ if $\mathscr{B}_p = \emptyset$ or $\{\mathbf{x}_q \mid q \in \mathscr{N}_p\} \cup \{\mathbf{x}_b \mid b \in \mathscr{B}_p\}$ if $\mathscr{B}_p \neq \emptyset$. From a cell-centered gradient, a vertex value can be approximated by an inverse distance average and linear approximation from adjacent cells.

Before we define an inflow-based gradient, a face gradient should be computed in order to obtain a flux in (2) at a face e_f, $f \in \mathscr{F}$. Let us denote $\mathscr{P}_f$ as a set of points around a face center: either $\{\mathbf{x}_p, \mathbf{x}_q\} \cup \mathbf{V}_f$ if $\exists!\ p, q \in \mathscr{C}$ such that $e_f \in \partial\Omega_p \cap \partial\Omega_q$

or $\{\mathbf{x}_p\} \cup \mathbf{V}_f$ if $\exists!\ p \in \mathscr{C}$ such that $f \in \mathscr{B}_p \neq \emptyset$, where $\mathbf{V}_f$ are vertices of a face e_f. Note that $\mathscr{P}_f$ is a generalization of diamond-cell strategy in a regular structured cube mesh [1]. A face value α_f and gradient $\boldsymbol{\beta}_f$ are obtained by minimizer of a functional $g(a_f, \mathbf{b}_f) = \sum_{\mathbf{x} \in \mathscr{P}_f} w_f(\mathbf{x}) \left| a_f + \mathbf{b}_f \cdot (\mathbf{x} - \mathbf{x}_f) - \phi(\mathbf{x}) \right|^2$, where a weight function $w_f(\mathbf{x}) = |\mathbf{x} - \mathbf{x}_f|^{-2}$ at the face center $\mathbf{x}_f$. Note that a face value α_f on a boundary face is a linearly extended value. Finally, we define an inflow-based gradient as an inverse distance average of face gradients only from inflow faces with an inverse distant $d_{pf} = |\mathbf{x}_f - \mathbf{x}_p|^{-1}$ and its sum $W_d = \sum_{f \in \mathscr{A}_p^-} d_{pf}$:

$$D_p^- \phi = W_d^{-1} \sum_{f \in \mathscr{A}_p^-} d_{pf} \boldsymbol{\beta}_f, \tag{3}$$

where $\mathscr{A}_p^- = \mathscr{B}_p^- \cup \mathscr{F}_p^-$, $\mathscr{B}_p^- = \{b \in \mathscr{B}_p \mid a_{pb} < 0\}$, and $\mathscr{F}_p^- = \{f \in \mathscr{F}_p \setminus \mathscr{B}_p \mid a_{pf} < 0\}$. If $\mathscr{A}_p^- = \emptyset$, then we define $D_p^- \phi = 0$.

Now, we compute a face value ϕ_{pf} in (2) from the inflow-based gradient. When a face value is computed at an internal face, a face value ϕ_{pf} in (2) is computed straightforwardly:

$$\begin{aligned} f \in \mathscr{F}_p \setminus \mathscr{B}_p,\ p \in \mathscr{C} \ &\Rightarrow\ \exists!\, q \in \mathscr{N}_p \text{ such that } e_f \in \partial\Omega_p \cap \partial\Omega_q \\ &\Rightarrow\ \phi_{pf} = \begin{cases} \phi_p + D_p^- \phi \cdot (\mathbf{x}_f - \mathbf{x}_p) & \text{if } a_{pf} \geq 0, \\ \phi_q + D_q^- \phi \cdot (\mathbf{x}_f - \mathbf{x}_q) & \text{if } a_{pf} < 0. \end{cases} \end{aligned} \tag{4}$$

When a face value is computed at a boundary face, we use the linear extrapolation and then a face value ϕ_{pb} in (2) is formulated by

$$b \in \mathscr{B}_p (\neq \emptyset),\ \ p \in \mathscr{C} \ \Rightarrow\ \phi_{pb} = \begin{cases} \phi_p + D_p^- \phi \cdot (\mathbf{x}_b - \mathbf{x}_p) & \text{if } a_{pb} \geq 0, \\ \alpha_b & \text{if } a_{pb} < 0. \end{cases} \tag{5}$$

Note that the boundary face value α_b, $b \in \mathscr{B}_p$ is obtained by imposing a linear extrapolation. From (4) and (5), we finally have the spatial discretization:

$$\begin{aligned} \int_{\Omega_p} \partial_t \phi = &- \sum_{f \in \mathscr{F}_p^-} \left(\phi_q + D_q^- \phi \cdot \mathbf{d}_{qf} - \phi_p \right) a_{pf} - \sum_{f \in \mathscr{F}_p^+} \left(D_p^- \phi \cdot \mathbf{d}_{pf} \right) a_{pf} \\ &- \sum_{b \in \mathscr{B}_p^-} \left(\alpha_b - \phi_p \right) a_{pb} - \sum_{b \in \mathscr{B}_p^+} \left(D_p^- \phi \cdot \mathbf{d}_{pb} \right) a_{pb}, \end{aligned} \tag{6}$$

where $\mathscr{B}_p^+ = \mathscr{B}_p \setminus \mathscr{B}_p^-$, $\mathscr{F}_p^+ = (\mathscr{F}_p \setminus \mathscr{B}_p) \setminus \mathscr{F}_p^-$, $\mathbf{d}_{qf} = \mathbf{x}_f - \mathbf{x}_q$, and for each $f \in \mathscr{F}_p \setminus \mathscr{B}_p$, $p \in \mathscr{C}$ there exists an index $q \in \mathscr{N}_p$ such that $e_f \subset \partial\Omega_p \cap \partial\Omega_q$. From a tedious derivation of (6) in an evenly spaced 1D domain, the inflow-based gradient in the above formula brings the second order upwind difference of the magnitude

of the gradient. Note that the first order upwind difference is used in well-known standard schemes [11, 13].

2.2 Semi-implicit Method

Let us denote an evenly divided time step $\Delta t = T/N$ for a fixed $N \in \mathbb{N}$ and $\phi_p^n = \phi(\mathbf{x}_p, n\Delta t)$. Inspired by [3, 7–9, 16], the outflow information is used explicitly and we propose to use the inflow information partly implicitly and partly iteratively because of a limitation of sharing variables in the 1-ring face neighborhood structure of decomposed domains:

$$\begin{aligned} &\frac{|\Omega_p|}{\Delta t}\left(\phi_p^{n,k} - \phi_p^{n-1}\right) + \sum_{f\in\mathscr{F}_p^-}\left(\phi_q^{n,k} + D_q^-\phi^{n,k-1}\cdot\mathbf{d}_{qf} - \phi_p^{n,k}\right)a_{pf}^{n-1} \\ &+ \sum_{b\in\mathscr{B}_p^-}\left(\alpha_b^{n,k-1} - \phi_p^{n,k}\right)a_{pb}^{n-1} + \sum_{f\in\mathscr{A}_p^+}\left(D_p^-\phi^{n-1}\cdot\mathbf{d}_{pf}\right)a_{pf}^{n-1} = 0, \end{aligned} \tag{7}$$

where $k = 1, \ldots, K$ and $\mathscr{A}_p^+ = \mathscr{B}_p^+ \cup \mathscr{F}_p^+$. The above system of equations can be written by

$$\left(\frac{|\Omega_p|}{\Delta t} - \sum_{f\in\mathscr{A}_p^-} a_{pf}^{n-1}\right)\phi_p^{n,k} + \sum_{f\in\mathscr{F}_p^-} a_{pf}^{n-1}\phi_q^{n,k} = R(\phi_p^{n-1}, \phi_p^{n,k-1}), \tag{8}$$

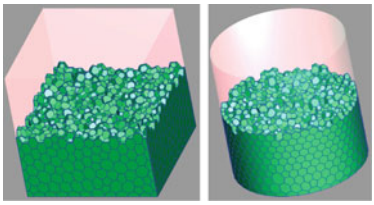
where the right-hand side R is a collection of explicit information:

$$\begin{aligned} R(\phi_p^{n-1}, \phi_p^{n,k-1}) \equiv{}& \frac{|\Omega_p|}{\Delta t}\phi_p^{n-1} - \sum_{b\in\mathscr{B}_p^-}\alpha_b^{n,k-1}a_{pb}^{n-1} \\ &- \sum_{f\in\mathscr{F}_p^-} D_q^-\phi^{n,k-1}\cdot\mathbf{d}_{qf}a_{pf}^{n-1} - \sum_{f\in\mathscr{A}_p^+} D_p^-\phi^{n-1}\cdot\mathbf{d}_{pf}a_{pf}^{n-1}. \end{aligned}$$

For all examples in Sect. 3, we fix $K = 1$ and update $\phi^n = \phi^{n,1}$ using $\phi^{n,0} = \phi^{n-1}$ in the above formulas.

3 Numerical Experiments

Two examples are presented to check an EOC of the proposed method. An algebraic multigrid method (AMG) in AVL FIRE® on decomposed computational domains with 1-ring face neighborhood structure is used to solve (8) for all examples. In



level	P_b	P_c
1	4,033	3,947
2	30,683	28,410
3	241,726	224,548
4	1,914,579	1,788,209

Fig. 1 The first and second from the *left figure* are polyhedron cells in a box (P_b) and a cylinder (P_c) shape generated by AVL FIRE® and the *right table* is the number of cells at each level. If one level gets higher, the average volume of cells is approximately 8 times smaller

Fig. 1, a box shape $\bar{\Omega} = [-0.05, 0.05]^3 \subset \mathbb{R}^3$ and a cylinder shape whose height 0.1 and radius is 0.05 are chosen to be a computational domain and polyhedron cells are generated in four levels to check EOC. A time step Δt in (7) for each level from 1 to 4 is fixed to be $3.0 \cdot 10^{-3}$, $1.5 \cdot 10^{-3}$, $7.5 \cdot 10^{-4}$, and $3.75 \cdot 10^{-4}$, respectively. A regularization parameter $\sigma = 10^{-12}$ in (2) is fixed for all examples.

The first example is a bidirectional flow from an analytically represented shape:

$$\partial_t \phi(\mathbf{x}, t) \pm |\nabla \phi(\mathbf{x}, t)| = \pm 1, \quad (\mathbf{x}, t) \in \Pi^{\pm} \times [0, T], \tag{9}$$

where a closed surface Π is given such that $\Pi = \partial\Pi^+ \cap \partial\Pi^-$, $\bar{\Pi}^+ \cup \bar{\Pi}^- = \bar{\Omega}$, $\Pi^+ \cap \Pi^- = \emptyset$ and an initial value $\phi(\mathbf{x}, 0)$ is positive on Π^+, negative on Π^-, and zero at $\mathbf{x} \in \Pi$. The bidirectional flow computes a signed distance function from Π using linear extrapolation at boundary. In Table 1, Π is chosen as a sphere whose radius is 0.02 and a cube whose edge is 0.04 and $T = 0.3$ is large enough to reach a steady state of (9) in a given box or cylinder shape domains in Fig. 1. From a sphere shape, the EOC from L^1-norm is second order but it is the first order from L^∞-norm. It is because L^∞-norm is sensitive on a singularity placed at the center of sphere. If the singularity is avoided in $L^\infty_\varepsilon = L^\infty(\Omega_\varepsilon)$ where $\Omega_\varepsilon = \{\mathbf{x} \in \Omega | |\mathbf{x}| > \varepsilon\}$ and $\varepsilon = 0.01$, the the EOC from L^∞_ε is around 2. From a cube shape, the EOC from L^1 and L^∞-norms is the first order which is caused by a lot of discontinuities of gradient in a solution.

In Table 2, we compare the wall clock time between semi-implicit and explicit method in the first example. The time step in an explicit method is computed by the same CFL condition in [2] and it is roughly three times smaller than the time steps used for the proposed semi-implicit method. The wall clock time of the proposed method only takes 18.75% of an explicit method in the average of $T_i / T_e * 100$ and it is caused by choosing a relatively large time step compared to an explicit method. Note that for the explicit method a second order total variation diminishing (TVD) Runge-Kutta method [4, 15] is used.

The second example is a propagation of surface which makes a given surface to shrink or expand along its normal direction:

$$\partial_t \phi(\mathbf{x}, t) \pm |\nabla \phi(\mathbf{x}, t)| = 0, \quad (\mathbf{x}, t) \in \Omega \times [0, T], \tag{10}$$

Table 1 The EOC of bidirectional flow (9); more details in Sect. 3

	Sphere				Cube			
	P_b		P_c		P_b		P_c	
Level	L^1	EOC	L^1	EOC	L^1	EOC	L^1	EOC
1	$1.90 \cdot 10^{-4}$	–	$1.70 \cdot 10^{-4}$	–	$1.09 \cdot 10^{-3}$	–	$8.30 \cdot 10^{-4}$	–
2	$5.24 \cdot 10^{-5}$	1.86	$4.29 \cdot 10^{-5}$	1.98	$5.29 \cdot 10^{-4}$	1.05	$4.38 \cdot 10^{-4}$	0.92
3	$1.30 \cdot 10^{-5}$	2.00	$1.08 \cdot 10^{-5}$	1.99	$2.54 \cdot 10^{-4}$	1.06	$1.94 \cdot 10^{-4}$	1.17
4	$3.10 \cdot 10^{-6}$	2.07	$2.60 \cdot 10^{-6}$	2.06	$1.28 \cdot 10^{-4}$	0.99	$9.55 \cdot 10^{-5}$	1.02
Level	L^∞	EOC	L^∞	EOC	L^∞	EOC	L^∞	EOC
1	$8.67 \cdot 10^{-4}$	–	$8.09 \cdot 10^{-4}$	–	$3.87 \cdot 10^{-3}$	–	$4.17 \cdot 10^{-3}$	–
2	$4.40 \cdot 10^{-4}$	0.98	$3.92 \cdot 10^{-4}$	1.04	$2.14 \cdot 10^{-3}$	0.85	$2.13 \cdot 10^{-3}$	0.97
3	$2.71 \cdot 10^{-4}$	0.70	$2.36 \cdot 10^{-4}$	0.73	$8.88 \cdot 10^{-4}$	1.27	$8.51 \cdot 10^{-4}$	1.32
4	$1.17 \cdot 10^{-4}$	1.20	$1.11 \cdot 10^{-4}$	1.08	$4.45 \cdot 10^{-4}$	1.00	$4.52 \cdot 10^{-4}$	0.91
Level	L_ε^∞	EOC	L_ε^∞	EOC				
1	$4.90 \cdot 10^{-4}$	–	$4.24 \cdot 10^{-4}$	–				
2	$1.84 \cdot 10^{-4}$	1.41	$1.93 \cdot 10^{-4}$	1.14	N/A			
3	$4.88 \cdot 10^{-5}$	1.92	$4.41 \cdot 10^{-5}$	2.13				
4	$1.43 \cdot 10^{-5}$	1.77	$1.22 \cdot 10^{-5}$	1.85				

Table 2 A comparison of wall clock time between semi-implicit (T_i) and explicit (T_e) method of solving (9) until $T = 0.003$; From the level 1 to 4, the numbers of CPUs are 2, 8, 32, and 128, respectively. The wall clock time is the average of 5 repeated computations

		P_b				P_c			
Level		1	2	3	4	1	2	3	4
Sphere	T_i	1.09	4.59	21.39	79.10	1.04	4.38	21.31	86.59
	T_e	5.79	24.70	113.49	491.35	5.49	23.48	112.98	406.89
Cube	T_i	1.08	4.55	21.18	96.16	1.03	4.27	21.48	86.90
	T_e	5.78	24.65	113.36	491.35	5.49	23.37	13.42	406.29

where an initial level set function is a signed distance function of spherical and octahedron shapes. In case of shrinking shapes, we use the initial shapes as two spheres whose centers are $(\pm 0.025, 0, 0)$ and radius is 0.02 or two octahedrons whose centers are same as the spheres and an edge is $0.02\sqrt{2}$ and the final time $T = 0.006$. In case of expanding shapes, we use the initial shapes as two spheres whose centers are $(\pm 0.025, 0, 0)$ and radius is 0.024 or two octahedrons whose centers are same as the spheres and an edge is $0.024\sqrt{2}$ and the final time $T = 0.006$. Note that the expanding two separated shapes merge as one shape at the final time. In this example, since the meaningful numerical results are only on the zero level set, we measure a local error from $L^1_{loc} \equiv L^1(\Gamma)$, where Γ is the zero level set of exact solution. In Table 3, the EOC from L^1_{loc}-norm is presented. The EOC of shrinking octahedrons is supposed to be the first order because of discontinuities of gradient

Table 3 The EOC of a propagation in a normal direction (10); more details in Sect. 3

	Shrinking spheres				Shrinking octahedrons			
	P_b		P_c		P_b		P_c	
Level	L^1_{loc}	EOC	L^1_{loc}	EOC	L^1_{loc}	EOC	L^1_{loc}	EOC
1	$2.34 \cdot 10^{-4}$	–	$2.65 \cdot 10^{-4}$	–	$7.47 \cdot 10^{-4}$	–	$6.05 \cdot 10^{-4}$	–
2	$6.37 \cdot 10^{-5}$	1.88	$6.86 \cdot 10^{-5}$	1.95	$4.09 \cdot 10^{-4}$	0.87	$3.77 \cdot 10^{-4}$	0.68
3	$1.43 \cdot 10^{-5}$	2.15	$1.37 \cdot 10^{-5}$	2.33	$1.51 \cdot 10^{-4}$	1.44	$1.35 \cdot 10^{-4}$	1.48
4	$3.05 \cdot 10^{-6}$	2.23	$2.78 \cdot 10^{-6}$	2.30	$4.41 \cdot 10^{-5}$	1.77	$3.90 \cdot 10^{-5}$	1.80
	Expanding spheres				Expanding octahedrons			
	P_b		P_c		P_b		P_c	
Level	L^1_{loc}	EOC	L^1_{loc}	EOC	L^1_{loc}	EOC	L^1_{loc}	EOC
1	$1.60 \cdot 10^{-4}$	–	$1.36 \cdot 10^{-4}$	–	$7.08 \cdot 10^{-4}$	–	$6.15 \cdot 10^{-4}$	–
2	$4.02 \cdot 10^{-5}$	1.99	$3.80 \cdot 10^{-5}$	1.84	$3.35 \cdot 10^{-4}$	1.08	$3.31 \cdot 10^{-4}$	0.89
3	$1.05 \cdot 10^{-5}$	1.94	$9.34 \cdot 10^{-6}$	2.03	$1.80 \cdot 10^{-4}$	0.89	$1.71 \cdot 10^{-4}$	0.95
4	$2.50 \cdot 10^{-6}$	2.07	$2.35 \cdot 10^{-6}$	1.99	$9.64 \cdot 10^{-5}$	0.90	$9.31 \cdot 10^{-5}$	0.88

on the zero level set but it seems to be higher than 1. The EOC of shrinking spheres is higher than expanding spheres because the solution of shrinking spheres do not have any singularities on the zero level set. As it is expected, the EOC of expanding octahedrons is close to the first order and it is because of discontinuities of gradient and linearly extrapolated boundary values.

4 Conclusion

We proposed a new semi-implicit level set method for motion in normal direction which is second order accurate on three-dimensional polyhedron meshes.

Acknowledgements The work was supported by grants VEGA 1/0808/15, VEGA 1/0728/15, APVV-0522-15.

References

1. Coudière, Y., Vila, J.P., Villedieu, P.: Convergence rate of a finite volume scheme for a two dimensional convection-diffusion problem. ESAIM Math. Model. Numer. Anal. **33**, 493–516 (1999)
2. Frolkovič, P., Mikula, K.: High-resolution flux-based level set method. SIAM J. Sci. Comput. **29**, 579–597 (2007)
3. Frolkovič, P., Mikula, K., Urbán, J.: Semi-implicit finite volume level set method for advective motion of interfaces in normal direction. Appl. Numer. Math. **95**, 214–228 (2015)

4. Gottlieb, S., Shu, C.W.: Total Variation Diminishing Runge-Kutta schemes. Math. Comput. **67**, 73–85 (1998)
5. Luo, J., Luo, Z., Chen, L., Tong, L., Wang, M.Y.: A semi-implicit level set method for structural shape and topology optimization. J. Comput. Phys. **227**, 5561–5581 (2008)
6. May, S., Berger, M.: An explicit implicit scheme for cut cells in embedded boundary meshes. J. Sci. Comput. 1–25 (2016). doi:https://doi.org/10.1007/s10915-016-0326-2
7. Mikula, K., Ohlberger, M.: A new level set method for motion in normal direction based on a semi-implicit forward-backward diffusion approach. SIAM J. Sci. Comput. **32**, 1527–1544 (2010)
8. Mikula, K., Ohlberger, M.: Inflow-implicit/outflow-explicit scheme for solving advection equations. In: Fořt, J., Fürst, J., Halama, J., Herbin, R., Hubert, F. (eds.) Finite Volumes for Complex Applications VI–Problems and Perspectives. Springer Proceedings in Mathematics 4, vol. 1, pp. 683–691. Springer, Berlin (2011)
9. Mikula, K., Ohlberger, M., Urbán, J.: Inflow-implicit/outflow-explicit finite volume methods for solving advection equations. Appl. Numer. Math. **85**, 16–37 (2014)
10. Osher, S., Fedkiw, R.: Level Set Methods and Dynamic Implicit Surfaces. Springer, Berlin (2000)
11. Osher, S., Sethian, J.A.: Fronts propagating with curvature dependent speed: algorithms based on Hamilton-Jacobi formultaions. J. Comput. Phys. **79**, 12–49 (1988)
12. Perić, M.: Flow simulation using control volumes of arbitrary polyhedral shape. ERCOFTAC Bull. **62**, 25–29 (2004)
13. Rouy, E., Tourin, A.: A viscosity solutions approach to shape-from-shading. SIAM J. Numer. Anal. **29**, 867–884 (1992)
14. Sethian, J.A.: Level Set Methods and Fast Marching Methods, Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materical Science. Cambridge University Press, New York (1999)
15. Shu, C.W., Osher, S.: Efficient implementation of essentially non-oscillatory shock-capturing schemes. J. Comput. Phys. **77**, 439–471 (1988)
16. Smereka, P.: Semi-implicit level set methods for curvature and surface diffusion motion. J. Sci. Comput. **19**, 439–456 (2003)

A Staggered Scheme for the Euler Equations

Thierry Goudon, Julie Llobell and Sebastian Minjeaud

Abstract We extend to the full Euler system the scheme introduced in [Berthelin, Goudon, Minjeaud, Math. Comp. 2014] for solving the barotropic Euler equations. This finite volume scheme is defined on staggered grids with numerical fluxes derived in the spirit of kinetic schemes. The difficulty consists in finding a suitable treatment of the energy equation while density and internal energy on the one hand, and velocity on the other hand, are naturally defined on dual locations. The proposed scheme uses the density, the velocity and the internal energy as computational variables and stability conditions are identified in order to preserve the positivity of the discrete density and internal energy. Moreover, we define averaged energies which satisfy *local* conservation equations. Finally, we provide numerical simulations of Riemann problems to illustrate the behaviour of the scheme.

Keywords Finite volumes · Conservation laws · Staggered grids · Euler equations

MSC (2010): 65M08 · 76M12 · 35L65 · 35Q31

T. Goudon · J. Llobell (✉) · S. Minjeaud
Université Côte d'Azur, CNRS, Inria, France
e-mail: thierry.goudon@inria.fr

J. Llobell
e-mail: llobell@unice.fr

S. Minjeaud
e-mail: sebastian.minjeaud@unice.fr

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_10

1 Introduction

This work aims at designing a scheme to numerically solve the 1D-Euler system:

$$\begin{aligned} &\partial_t \begin{pmatrix} \rho \\ \rho u \\ \rho E \end{pmatrix} + \partial_x \begin{pmatrix} \rho u \\ \rho u^2 + p \\ \rho E u + p u \end{pmatrix} = 0, \quad (t,x) \in [0,\infty) \times \mathbb{R}, \\ &E = u^2/2 + e, \quad p = (\gamma - 1)\rho e, \end{aligned} \tag{1}$$

where ρ, u, E and p stand for the density, the velocity, the total energy and the pressure respectively, $u^2/2$ and e are the kinetic and internal energies, and $\gamma > 1$ is the adiabatic exponent.

We wish to extend to (1) the scheme designed in [1] for the barotropic Euler equations. This scheme works on staggered grids—meaning that densities and velocities are not collocated—and this raises a difficulty for (1) as the definition of the total energy mixes quantities, namely the velocity and the internal energy, naturally defined on different grids. To address this issue, it is convenient to work with the internal energy equation, namely

$$\partial_t(\rho e) + \partial_x(\rho e u) = -p\partial_x u, \tag{2}$$

instead of the evolution equation for ρE, since discrete densities, pressures, and internal energies are naturally stored at the same locations. This formulation has also the advantage of making more direct the analysis of the positivity of e. Unfortunately, as it is well-known, this non conservative formulation is not equivalent to (1) when the solution presents discontinuities. We shall follow the approach discussed in [2] by introducing in (2) correction terms accounting for the kinetic energy balance. Then, the scheme introduced in [2] can be shown: a) to be consistent with (a weak form of) the total energy equation as the space step δx goes to zero and b) to conserve the *global* discrete total energy. Our purpose is two-fold. First of all, we shall adapt the scheme of [1] for dealing with (1). Second of all, we introduce averaged energies which satisfy *local* conservation equations. Finally we provide some numerical simulations in Sect. 5.

2 Staggered Scheme

Let $(x_j)_j$ be a subdivision of the 1D computational domain and denote the size of the cells by $\delta x_{j+\frac{1}{2}} = x_{j+1} - x_j$. The cell centers, $x_{j+\frac{1}{2}} = (x_j + x_{j+1})/2$, define the dual mesh and we set $\delta x_j = (\delta x_{j-\frac{1}{2}} + \delta x_{j+\frac{1}{2}})/2$. The discrete densities $\rho_{j+\frac{1}{2}}$ and internal energies $e_{j+\frac{1}{2}}$ are stored at the centers $x_{j+\frac{1}{2}}$ whereas the velocities u_j are located at the edges x_j. The time dicretization is explicit and we use the convention

that, with q the evaluation of a certain quantity at time t, $\overline{q}$ stands for its update at time $t + \delta t$.

- The density $\rho_{j+\frac{1}{2}}$ is updated using the discrete mass balance equation:

$$\frac{\overline{\rho}_{j+\frac{1}{2}} - \rho_{j+\frac{1}{2}}}{\delta t} + \frac{\mathscr{F}_{j+1} - \mathscr{F}_j}{\delta x_{j+\frac{1}{2}}} = 0.$$

The mass fluxes are defined by $\mathscr{F}_j = \mathscr{F}_j^+ + \mathscr{F}_j^-$ where $\mathscr{F}_j^+ = \mathscr{F}^+(\rho_{j-\frac{1}{2}}, e_j, u_j)$, $\mathscr{F}_j^- = \mathscr{F}^-(\rho_{j+\frac{1}{2}}, e_j, u_j)$ and $e_j = (e_{j-\frac{1}{2}} + e_{j+\frac{1}{2}})/2$. Denoting $c(e) = \sqrt{(\gamma-1)\gamma e}$ the sound speed, the definition of the numerical fluxes $\mathscr{F}^\pm$ is extracted from [1]

$$\mathscr{F}^+(\rho, e, u) = \begin{cases} 0 & \text{if } u \leqslant -c(e), \\ \dfrac{\rho(u + c(e))^2}{4c(e)} & \text{if } |u| < c(e), \\ \rho u & \text{if } u > c(e), \end{cases} \quad \text{and } \mathscr{F}^-(\rho, e, u) = -\mathscr{F}^+(\rho, e, -u).$$

In the sequel we use the following two properties: $\forall u \in \mathbb{R},\ \forall \rho,\ e \geqslant 0$,

$$0 \leqslant \mathscr{F}^+(\rho, e, u) \leqslant \rho[\lambda_+(e, u)]^+ \quad \text{and} \quad -\rho[\lambda_-(e, u)]^- \leqslant \mathscr{F}^-(\rho, e, u) \leqslant 0, \tag{3}$$

where $\lambda_\pm(e, u) = u \pm c(e)$ and $[z]^\pm = \frac{1}{2}(|z| \pm z)$.

- The velocity u_j is then updated using the discrete momentum balance equation:

$$\frac{\overline{\rho}_j \overline{u}_j - \rho_j u_j}{\delta t} + \frac{\mathscr{G}_{j+\frac{1}{2}} - \mathscr{G}_{j-\frac{1}{2}}}{\delta x_j} + \frac{\Pi_{j+\frac{1}{2}} - \Pi_{j-\frac{1}{2}}}{\delta x_j} = 0. \tag{4}$$

The momentum flux $\mathscr{G}_{j+\frac{1}{2}}$ and the pressure $\Pi_{j+\frac{1}{2}}$ are defined by:

$$\mathscr{G}_{j+\frac{1}{2}} = u_j \mathscr{F}^+_{j+\frac{1}{2}} + u_{j+1} \mathscr{F}^-_{j+\frac{1}{2}} \quad \text{and} \quad \Pi_{j+\frac{1}{2}} = (\gamma - 1)\rho_{j+\frac{1}{2}} e_{j+\frac{1}{2}}.$$

The quantities ρ_j and $\mathscr{F}^\pm_{j+\frac{1}{2}}$ are expressed as mean values of $\rho_{j\pm\frac{1}{2}}$ and $\mathscr{F}^\pm_j, \mathscr{F}^\pm_{j+1}$:

$$\rho_j = \frac{\delta x_{j+\frac{1}{2}} \rho_{j+\frac{1}{2}} + \delta x_{j-\frac{1}{2}} \rho_{j-\frac{1}{2}}}{2\delta x_j} \quad \text{and} \quad \mathscr{F}^\pm_{j+\frac{1}{2}} = \frac{\mathscr{F}^\pm_{j+1} + \mathscr{F}^\pm_j}{2}. \tag{5}$$

- The internal energy $e_{j+\frac{1}{2}}$ is updated using the following discrete equation:

$$\frac{\overline{\rho}_{j+\frac{1}{2}} \overline{e}_{j+\frac{1}{2}} - \rho_{j+\frac{1}{2}} e_{j+\frac{1}{2}}}{\delta t} + \frac{\mathscr{E}_{j+1} - \mathscr{E}_j}{\delta x_{j+\frac{1}{2}}} + \Pi_{j+\frac{1}{2}} \frac{\overline{u}_{j+1} - \overline{u}_j}{\delta x_{j+\frac{1}{2}}} = S_{j+\frac{1}{2}}. \tag{6}$$

The internal energy flux $\mathscr{E}_j$ is given by $\mathscr{E}_j = e_{j-\frac{1}{2}}\mathscr{F}_j^+ + e_{j+\frac{1}{2}}\mathscr{F}_j^-$. According to [2], the rhs $S_{j+\frac{1}{2}}$ is designed to account for the rest term that appears in the discrete kinetic energy balance and that do not vanish when δx goes to zero.

- To be more specific, the kinetic energy balance is obtained by multiplying (4) by $\overline{u}_j$. We find, see [1, 2]:

$$\frac{\overline{\rho}_j \dfrac{\overline{u}_j^2}{2} - \rho_j \dfrac{u_j^2}{2}}{\delta t} + \frac{\mathscr{H}_{j+\frac{1}{2}} - \mathscr{H}_{j-\frac{1}{2}}}{\delta x_j} + \frac{\Pi_{j+\frac{1}{2}} - \Pi_{j-\frac{1}{2}}}{\delta x_j}\overline{u}_j = -R_j,$$

where the kinetic energy flux is given by $\mathscr{H}_{j+\frac{1}{2}} = \dfrac{u_j^2}{2}\mathscr{F}^+_{j+\frac{1}{2}} + \dfrac{u_{j+1}^2}{2}\mathscr{F}^-_{j+\frac{1}{2}}$ and

$$R_j = \frac{1}{2\delta t}\overline{\rho}_j\left(\overline{u}_j - u_j\right)^2 + \frac{1}{\delta x_j}\left(\frac{(u_j - u_{j-1})^2}{2}\mathscr{F}^+_{j-\frac{1}{2}} - \frac{(u_{j+1} - u_j)^2}{2}\mathscr{F}^-_{j+\frac{1}{2}}\right)$$
$$+ \frac{1}{\delta x_j}(\overline{u}_j - u_j)(u_j - u_{j-1})\mathscr{F}^+_{j-\frac{1}{2}} + \frac{1}{\delta x_j}(\overline{u}_j - u_j)(u_{j+1} - u_j)\mathscr{F}^-_{j+\frac{1}{2}}.$$

It is thus quite natural to define the source term in the following way:

$$S_{j+\frac{1}{2}} = \frac{\delta x_{j+1}R_{j+1} + \delta x_j R_j}{2\delta x_{j+\frac{1}{2}}}.$$

The scheme presented above is close to the 1D version of the scheme presented in [3] but the two schemes differ by two points. Firstly, the mass fluxes in [3] are upwinded with respect to the material velocity (in other words, it corresponds to the choice $\mathscr{F}^{\pm}(\rho, e, u) = \pm\rho[u]^{\pm}$). Secondly, the time steppings are different: even if both schemes are explicit, the variables are not updated in the same order. We solve the discrete equations in the following way: $\rho \to u \to e$ whereas [3] proceeds with $\rho \to e \to u$. In particular, here the corrective term $S_{j+\frac{1}{2}}$ does not need any time shift since the updated velocity $\overline{u}$ is known when solving (6).

3 Stability Conditions

We now turn to the study of the stability conditions which ensure the positivity of the density and the internal energy.

Proposition 1 *Assuming that* $e_{j+\frac{1}{2}} \geqslant 0$, $\rho_{j+\frac{1}{2}} \geqslant 0$, $\forall j$ *and the following CFL-like conditions hold for all* j

$$\frac{\delta t}{\delta x_{j+\frac{1}{2}}}\left([u_{j+1}]^+ + \frac{c(e_{j+\frac{3}{2}}) + c(e_{j+\frac{1}{2}})}{\sqrt{2}} + [u_j]^- + \frac{c(e_{j+\frac{1}{2}}) + c(e_{j-\frac{1}{2}})}{\sqrt{2}}\right) \leqslant \frac{1}{\gamma}, \tag{7}$$

$$\frac{\delta t}{\delta x_{j+\frac{1}{2}}} c(e_{j+\frac{1}{2}+k}) \leqslant \frac{(\gamma-1)}{2\sqrt{2}}, \quad \forall k \in \{-1, 0, 1\}, \tag{8}$$

then $\overline{e}_{j+\frac{1}{2}} \geqslant 0$ *and* $\overline{\rho}_{j+\frac{1}{2}} \geqslant 0$.

Proof We assume that $e_{j+\frac{1}{2}} \geqslant 0$, $\rho_{j+\frac{1}{2}} \geqslant 0$ and that (7) and (8) holds for all j. We start by observing that:

$$[\lambda_\pm(e,u)]^\pm \leqslant [u]^\pm + c(e) \tag{9}$$

$$\sqrt{2}\, c(e_j) \leqslant c(e_{j-\frac{1}{2}}) + c(e_{j+\frac{1}{2}}) \tag{10}$$

Positivity of the density: As proved in [1], the positivity of $\overline{\rho}_{j+\frac{1}{2}}$ comes from the inequality

$$\frac{\delta t}{\delta x_{j+\frac{1}{2}}}\left([\lambda_+(e_{j+1}, u_{j+1})]^+ + [\lambda_-(e_j, u_j)]^-\right) \leqslant 1.$$

It is directly implied by (7) since $\gamma > 1$ and (9) holds.
Positivity of the internal energy: We rewrite the terms $(-1)^i \Pi_{j+\frac{1}{2}} \overline{u}_{j+i}$, $i \in \{0, 1\}$, involved in (6), by making the discrete time derivative $(\overline{u}_{j+i} - u_{j+i})$ appear. Then, we make use of the Young inequality as follows:

$$\begin{aligned}(-1)^i \Pi_{j+\frac{1}{2}} \overline{u}_{j+i} &= (-1)^i(\gamma-1)\left(\rho_{j+\frac{1}{2}} e_{j+\frac{1}{2}}\left(\overline{u}_{j+i} - u_{j+i}\right) + \rho_{j+\frac{1}{2}} e_{j+\frac{1}{2}} u_{j+i}\right) \\ &\geqslant -\rho_{j+\frac{1}{2}}\left(\frac{c(e_{j+\frac{1}{2}})}{2\sqrt{2}\gamma}(\overline{u}_{j+i} - u_{j+i})^2 + (\gamma-1)e_{j+\frac{1}{2}}\left(\frac{c(e_{j+\frac{1}{2}})}{\sqrt{2}} - (-1)^i u_{j+i}\right)\right).\end{aligned}$$

Next, we write $\overline{\rho}_{j+\frac{1}{2}} \overline{e}_{j+\frac{1}{2}} \geqslant \mathrm{T}_0 + \mathrm{T}_1^0 + \mathrm{T}_1^1$ where:

$$\mathrm{T}_0 = \rho_{j+\frac{1}{2}} e_{j+\frac{1}{2}}\left(1 - \frac{\delta t}{\delta x_{j+\frac{1}{2}}}(\gamma-1)\left(2\frac{c(e_{j+\frac{1}{2}})}{\sqrt{2}} - u_j + u_{j+1}\right)\right) - \delta t \frac{\mathscr{E}_{j+1} - \mathscr{E}_j}{\delta x_{j+\frac{1}{2}}},$$

$$\mathrm{T}_1^i = \frac{\delta t}{2}\frac{\delta x_{j+i}}{\delta x_{j+\frac{1}{2}}} R_{j+i} - \frac{\delta t}{\delta x_{j+\frac{1}{2}}}\frac{c(e_{j+\frac{1}{2}})}{2\sqrt{2}\gamma}\rho_{j+\frac{1}{2}}(\overline{u}_{j+i} - u_{j+i})^2.$$

Thus, to guarantee that $\overline{e}_{j+\frac{1}{2}}$ is non negative it is sufficient to ensure that these three terms are non negative. This holds under the assumptions (7) and (8).

Indeed, using the definition of the flux $\mathscr{E}_j$ and owing to (3), we obtain

$$\mathrm{T}_0 \geqslant \rho_{j+\frac12} e_{j+\frac12} \left(1 - \frac{\delta t}{\delta x_{j+\frac12}} (\gamma - 1) \left([u_j]^- + \frac{c(e_{j+\frac12})}{\sqrt{2}} + [u_{j+1}]^+ + \frac{c(e_{j+\frac12})}{\sqrt{2}} \right) \right) - \frac{\delta t}{\delta x_{j+\frac12}} \rho_{j+\frac12} e_{j+\frac12} \left([\lambda_+(e_{j+1}, u_{j+1})]^+ + [\lambda_-(e_j, u_j)]^- \right)$$

where, due to (7), the rhs is non negative by virtue of (9) and (10) .

Next, we turn to T_1^i. Using twice the Young inequality and bearing in mind the definition of $\overline{\rho}_j$, we observe that

$$\frac{\delta t}{2} \frac{\delta x_{j+i}}{\delta x_{j+\frac12}} R_{j+i} \geqslant \frac{\delta x_{j+i}}{4 \delta x_{j+\frac12}} (\overline{u}_{j+i} - u_{j+i})^2 \left(\rho_{j+i} - \frac{\delta t}{\delta x_{j+i}} (\mathscr{F}^+_{j+i+\frac12} - \mathscr{F}^-_{j+i-\frac12}) \right).$$

Hence, we have

$$\mathrm{T}_1^i \geqslant \frac{\delta x_{j+i}}{4 \delta x_{j+\frac12}} (\overline{u}_{j+i} - u_{j+i})^2 \left(\rho_{j+i} - \frac{\delta t}{\delta x_{j+i}} (\mathscr{F}^+_{j+i+\frac12} - \mathscr{F}^-_{j+i-\frac12}) - \frac{\delta t}{\gamma} \frac{2}{\sqrt{2}} \frac{\rho_{j+\frac12} c(e_{j+\frac12})}{\delta x_{j+i}} \right).$$

Coming back to (5), we write $T_1^i \geqslant \dfrac{(\overline{u}_{j+i} - u_{j+i})^2}{4 \delta x_{j+\frac12}} \left(T_2^{i,0} + T_2^{i,1} \right)$ where, for $k = 0, 1$,

$$\mathrm{T}_2^{i,k} = \frac{\delta x_{j+i+k-\frac12}}{2} \rho_{j+i+k-\frac12} - \delta t \frac{\mathscr{F}^+_{j+i+k} - \mathscr{F}^-_{j+i+k-1}}{2} - \frac{\delta t}{\gamma} \frac{2}{\sqrt{2}} \rho_{j+i+k-\frac12} c(e_{j+\frac12}).$$

Note that a non negative term has been added to obtain a symmetric formulation in the above inequality. Due to (3) and (7) we get

$$\mathscr{F}^+_{j+i+k} - \mathscr{F}^-_{j+i+k-1} \leqslant \frac{\delta x_{j+i+k-\frac12}}{\gamma \delta t} \rho_{j+i+k-\frac12},$$

and this allows us to write

$$\mathrm{T}_2^{i,k} \geqslant \frac{\delta x_{j+i+k-\frac12}}{2\gamma} \rho_{j+i+k-\frac12} \left(\gamma - 1 - \frac{\delta t}{\delta x_{j+i+k-\frac12}} \frac{4}{\sqrt{2}} c(e_{j+\frac12}) \right).$$

We conclude by observing that this term is non negative by virtue of (8).

4 Numerical Diffusion and Energy Conservation

It is worth discussing the expression of the numerical diffusion produced by our scheme, see also the appendix in [1]. Let us set the following non negative quantity

$$C_j = \begin{cases} -u_j & \text{if } u_j \leqslant -c(e_j) \\ \dfrac{u_j^2 + c(e_j)^2}{4c(e_j)} & \text{if } |u_j| < c(e_j) \\ u_j & \text{if } u_j > c(e_j) \end{cases}$$

and the following notations for averaged quantities

$$\{q\}_j = \frac{q_{j-\frac{1}{2}} + q_{j+\frac{1}{2}}}{2} \quad \text{and} \quad \{q\}_{j+\frac{1}{2}} = \frac{q_j + q_{j+1}}{2}.$$

Denoting $\mathscr{F}^{|.|} = \mathscr{F}^+ - \mathscr{F}^-$, wich is a positive quantity, the mass and the momentum fluxes can be cast as the sum of a centered term and a diffusion term:

$$\mathscr{F}_j = \{\rho\}_j u_j - \frac{C_j}{2}\left(\rho_{j+\frac{1}{2}} - \rho_{j-\frac{1}{2}}\right),$$
$$\mathscr{G}_{j+\frac{1}{2}} = \{\mathscr{F}\}_{j+\frac{1}{2}}\{u\}_{j+\frac{1}{2}} - \frac{\{\mathscr{F}^{|.|}\}_{j+\frac{1}{2}}}{2}\left(u_{j+1} - u_j\right).$$

Concerning the internal energy and kinetic energy fluxes, they become:

$$\mathscr{E}_j = \{\rho e\}_j u_j - \frac{C_j}{2}\left(e_{j+\frac{1}{2}}\rho_{j+\frac{1}{2}} - e_{j-\frac{1}{2}}\rho_{j-\frac{1}{2}}\right),$$
$$\mathscr{H}_{j+\frac{1}{2}} = \{\mathscr{F}\}_{j+\frac{1}{2}}\{\frac{u^2}{2}\}_{j+\frac{1}{2}} - \frac{\{\mathscr{F}^{|.|}\}_{j+\frac{1}{2}}}{2}\left(\frac{u_{j+1}^2}{2} - \frac{u_j^2}{2}\right).$$

As a by-product, it is remarkable that the scheme properly deals with 1D-contact discontinuities: if the discrete velocity and pressure are constant in the neighborhood of $x_{j+\frac{1}{2}}$, *ie* $u_{j-1} = u_j = u_{j+1} = u_{j+2} = u$ and $\Pi_{j-1/2} = \Pi_{j+1/2} = \Pi_{j+3/2} = \Pi$, then the scheme guaranties that they remain constant in the neighborhood of this point at the next time, *ie* $\overline{\Pi}_{j+1/2} = \Pi$ and $\overline{u}_{j+1} = u = \overline{u}_j$.

Let us now introduce the averaged energy in $x_{j+\frac{1}{2}}$ and x_j defined by:

$$E_{j+\frac{1}{2}} = e_{j+\frac{1}{2}} + \frac{\delta x_j \rho_j \dfrac{u_j^2}{2} + \delta x_{j+1}\rho_{j+1}\dfrac{u_{j+1}^2}{2}}{2\delta x_{j+\frac{1}{2}}\rho_{j+\frac{1}{2}}}$$

and

$$E_j = \frac{u_j^2}{2} + \frac{\delta x_{j+\frac{1}{2}}\rho_{j+\frac{1}{2}}e_{j+\frac{1}{2}} + \delta x_{j-\frac{1}{2}}\rho_{j-\frac{1}{2}}e_{j-\frac{1}{2}}}{2\delta x_j \rho_j}.$$

To obtain conservative equations for those quantities, we introduce the fluxes

$$\mathscr{T}_j = \mathscr{E}_j + \frac{\mathscr{H}_{j+\frac12} + \mathscr{H}_{j-\frac12}}{2} \text{ and } \mathscr{T}^*_{j+\frac12} = \frac{\mathscr{E}_{j+1} + \mathscr{E}_j}{2} + \mathscr{H}_{j+\frac12} - \frac{\delta x_{j+1} R_{j+1} - \delta x_j R_j}{4}.$$

Next we get the following consistent balance equations for $\rho_j E_j$ and $\rho_{j+\frac12} E_{j+\frac12}$:

$$\frac{\overline{\rho}_{j+\frac12}\overline{E}_{j+\frac12} - \rho_{j+\frac12} E_{j+\frac12}}{\delta t} + \frac{\mathscr{T}_{j+1} - \mathscr{T}_j}{\delta x_{j+\frac12}} + \frac{\overline{u}_{j+1}\{\Pi\}_{j+1} - \overline{u}_j\{\Pi\}_j}{\delta x_{j+\frac12}} = 0$$

and

$$\frac{\overline{\rho}_j \overline{E}_j - \rho_j E_j}{\delta t} + \frac{\mathscr{T}^*_{j+\frac12} - \mathscr{T}^*_{j-\frac12}}{\delta x_j} + \frac{\Pi_{j+\frac12}\{\overline{u}\}_{j+\frac12} - \Pi_{j-\frac12}\{\overline{u}\}_{j-\frac12}}{\delta x_j} = 0.$$

5 Numerical Simulations of Riemann Problems

We perform the numerical resolutions of some Riemann problems – see [4] – on the computational domain [0, 1]. The number of grid points is equal to 1000 and the time step is given by $\delta t = \delta x/100$. We take $\gamma = 1.4$. The initial data $\rho,\ u,\ p$ are piecewise constant functions with a discontinuity located at $x_0 = 0.5$, according to the table below. In Fig. 1, we represent the pressure $p_{j+\frac12}$, velocity u_j and internal energy $e_{j+\frac12}$ at the final time T (also given in the table below).

	ρ_l	ρ_r	u_l	u_r	p_l	p_r	T
Test #1	1	0.125	0	0	1	0.1	0.25
Test #2	1	1	0	0	1000	0.01	0.012
Test #3	5.99924	5.99242	19.5975	−6.19633	460.894	46.0950	0.035

Test #1, the so-called Sod test problem, is a mild test whose solution consists of a left rarefaction, a contact disconinuity and a right shock. Test #2 is a more severe test problem whose solution contains a left rarefaction, a contact discontinuity and a right shock. Test #3 corresponds to the collision of two strong shocks and consists of a left facing shock (travelling very slowly to the right), a right travelling contact discontinuity and a right travelling shock wave.

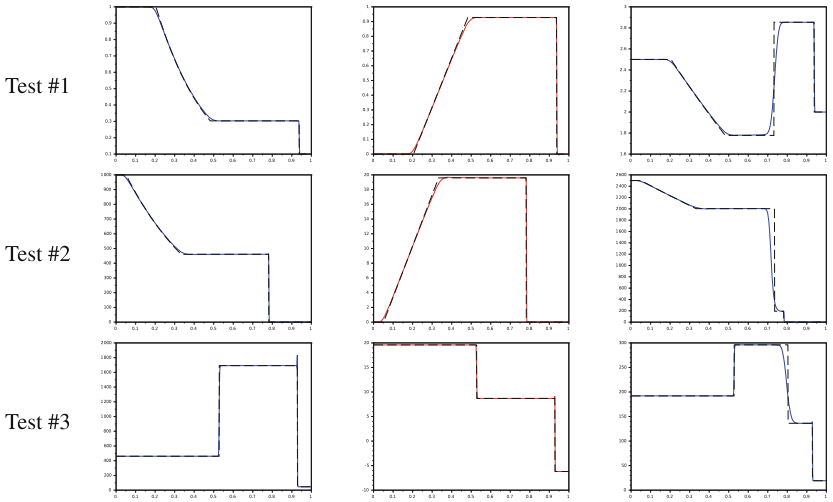


Fig. 1 Numerical (*solid lines*) and exact (*dotted lines*) solutions: pressure (*left*), velocity (*middle*), internal energy (*right*)

References

1. Berthelin, F., Goudon, T., Minjeaud, S.: Kinetic schemes on staggered grids for barotropic Euler models: entropy-stability analysis. Math. Comput. **84**, 2221–2262 (2015)
2. Herbin, R., Kheriji, W., Latche, J.C.: Staggered schemes for all speed flows. ESAIM Proc. **35**, 122–150 (2012)
3. Herbin, R., Latche, J.C., Nguyen, T.: Consistent Explicit Staggered Schemes for Compressible Flows. Part II: The Euler Equation. hal-00821069 (2013)
4. Toro, E.F.: Riemann Solvers and Numerical Methods for Fluid Dynamics, 3rd edn. Springer, Berlin (2009)

A Numerical Scheme for the Propagation of Internal Waves in an Oceanographic Model

Christian Bourdarias, Stéphane Gerbi and Ralph Lteif

Abstract In this paper, we introduce a new reformulation of the Green-Naghdi model in the Camassa-Holm regime for the propagation of internal waves over a flat topography to improve the frequency dispersion of the original model. We develop a second order splitting scheme where the hyperbolic part of the system is treated with a high-order finite volume scheme and the dispersive part is treated with a finite difference approach. Numerical simulations are then performed to validate the model.

Keywords Green Naghdi model · Nonlinear shallow water · Splitting method · Finite volume · Finite Difference · WENO reconstruction

1 Introduction

This study deals with the propagation of internal waves in the uni-dimensional setting located at the interface between two layers of fluids of different densities. The fluids are assumed to be incompressible, homogeneous, and immiscible, limited from above by a rigid lid and from below by a flat bottom. This type of fluid dynamics problem is encountered by researchers in oceanography when they study the wave near the shore. Because of the difference in the salinity of the different layers of water near the shore, it is useful to model the flow of salted water by a two layers incompressible fluids flow. The usual way of describing such a flow is to use the 3D-Euler equations

C. Bourdarias · S. Gerbi
LAMA, UMR 5127 CNRS, Université Savoie Mont Blanc,
73376 Le Bourget du Lac Cedex, France
e-mail: christian.bourdarias@univ-smb.fr

S. Gerbi
e-mail: stephane.gerbi@univ-smb.fr

R. Lteif (✉)
Laboratory of Mathematics-EDST, Lebanese University, Beirut, Lebanon
e-mail: ralphlteif_90@hotmail.com

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_11

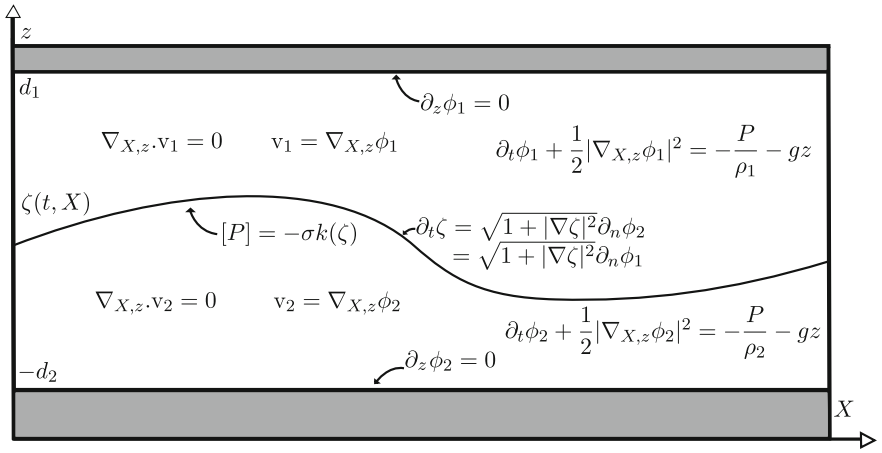


Fig. 1 Domain of study and governing equations

for the different layers adding some thermodynamic and dynamic conditions at the interface. This system will be called the *full Euler system* (Fig. 1) .

We introduce dimensionless variables and the two scale parameters μ, the shallowness parameter and ε, the nonlinearity parameter, defined by: $\mu = \frac{d_1^2}{\lambda^2}$, $\varepsilon = \frac{a}{d_1}$ where a is a typical length of the vertical oscillation of the interface, λ is a typical wavelength. We also define the dimensionless parameters $\gamma = \frac{\rho_1}{\rho_2}$ and $\delta = \frac{d_1}{d_2}$ representing respectively the ratio between the densities and the depth of the two layers.

In this work, we present a splitting technique for the numerical resolution of the GN model in the Camassa-Holm (or medium amplitude internal waves) regime, $\varepsilon = \mathcal{O}(\sqrt{\mu})$, obtained and fully justified by Duchêne, Israwi and Talhouk in [4]. In the Camassa-Holm regime, the authors has proved the existence and well-posedeness of the resulting system and its consistecy with the *full Euler system* in the sense that its solution remain close to the exact solution of the *full Euler system* with corresponding initial data up to the order $\mathcal{O}(\mu^2)$.

This model is first recast under a new formulation more suitable for numerical resolution with the same order of precision as the standard one but with improved frequency dispersion. For this sake, we introduce a one parameter family depending on $\alpha > 0$. The choice of the parameter α is motivated by the exact agreement between the phase velocity dispersion relation of the *full Euler system* and the improved Green-Naghdi system (1) for fixed values for γ and δ and for large wavenumbers (k around 4–5). This parameter is denoted α_{opt} and is obtained by an algebraic equation, see [2] for details.

We obtain the Green-Naghdi model in the Camassa-Holm with improved dispersion whose unknowns are ζ the mean elevation of the interface and v the shear mean velocity:

$$\begin{cases} \partial_t \zeta + \partial_x \big(f(\varepsilon\zeta)v\big) \;=\; 0, \\ (I + \mu\nu\alpha T[0])\big[\partial_t v + \varepsilon\varsigma v\partial_x v + \dfrac{\alpha-1}{\alpha}\big((\gamma+\delta)\partial_x\zeta + \varepsilon\partial_x(q_3(\varepsilon\zeta)v^2)\big)\big] \\ + \dfrac{1}{\alpha}\big((\gamma+\delta)\partial_x\zeta + \varepsilon\partial_x(q_3(\varepsilon\zeta)v^2)\big) + \mu\varepsilon Q_1(v) + \mu\varepsilon\nu Q_2(\zeta) + \mu\varepsilon\nu Q_3(\zeta) = 0. \end{cases} \tag{1}$$

with $f(X) = \frac{(1-X)(\delta^{-1}+X)}{1-X+\gamma(\delta^{-1}+X)}$, $T[0]V = -\partial_x^{2V}$, $S[\zeta]V = -\kappa_2\partial_x(\zeta\,\partial_x V)$, $q_3(\varepsilon\zeta) = \frac{1}{2}\big(f'(\varepsilon\zeta) - \varsigma\big)$,
$Q_1(v) = \kappa\,\partial_x((\partial_x v)^2)$, $Q_2(\zeta) = -S[\zeta](I + \mu\nu\alpha T[0])^{-1}\big((\gamma+\delta)\partial_x\zeta\big)$,
$Q_3(\zeta) = \kappa_1\zeta T[0](I + \mu\nu\alpha T[0])^{-1}[(\gamma+\delta)\partial_x\zeta]$.

2 Numerical Methods

As pointed out by many authors [1, 8] the improved dispersion Green-Naghdi Eqs. (1) is well-adapted to the implementation of a splitting scheme separating the hyperbolic and the dispersive parts of the equations.

2.1 The Splitting Method

We decompose the solution operator $S(.)$ associated to (1) at each time step Δt by the following second order operator splitting:

$$S(\Delta t) = S_1(\Delta t/2)S_2(\Delta t)S_1(\Delta t/2)$$

where $S_1(.)$ is the solution operator associated to the conservative part, and $S_2(.)$ the solution operator associated to the dispersive part of the Eq. (1). In this study, S_1 is computed using a finite volume method while S_2 is computed using a classical finite-difference method.

• $S_1(t)$ is the solution operator associated to the conservative part namely the nonlinear shallow water equations, NSWE:

$$\begin{cases} \partial_t \zeta + \partial_x \big(f(\varepsilon\zeta)v\big) \;=\; 0, \\ \partial_t v + \partial_x\Big(\dfrac{\varepsilon f'(\varepsilon\zeta)}{2}v^2 + (\gamma+\delta)\zeta\Big) = 0. \end{cases} \tag{2}$$

Under the hyperbolicity condition for the shallow water system provided in [6], this system is strictly hyperbolic.

• $S_2(t)$ is the solution operator associated to the remaining (dispersive) part of the equations.

$$\begin{cases} \partial_t \zeta \;=\; 0, \\ \\ (I + \mu\nu\alpha T[0])\big[\partial_t v - \dfrac{1}{\alpha}\big((\gamma+\delta)\partial_x\zeta + \varepsilon\partial_x(q_3(\varepsilon\zeta)v^2)\big)\big] \\ +\dfrac{1}{\alpha}\big((\gamma+\delta)\partial_x\zeta + \varepsilon\partial_x(q_3(\varepsilon\zeta)v^2)\big) + \mu\varepsilon Q_1(v) + \mu\varepsilon\nu Q_2(\zeta) + \mu\varepsilon\nu Q_3(\zeta) = 0. \end{cases} \tag{3}$$

2.2 Finite Volume Scheme

In what follows, we consider the numerical approximation of the hyperbolic system of conservation laws (2). We have constructed three finite volume schemes: first order, second order "MUSCL" type method and finally 5th order WENO method and tested their accuracy by using the exact (up to the order $\mathscr{O}(\mu^2)$) solitary wave solutions of the one layer Green-Naghdi equations over a flat bottom (see [8]). We do not present the results in this paper but the 5th order method is clearly much more accurate. However we do not obtain the predicted order with respect with the spatial mesh size. This might be due to the fact that the given analytic solution satisfies the model up to an $\mathscr{O}(\mu^2)$ remainder. We believe that this splitting strategy may be also applied in the variable bottom case. This is the subject of a future work.

2.2.1 Higher Order Finite-Volume Scheme: WENO5-RK4

To reach higher order accuracy in smooth regions and a good resolution around discontinuities, we implement fifth-order accuracy WENO reconstruction, following [7]. To automatically achieve high order accuracy and non-oscillatory property near discontinuities, WENO schemes use the idea of adaptive stencils in the reconstruction procedure based on the local smoothness of the numerical solution.

As far as time discretization is concerned, we use the fourth-order explicit RungeKutta "RK4" method.

2.3 Finite Difference Scheme for the Dispersive Part

The finite volume-finite difference mix imply to switch between the cell-averaged and nodal values for each unknown and at each time step. To this end, we use the fifth-order accuracy WENO reconstruction, that allows to approximate the nodal values (i.e. finite difference unknowns) $(U_i^n)_{i=1,N+1}$ in terms of the cell-averaged values (i.e. finite volume unknowns) $(\overline{U}_i^n)_{i=1,N}$.

The finite difference discretization of the system (3) leads to the following discrete problem:

$$\begin{cases} \dfrac{\zeta^{n+1}-\zeta^n}{\Delta t} = 0, \\ \dfrac{v^{n+1}-v^n}{\Delta t} - \dfrac{1}{\alpha}(\gamma+\delta)D_1(\zeta^n) - 2\dfrac{\varepsilon}{\alpha}q_3(\varepsilon\zeta^n)v^n D_1(v^n) - \dfrac{\varepsilon^2}{\alpha}q_3'(\varepsilon\zeta^n)D_1(\zeta^n)(v^n)(v^n) \\ +(I-\mu\nu\alpha D_2)^{-1}\Big[\dfrac{1}{\alpha}(\gamma+\delta)D_1(\zeta^n) + 2\dfrac{\varepsilon}{\alpha}q_3(\varepsilon\zeta^n)v^n D_1(v^n) + \dfrac{\varepsilon^2}{\alpha}q_3'(\varepsilon\zeta^n)D_1(\zeta^n)(v^n)(v^n) \\ +\mu\varepsilon Q_1(v^n) + \mu\nu\varepsilon Q_2(\zeta^n) + \mu\varepsilon\nu Q_3(\zeta^n)\Big] = 0, \end{cases} \tag{4}$$

with

$$Q_1(v^n) = 2\kappa D_1(v^n)D_2(v^n),$$

$$Q_2(\zeta^n) = \kappa_2 D_1\Big[\zeta^n D_1\Big((I-\mu\nu\alpha D_2)^{-1}(\gamma+\delta)D_1(\zeta^n)\Big)\Big],$$

$$Q_3(\zeta^n) = -\kappa_1\zeta^n D_2\Big[(I-\mu\nu\alpha D_2)^{-1}(\gamma+\delta)D_1(\zeta^n)\Big].$$

The system (4) is solved at each time step using a classical finite-difference technique, where the matrices D_1 and D_2 are the classical centered discretizations of the derivatives ∂_x and ∂_x^2 given below:

$$(\partial_x U)_i = \frac{1}{12\Delta x}(-U_{i+2} + 8U_{i+1} - 8U_{i-1} + U_{i-2}),$$

$$(\partial_x^2 U)_i = \frac{1}{12\Delta x^2}(-U_{i+2} + 16U_{i+1} - 30U_i + 16U_{i-1} - U_{i-2}).$$

For time discretization, the fourth-order formula "DF4" is associated to a fourth-order classical Runge-Kutta "RK4" scheme, and thus one obtains the "DF4-RK4" scheme.

We only treat either periodic boundary conditions or reflective boundary conditions for the hyperbolic and dispersive parts of the splitting scheme. Suitable relations are imposed on both cell-averaged and nodal quantities.

3 Numerical Validations: Kelvin-Helmholtz Instabilities

In this section, we present a numerical experiment to validate the numerical efficiency and accuracy of the improved Green-Naghdi model (1). We use the WENO5 reconstruction for the hyperbolic part of the splitting scheme and a fourth order finite difference scheme "DF4" for the dispersive part, both associated to a fourth-order classical Runge-Kutta "RK4" time scheme called "WENO5-DF4-RK4". We

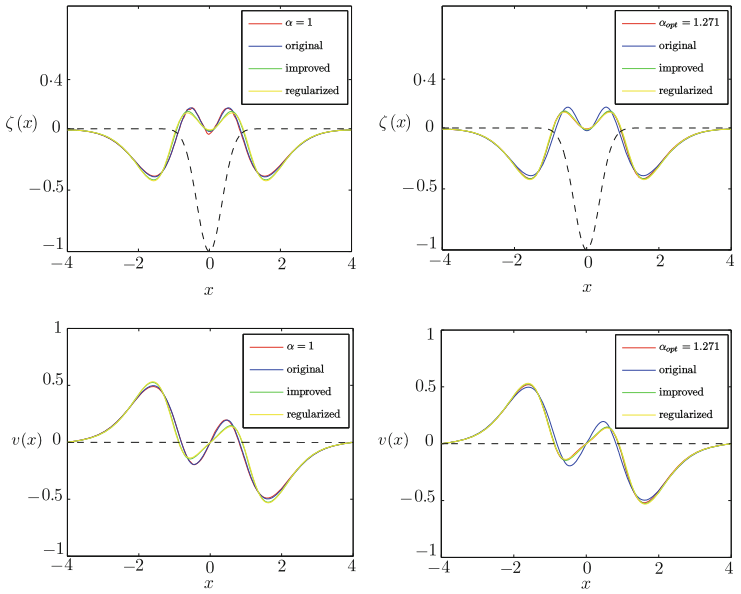


Fig. 2 Comparison with the Green-Naghdi models, with surface tension, at time $t = 2$, for $\alpha = 1$ (*left*) and $\alpha_{opt} = 1.271$ (*right*)

would like to highlight the importance of the choice of the parameter α in order to improve the frequency dispersion of the model (1), through the simulation of a sufficiently regular initial wave, following the numerical experiments performed in [5]. In the aforementioned paper they introduce a new class of Green-Naghdi type models for the propagation of internal waves with improved frequency dispersion in order to prevent high-frequency Kelvin-Helmholtz instabilities. These models are obtained by regularizing the original Green-Naghdi one by slightly modifying the dispersion components using a class of Fourier multipliers. They represent three different choices of the Fourier multipliers, each one yields to a specific Green-Naghi model which they denote as follows: **"original"** as the classical Green-Naghi model introduced in [3], **"regularized"** which is a well-posed system for sufficiently small and regular data, even in absence of surface tension, **"improved"** whose dispersion relation is the same as the one of the *full Euler system*. In order to compare with the numerical experiments done in [5], we choose the initial data $\zeta(0, x) = -e^{-4|x|^2}$ and $v(0, x) = 0$ (represented by the dashed lines). The computational domain is the interval $x \in (-4, 4)$ discretized with 512 cells using periodic boundary conditions.

The dimensionless parameters are set as follows: $\mu = 0.1$, $\varepsilon = 0.5$, $\delta = 0.5$, $\gamma = 0.95$. With thess values and choosing the wavenumber $k = 5$, we obtain $\alpha_{opt} = 1.271$.

Figures 2 and 3 show the comparisons between our numerical solution for $\alpha = 1$ (left) and $\alpha_{opt} = 1.271$ (right) and the Green-Naghdi models solutions obtained in [5], with a small amount of surface tension, at time $t = 2$ and $t = 3$ respectively.

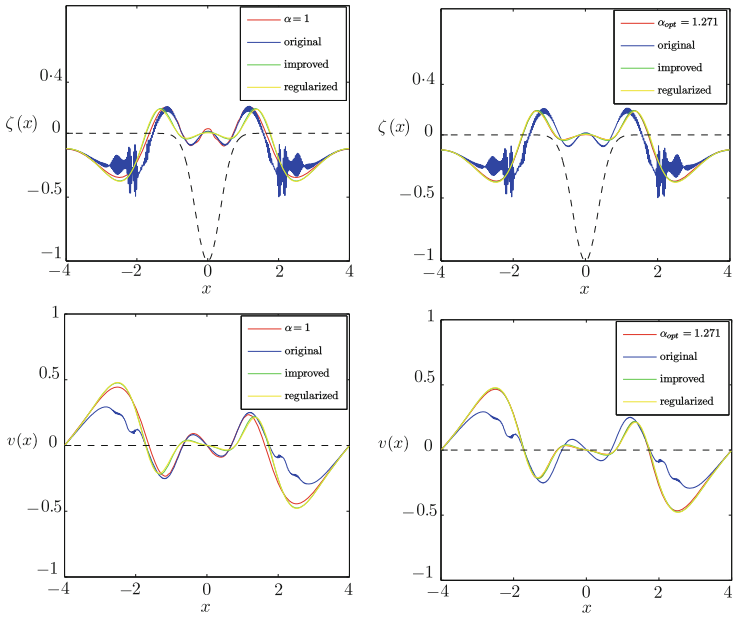


Fig. 3 Comparison with the Green-Naghdi models, with surface tension, at time $t = 3$, for $\alpha = 1$ (*left*) and $\alpha_{opt} = 1.271$ (*right*)

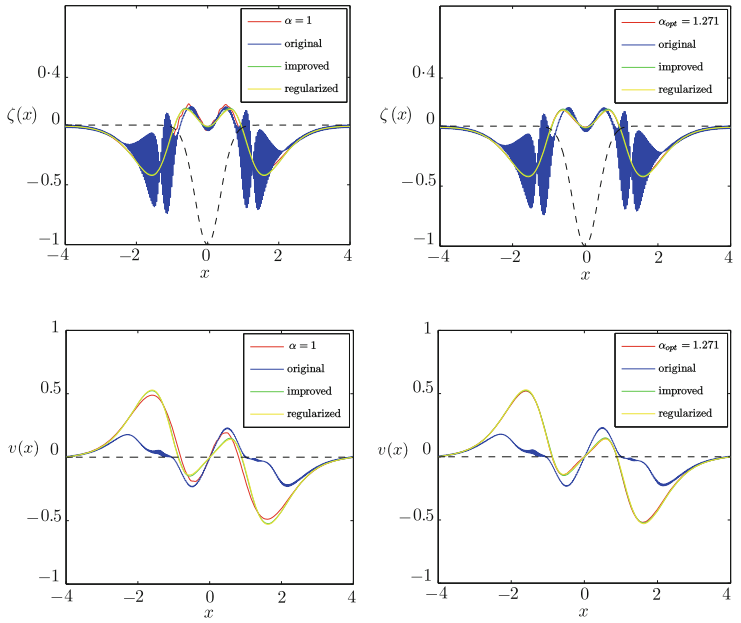


Fig. 4 Comparison with the Green-Naghdi models, without surface tension ($\text{bo}^{-1} = 0$), at time $t = 2$, for $\alpha = 1$ (*left*) and $\alpha_{opt} = 1.271$ (*right*)

We observe an excellent agreement between our numerical solution computed for $\alpha_{opt} = 1.271$ and both "improved" and "regularized" models at $t = 2$ and $t = 3$. As expected, at $t = 3$ the original model induces Kelvin-Helmholtz instabilities. Meanwhile, the flows predicted by the regularized and improved models and by our model (1) with $\alpha_{opt} = 1.271$ remain smooth and are very similar. Similarly, Fig. 4 shows an excellent agreement between the numerical solutions computed for $\alpha_{opt} = 1.271$ with the "improved" and "regularized" models, without surface tension at time $t = 2$, while the flow of the original model is completely destroyed due to Kelvin-Helmholtz instabilities.

The overall observations show the importance of the choice of the parameter α in improving the frequency dispersion. Indeed, when choosing $\alpha_{opt} = 1.271$, we observe an excellent matching between our numerical solutions and those obtained by the "improved" model before the latter is completely destroyed in absence of surface tension due to the Kelvin-Helmholtz instabilities. As well, our numerical solution matches the one computed by the "regularized" model even for a large time and with or without surface tension. This is not the case when choosing $\alpha = 1$. In fact, the "improved" model has exactly the same dispersion relation as the one of the *full Euler system* and for all wave numbers (see [5, Sect. 3] for more details) and the dispersion relation of the "regularized" model fit the one of the *full Euler system* to an $\mathcal{O}(\mu^3)$ order. This explains the reason behind the matching when choosing an optimal value for α and highlight the advantage of the proposed approach in improving the frequency dispersion.

References

1. Bonneton, P., Chazel, F., Lannes, D., Marche, F., Tissier, M.: A splitting approach for the fully nonlinear and weakly dispersive Green-Naghdi model. J. Comput. Phys. **230**(4), 1479–1498 (2011)
2. Bourdarias, C., Gerbi, S., Lteif, R.: A numerical scheme for an improved Green-Naghdi model in the Camassa-Holm regime for the propagation of internal waves. arXiv:1702.02810 (2017)
3. Choi, W., Camassa, R.: Fully nonlinear internal waves in a two-fluid system. J. Fluid Mech. **396**, 1–36 (1999)
4. Duchêne, V., Israwi, S., Talhouk, R.: A new fully justified asymptotic model for the propagation of internal waves in the Camassa-Holm regime. SIAM J. Math. Anal. **47**(1), 240–290 (2015)
5. Duchêne, V., Israwi, S., Talhouk, R.: A new class of two-layer Green-Naghdi systems with improved frequency dispersion. arXiv:1503.02397. To appear in Studies in Applied Mathematics (2017)
6. Guyenne, P., Lannes, D., Saut, J.C.: Well-posedness of the Cauchy problem for models of large amplitude internal waves. Nonlinearity **23**(2), 237–275 (2010)
7. Jiang, G.S., Shu, C.W.: Efficient implementation of weighted ENO schemes. J. Comput. Phys. **126**(1), 202–228 (1996)
8. Lannes, D., Marche, F.: A new class of fully nonlinear and weakly dispersive Green-Naghdi models for efficient 2D simulations. J. Comput. Phys. **282**, 238–268 (2015)

A Splitting Scheme for Three-Phase Flow Models

Hamza Boukili and Jean-Marc Hérard

Abstract A fractional step method that provides approximate solutions of a three-phase flow model is presented herein. The three-fluid model enables to handle smooth or discontinuous unsteady solutions. The numerical method is grounded on the use of the entropy inequality that governs smooth solutions of the set of PDEs. The evolution step relies on an explicit scheme, while implicit schemes are embedded in the relaxation step. The main properties of the scheme are given. Numerical approximations of two basic Riemann problems are eventually presented.

Keywords Three-phase flow · Entropy · Shocks · Vapour explosion · Finite volumes

1 Introduction

In order to perform numerical simulations of vapour explosion, a phenomenum resulting from the violent interaction between a hot liquid metal and a coolant (usually liquid water and its vapour), flow models with at least three phases are mandatory. Owing to the high velocity and high pressure levels arising in these situations, and also due to the occurence of strong shock waves, models should at least enable highly unsteady simulations, and should be such that unique and well defined jump conditions hold through discontinuities. However, only few contributions arise from the literature on that topic. Among these, one may at least mention [6–8, 10, 11]. Actually, we will focus here on the barotropic class defined in [8], and will present a possible fractional step method in order to compute approximate solutions of the

H. Boukili · J.-M. Hérard (✉)
EDF Lab Chatou, 6, Quai Watier, 78400 Chatou, France
e-mail: jean-marc.herard@edf.fr

H. Boukili
e-mail: hamza.boukili@edf.fr

H. Boukili · J.-M. Hérard
I2M, Aix Marseille Université, 39 Rue Joliot Curie, 13453 Marseille, France

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_12

latter model. Some among identified difficulties concern the way to cope with pressure relaxation effects and to preserve positive values of densities and statistical fractions; moreover, schemes should be such that they provide convergent and consistant approximations of shock patterns. Possible extensions to the non barotropic framework and problems arising with mass transfer terms are not addressed here.

2 Three-Phase Flow Model

Governing equations

In the sequel, $\alpha_k \in [0,1]$, $\rho_k, m_k = \alpha_k \rho_k, U_k$, respectively denote the mean statistical fraction, the mean density, the partial mass and the mean velocity of phase k (phase 1 denotes liquid metal). The mean pressure $P_k(\rho_k)$ is an increasing function with:

$$\lim_{x\to\infty} P_k(x) = +\infty \quad ; \quad \lim_{x\to 0} P_k(x) = 0$$

and we note as usual $c_k^2 = P_k'(\rho_k)$. The set of PDEs that is considered is (see [8]):

$$\begin{cases} \frac{\partial \alpha_k}{\partial t} + \mathscr{V}_i(W)\frac{\partial \alpha_k}{\partial x} = \phi_k(W)\ ; \\ \frac{\partial m_k}{\partial t} + \frac{\partial m_k U_k}{\partial x} = 0\ ; \\ \frac{\partial m_k U_k}{\partial t} + \frac{\partial m_k U_k^2 + \alpha_k P_k}{\partial x} + \Sigma_{l=1,l\neq k}^{3} \Pi_{kl}(W)\frac{\partial \alpha_l}{\partial x} = m_k S_k(W)\ . \end{cases} \tag{1}$$

It may be alternatively rewritten in a more condensed form:

$$\frac{\partial W}{\partial t} + \frac{\partial F(W)}{\partial x} + G(W)\frac{\partial H(W)}{\partial x} = S(W) \tag{2}$$

where the main variable W and fluxes $F(W)$, $H(W)$ are defined as:

$$W = (\alpha_2, \alpha_3, m_1, m_2, m_3, m_1U_1, m_2U_2, m_3U_3)^t$$

$$F(W) = (0, 0, m_1U_1, m_2U_2, m_3U_3, m_1U_1^2 + \alpha_1 P_1, m_2U_2^2 + \alpha_2 P_2, m_3U_3^2 + \alpha_3 P_3)^t$$

$$H(W) = (\alpha_2, \alpha_3, 0, 0, 0, 0, 0, 0)^t$$

$G(W)$ being implicitly defined by (1). The statistical fraction α_1 complies with: $\alpha_1 = 1 - \alpha_2 - \alpha_3$. We restrict herein to the case where: $\mathscr{V}_i(W) = U_1$, with (see [8]):

$$\begin{cases} \Pi_{12}(W) = \Pi_{21}(W) = \Pi_{23}(W) = P_2\ ; \\ \Pi_{13}(W) = \Pi_{31}(W) = \Pi_{32}(W) = P_3\ . \end{cases} \tag{3}$$

Closure laws for $\phi_k(W)$, $S_k(W)$ take the form:

$$\begin{cases} \phi_k(W) = d(W)\Sigma_{l=1}^{3}\left((P_k - P_l)\right) \ ; \\ m_k S_k(W) = \Sigma_{l=1}^{3}\left(e_{kl}(W)(U_l - U_k)\right) \end{cases} \tag{4}$$

where $d(W)$ and $e_{kl}(W) = e_{lk}(W)$ are positive bounded functions. Meaningful pressure relaxation time scales $d(W)$ arise from [5]. Other relaxation time scales $e_{kl}(W)$ embedded in momentum transfer terms may be found in the standard literature. We also define: $\psi_k'(\rho_k) = \frac{P_k(\rho_k)}{\rho_k^2}$, and the entropy of the mixture:

$$\eta = \Sigma_{k=1}^{3}\left(m_k U_k^2/2 + \psi_k(\rho_k)\right),$$

together with the entropy flux: $f_\eta(W) = \Sigma_{k=1}^{3}\left(\frac{U_k^2}{2} + \psi_k(\rho_k) + \frac{P_k}{\rho_k}\right) m_k U_k$. Actually this three-phase flow model inherits from similar properties as the Baer Nunziato two-phase flow model [1] (see [2, 3] for a slightly broader class).

Main properties

We recall first the main properties of the latter system (see [8]):

Property 1

- *Structure of the convective subset:*
  The homogeneous convective subset (left hand side of (2)) is hyperbolic unless $|U_1 - U_k| = c_k$. *Its eigenvalues are:*

$$\lambda_{0,1}(W) = U_1 \ \ ; \ \ \lambda_{2,3}(W) = U_1 \pm c_1 \ \ ; \ \ \lambda_{4,5}(W) = U_2 \pm c_2 \ \ ; \ \ \lambda_{6,7}(W) = U_3 \pm c_3$$

  The $0-1$*-wave is linearly degenerate, while other fields are genuinely non linear.*

- *Entropy inequality:*
  Smooth solutions of (2) comply with the entropy inequality:

$$\frac{\partial \eta(W)}{\partial t} + \frac{\partial f_\eta(W)}{\partial x} \leq 0. \tag{5}$$

- *Jump conditions:*
  Within each isolated wave, system (2) admits unique jump conditions.

We may now consider the $0-1$ coupling wave, which is the key point of the homogeneous model. Actually, six independent Riemann invariants arise which are given below. These will be used in order to construct exact solutions of the one-dimensional Riemann problem associated with (2) when neglecting source terms.

Proposition 1 *Riemann invariants of the* $0-1$ *coupling wave are:*

$$I^1_{0,1}(W) = U_1 \quad ; \quad I^2_{0,1}(W) = m_2(U_2 - U_1) \quad ; \quad I^3_{0,1}(W) = m_3(U_3 - U_1);$$
$$I^4_{0,1}(W) = \frac{(U_1 - U_2)^2}{2} + \int_0^{\rho_2} (\frac{c_2^2(x)}{x} dx) \quad ; \quad I^5_{0,1}(W) = \frac{(U_1 - U_3)^2}{2} + \int_0^{\rho_3} (\frac{c_3^2(x)}{x} dx);$$
$$I^6_{0,1}(W) = m_2(U_2 - U_1)^2 + m_3(U_3 - U_1)^2 + \Sigma_{k=1}^3(\alpha_k P_k).$$

The proof is straightforward though cumbersome.

3 Numerical Scheme

A fractional step method is introduced in order to compute approximate solutions of (2). The latter method complies with the entropy inequality (5). A Finite Volume scheme is built, considering a classical one-dimensional mesh, where Δx_i denotes the size of cell Ω_i. The first step involves an explicit scheme, whereas the scheme in the second -relaxation- step is implicit.

Time scheme

- **Step** 1. A first evolution step computes approximations of solutions of the convective subset; for given W_i^n, the state variable is updated following:

$$\begin{cases} \Delta x_i(W_i^{n+1,-} - W_i^n) + \Delta t^n \left(\mathscr{F}_{i+1/2}(W_i^n, W_{i+1}^n) - \mathscr{F}_{i-1/2}(W_{i-1}^n, W_i^n)\right) \\ \quad + \Delta t^n G(W_i^n) \left(\mathscr{H}_{i+1/2}(W_i^n, W_{i+1}^n) - \mathscr{H}_{i-1/2}(W_{i-1}^n, W_i^n)\right) = 0. \end{cases} \tag{6}$$

- **Step** 2. The second step takes all source terms into account, for given $W_i^{n+1,-}$, and computes W_i^{n+1} solution of:

$$(W_i^{n+1} - W_i^{n+1,-}) - \Delta t^n \mathscr{S}(W_i^{n+1,-}, W_i^{n+1}) = 0 \tag{7}$$

where: $\mathscr{S}(W_i^{n+1,-}, W_i^{n+1}) = (\phi_{i,2}, \phi_{i,3}, 0, 0, 0, S_{i,1}, S_{i,2}, S_{i,3})^t$, with:

$$\phi_{i,k} = \Sigma_{l=1}^3 \left(d(W_i^{n+1,-})(P_k(W_i^{n+1}) - P_l(W_i^{n+1})\right)$$

and:

$$S_{i,k} = \Sigma_{l=1}^3 \left(e_{kl}(W_i^{n+1,-})(U_l(W_i^{n+1}) - U_k(W_i^{n+1})\right)$$

for $k = 1, 2, 3$.

Numerical fluxes in the evolution step

We restrict herein to simple first-order Rusanov-type fluxes defined as follows:

$$\mathscr{F}_{ij}(W_i, W_j) = \left(F(W_i) + F(W_j) - R_{ij}(W_j - W_i)\right)/2\,,$$

together with:

$$\mathscr{H}_{ij}(W_i, W_j) = \left(H(W_i) + H(W_j)\right)/2$$

R_{ij} is defined as : $max_{i,j}(r(W_i), r(W_j))$, where $r(W)$ denotes the spectral radius of the whole jacobian matrix $(\frac{\partial F(W)}{\partial W} + G(W)\frac{\partial H(W)}{\partial W})$.

Property 2

- *For given strictly positive values* $(\alpha_k)_i^n$ and $(m_k)_i^n$, *the evolution step computes positive values* $(\alpha_k)_i^{n+1,-}$ and $(m_k)_i^{n+1,-}$ *if and only if the time step complies with the classical CFL-like condition:*

$$\Delta t^n max_{j=1\to N_{cell}}(R_{j-1/2} + R_{j+1/2})/(2\Delta x_j) = CFL < 1 \tag{8}$$

- *Assume that* $(\alpha_k)_i^{n+1,-}$ and $(m_k)_i^{n+1,-}$ *are positive. Then the discrete relaxation Step 2 computes a unique set of positive values* $(\alpha_k)_i^{n+1}$ and $(m_k)_i^{n+1}$, *and a unique set* $(U_1, U_2, U_3)_i^{n+1}$ *without any restriction on the time step.*

The proof for the first part involving the evolution Step 1 is classical. Actually, $(m_k)_i^{n+1,-}$ is a convex combination of partial masses $(m_k)_i^n$ and $(m_k)_{i\pm1}^n$, as soon as condition (8) holds. A similar result holds for $(\alpha_k)_i^{n+1,-}$. Moreover, when turning to Step 2, it may be easily checked that the linear system that provides $(U_1, U_2, U_3)^{n+1}$ admits a unique solution, since the determinant δ of the local discrete system:

$$\begin{aligned}\delta_i = m_1m_2m_3 + (\hat{e}_{13} + \hat{e}_{23})m_1m_2 + (\hat{e}_{12} + \hat{e}_{23})m_1m_3 + (\hat{e}_{12} + \hat{e}_{13})m_2m_3 \\ +(\hat{e}_{12}\hat{e}_{13} + \hat{e}_{12}\hat{e}_{23} + \hat{e}_{13}\hat{e}_{23})(m_1 + m_2 + m_3)\,,\end{aligned} \tag{9}$$

where $\hat{e}_{kl}$ and m_k respectively stand for $\Delta t^n e_{kl}(W_i^{n+1,-})$ and $(m_k)_i^{n+1,-}$, is strictly positive. Moreover, the relation $(m_k)_i^{n+1} = (m_k)_i^{n+1,-}$ guarantees positive values of partial masses. Eventually, the proof of existence and uniqueness of positive values of $(\alpha_k)_i^{n+1}$ is more intricate; it requires solving a non linear system with respect to $(x, y) = ((\alpha_2)_i^{n+1}, (\alpha_3)_i^{n+1})$ under the constraints: $x > 0, y > 0, 1 - x - y > 0$. We emphasize that similar schemes have been used for two-phase flow models [9].

4 Numerical Results

We focus here on simple EOS that read: $P_k(\rho_k) = P_k^0(\rho_k)^{\gamma_k}$. Two distinct Riemann problems are investigated, these being representative of what happens in water-vapour explosion. The time step complies with the CFL-like condition (8). We have

set in all cases: $CFL = 1/2$. The initial discontinuity separating states W_L and W_R is located at $x = 1/2$. We restrict here to uniform meshes, and we consider very large relaxation time scales, setting $d(W) = 0$ and $e_{kl}(W) = 0$.

Riemann problem 1: The first test case is a classical shock tube problem, where the initial data are such that velocities are null everywhere at the beginning of the computation, whatever the phase is. More precisely, we define W_L and W_R such that:

$$(\alpha_2)_L = 0.4 \quad ; \quad (\alpha_3)_L = 0.5 \quad ; \quad (\alpha_2)_R = 0.2 \quad ; \quad (\alpha_2)_R = 0.3;$$

$$(U_k)_L = (U_k)_R = 0. \quad ; \quad (\rho_k)_L = 1. \quad ; \quad (\rho_k)_R = 1/8.$$

where EOS are such that: $\gamma_1 = 7/5, \gamma_2 = 1.005, \gamma_3 = 1.001$ and $P_k^0 = 1.10^5$. Phasic pressures are plotted on Fig. 1, while $\mathscr{P} = I_{0,1}^6(W)$ and the effective pressure of the mixture acting on wall boundaries $P_{wall} = \Sigma_{k=1\to 3}\alpha_k P_k$ are given on Fig. 2. $\mathscr{P}$ is clearly well preserved through the right-going $0-1$-wave -which is located around $= 0.702$- unlike P_{wall}, which was expected. Velocity profiles have been added on Fig. 3. The finest mesh contains 80000 regular cells.

Riemann problem 2: The second test case is a simple Riemann problem where the initial data W_L and W_R are chosen such that:

$$I_{0,1}^m(W_L) = I_{0,1}^m(W_R)$$

for $m = 1 \to 6$, see property (3). EOS are such that: $\gamma_1 = 3/2, \gamma_2 = 2, \gamma_3 = 5/2$, and we still set: $P_k^0 = 1.10^5$. Actually, this is a very tough test case, which is much more discriminating than most of other Riemann problems that involve all waves. A simple though efficient way to measure errors in this particular case consists in computing the L^1 norm of independent variables $I_{0,1}^m(W)$. Obviously, and as expected, the rough

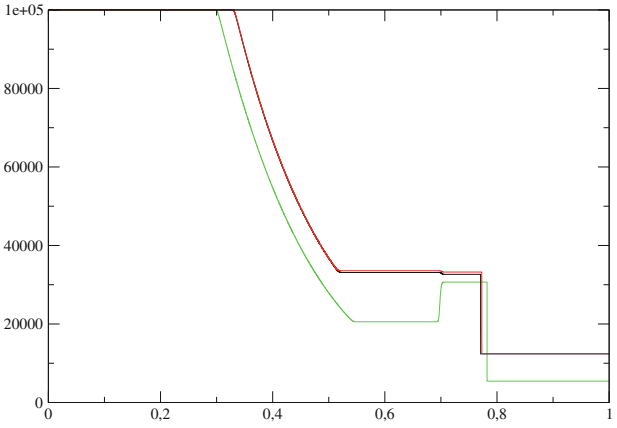


Fig. 1 Riemann problem 1. Pressure profiles on the finest mesh: P_1 (*green*), P_2 (*black*), P_3 (*red*)

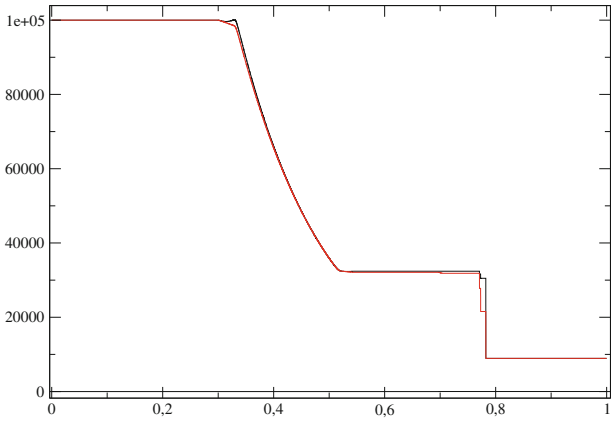


Fig. 2 Riemann problem 1. Pressure profiles on the finest mesh: $\mathscr{P} = I^6_{0,1}(W)$ (*black*), P_{wall} (*red*)

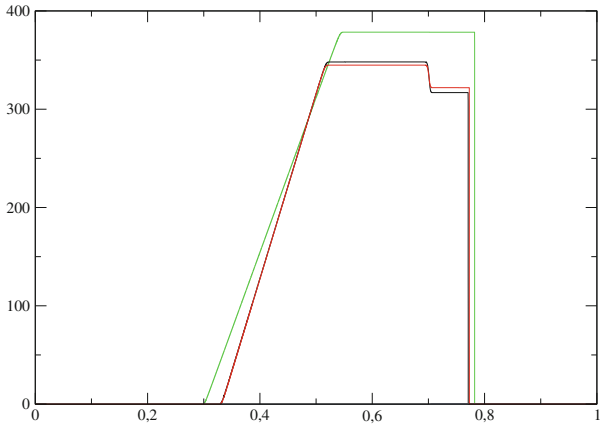


Fig. 3 Riemann problem 1. Velocity profiles on the finest mesh: U_1 (*green*), U_2 (*black*), U_3 (*red*)

Rusanov scheme yields rather high levels of error (close to 0.1% on the coarsest mesh, see Fig. 4). Nonetheless, and as expected, the error in L^1 norm varies as $h^{1/2}$, since the $0-1$-wave is LD (see Fig. 5).

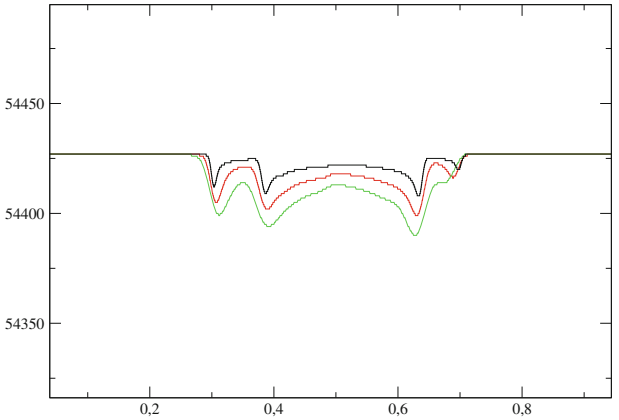


Fig. 4 Riemann problem 2. Pressure profiles for $\mathscr{P} = I^6_{0,1}(W)$ on three distinct meshes: 8000 cells (*black*), 2000 cells (*red*), 800 cells (*green*)

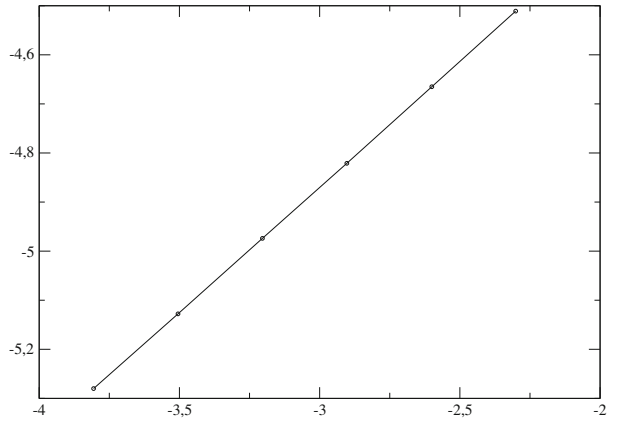


Fig. 5 Riemann problem 2. L^1 norm of the error for $\mathscr{P} = I^6_{0,1}(W)$ vs the mesh size h, using log/log scale. Coarsest and finest meshes contain 100 and 6400 regular cells respectively

Acknowledgements The first author receives financial support by ANRT through an EDF/CIFRE grant number 2016/0611. Computational facilities were provided by EDF.

References

1. Baer, M.R., Nunziato, J.W.: A two phase mixture theory for the deflagration to detonation transition (DDT) in reactive granular materials. Int. J. Multiphase Flow **12–6**, 861–889 (1986)
2. Bo, W., Jin, H., Kim, D., Liu, X., Lee, H., Pestiau, N., Yu, Y., Glimm, J., Grove, J.W.: Comparison and validation of multiphase closure models. Comput. Math. Appl. **56**, 1291–1302 (2008)

3. Coquel, F., Gallouët, T., Hérard, J.M., Seguin, N.: Closure laws for a two fluid two-pressure model. C. R. Acad. Sci. Paris, vol. I-332, pp. 927–932 (2002)
4. Flätten, T., Lund, H.: Relaxation two-phase flow models and the subcharacteristic condition. Math. Models Methods Appl. Sci., **21**(12), 2379–2407 (2011)
5. Gavrilyuk, S.: The structure of pressure relaxation terms: the one-velocity case, EDF report H-I83-2014-0276-EN (2014)
6. Giambo, S., La Rosa, V.: A hyperbolic three-phase relativistic flow model. ROMAI J. **11**, 89–104 (2015)
7. Hérard, J.M.: A hyperbolic three-phase flow model. Comptes Rendus Mathématique **342**, 779–784 (2006)
8. Hérard, J.M.: A class of compressible multiphase flow models. Comptes Rendus Mathématique **354**, 954–959 (2016)
9. Hérard, J.M., Hurisse, O.: A fractional step method to compute a class of compressible gas-liquid flows. Comput. Fluids **55**, 57–69 (2012)
10. Müller, S., Hantke, M., Richter, P.: Closure conditions for non-equilibrium multi-component models. Continuum Mech. Thermodynamics **28**, 1157–1190 (2016)
11. Romenski, E., Belozerov, A.A., Peshkov, I.M.: Conservative formulation for compressible multiphase flows, pp. 1–21 (2014). http://arxiv.org/abs/1405.3456

Modelling and Simulation of Non-hydrostatic Shallow Flows

M. J. Castro, C. Escalante and T. Morales de Luna

Abstract We consider the non-hydrostatic system derived by Yamazaki et al. for shallow flows. This model consists in the well known shallow water model which is coupled with two additional equations corresponding to non-hydrostatic terms. We develop a second-order well-balanced numerical method which combines finite-volume and finite-difference schemes. The numerical scheme has been implemented in GPUs and has been applied to idealized and challenging experimental test cases. The test cases show the accuracy and efficiency of the scheme.

Keywords Non-hydrostatic · Shallow-water · Finite-difference · Finite-volume · GPU

MSC (2010): 65N08 · 65N06 · 35Q35

1 Introduction

When modelling and simulating geophysical flows, the Nonlinear Shallow-Water equations, hereinafter SWE, is often a good choice as an approximation of the Navier-Stokes equations. Nevertheless, SWE do not take into account effects associated with dispersive waves. In recent years, effort has been done in the derivation of relatively simple mathematical models for shallow water flows that include long nonlinear water waves. Although such models are more expensive, from the computational point of view, the increasing computational power of current computers allows to

M.J. Castro · C. Escalante
Dpto. de Análisis Matemático, Universidad de Málaga, 29071 Málaga, Spain
e-mail: castro@anamat.cie.uma.es

C. Escalante
e-mail: escalante@uma.es

T. Morales de Luna (✉)
Dpto. de Matemáticas, Universidad de Córdoba, 14071 Córdoba, Spain
e-mail: tomas.morales@uco.es

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_13

consider Boussinesq Type Models. See for instance the works in [2, 4, 9–11, 13, 15] among others.

The challenge is to improve nonlinear dispersive properties of the model by including information on the vertical structure of the flow while designing fast and efficient algorithms for its simulation.

Here we shall use the approach introduced by Yamazaki in [16]. The model will be solved numerically using a two step algorithm: on a first step we solve the SWE in conservative form and on the second step we include the non-hydrostatic effects.

Numerical tests and comparison with experimental data show the accuracy and efficiency of the approach. The overall computational cost of the algorithm is no higher than 2.4 times the computational cost of classical SWE.

2 Description of the Model

We consider the non-hydrostatic model introduced in [16]. The governing equations are derived from the incompressible Navier-Stokes equations. As usual in shallow water models, the equations are obtained by a process of depth averaging on the vertical direction z. Nevertheless, opposed to what is done for SWE, the pressure is not assumed hydrostatic. Following Stelling and Zijlema and Casulli [8], total pressure is decomposed into a sum of hydrostatic and non-hydrostatic pressures. In this process, vertical velocity is assumed to have linear vertical profile. Moreover, for the vertical momentum equation, the vertical advective and dissipative terms are assumed small compared to their horizontal counterparts and thus neglected. The objective is to consider the easiest model that takes into account dispersive effects and the key point we seek is efficiency. Nevertheless what will be presented here can be easily adapted in the general case where these terms are included. Numerical simulations have shown that, for the purpose addressed here, they are not indeed relevant.

The resulting model can be written as

$$\begin{cases} \partial_t h + \nabla \cdot \boldsymbol{q} = 0, \\ \partial_t \boldsymbol{q} + \operatorname{div}\left(\dfrac{\boldsymbol{q} \otimes \boldsymbol{q}}{h}\right) + \nabla\left(\dfrac{1}{2} g h^2 + \dfrac{1}{2} h p\right) = (gh + p)\, \nabla H - \tau, \\ h \partial_t w = p, \\ h \nabla \cdot \boldsymbol{q} - \boldsymbol{q} \cdot \nabla (2\eta - h) + 2hw = 0. \end{cases} \tag{1}$$

where t is time and g is gravitational acceleration. $\mathbf{u} = (u, v)$ contains the depth averaged velocities components in the x and y directions respectively. w is the depth averaged velocity component in the z direction. $\mathbf{q} = h\mathbf{u}$ is the discharge vector in the x and y directions. p is the non-hydrostatic pressure at the bottom. The flow depth is $h = \eta + H$ where η is the surface elevation measured from the still-water level, H is the still water depth and τ is a friction law term (see Fig. 1). Operators ∇

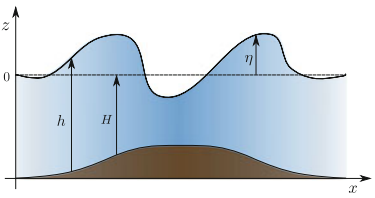


Fig. 1 Sketch and description of the variables

and $\nabla\cdot$ denote the gradient vector field and the divergence respectively in the (x, y) direction.

3 Numerical Scheme

For the sake of simplicity, we shall consider here just the one-dimensional case. System (1) can then be written in the compact form

$$\begin{cases} \partial_t \boldsymbol{U} + \partial_x \boldsymbol{F}_{SW}(\boldsymbol{U}) = \boldsymbol{G}_{SW}(\boldsymbol{U})\partial_x H + \mathscr{T}_{NH}(h, \partial_x h, H, \partial_x H, p, \partial_x p) - \boldsymbol{\tau}, \\ h\partial_t w = p, \\ \mathscr{B}(\boldsymbol{U}, \partial_x \boldsymbol{U}, H, \partial_x H, w) = 0, \end{cases} \tag{2}$$

where we introduce the notation

$$\boldsymbol{U} = \begin{pmatrix} h \\ q \end{pmatrix}, \ \boldsymbol{F}_{SW}(\boldsymbol{U}) = \begin{pmatrix} q \\ \dfrac{q^2}{h} + \dfrac{1}{2}gh^2 \end{pmatrix}, \ \boldsymbol{G}_{SW}(\boldsymbol{U}) = \begin{pmatrix} 0 \\ gh \end{pmatrix},$$

and for the friction term vector, the well-known Manning empirical formula is used.

Finally,

$$\mathscr{T}_{NH}(h, \partial_x h, H, \partial_x H, p, \partial_x p) = \begin{pmatrix} 0 \\ -\dfrac{1}{2}\left(h\partial_x p + p\partial_x(2\eta - h)\right) \end{pmatrix},$$

and

$$\mathscr{B}(\boldsymbol{U}, \boldsymbol{U}_x, H, H_x, w) = h\partial_x q - q\partial_x (2\eta - h) + 2hw.$$

The model is now solved numerically using a two-step algorithm: first the hyperbolic problem (SWE) is solved, then, in a second step, non-hydrostatic terms will be taken into account.

Remark 1 As pointed out in [14], in shallow water, complex events can be observed related to turbulent processes. One of these processes corresponds to the breaking of waves near the coast. The model presented here cannot describe this process without

an additional term which allows the model dissipate the required amount of energy on such situations. We refer to [5] for further details.

3.1 First Step: SWE

We consider first the hyperbolic problem (SWE) given by

$$\partial_t \boldsymbol{U} + \partial_x \boldsymbol{F}_{SW}(\boldsymbol{U}) = \boldsymbol{G}_{SW}(\boldsymbol{U})\partial_x H. \tag{3}$$

This system is solved numerically by using a finite volume method. As usual, we subdivide the horizontal spatial domain into standard computational cells $I_i = [x_{i-1/2}, x_{i+1/2}]$ with lengths Δx_i and define

$$\boldsymbol{U}_i(t) = \frac{1}{\Delta x_i} \int_{I_i} \mathbf{U}(x,t)dx,$$

the cell average of the function $\boldsymbol{U}(x,t)$ on cell I_i at time t. We shall also denote by x_i the center of the cell I_i. For the sake of simplicity, let us assume that all cells have the same length Δx.

Then, an efficient second-order well-balanced PVM path-conservative finite-volume method [6, 12] is applied.

For the sake of brevity, we omit here the details and refer to [5, 6] for the detail.

3.2 Second Step: Non-hydrostatic Terms

Regarding non-hydrostatic terms, we consider a staggered-grid formed by the points $x_{i-1/2}$, $x_{i+1/2}$ of the interfaces for each cell I_i, and denote the point values of the functions p and w on point $x_{i+1/2}$ at time t by

$$p_{i+1/2}(t) = p(x_{i+1/2}, t),\ w_{i+1/2}(t) = w(x_{i+1/2}, t).$$

Now, a second order compact finite-difference scheme is applied to

$$\begin{cases} \partial_t \boldsymbol{U} = \mathscr{T}_{NH}(h, \partial_x h, H, \partial_x H, p, \partial_x p) - \boldsymbol{\tau}, \\ h\partial_t w = p, \\ \mathscr{B}(\boldsymbol{U}, \boldsymbol{U}_x, H, H_x, w) = 0, \end{cases} \tag{4}$$

where the values obtained in previous step are used as initial condition for the system.

The resulting linear system is solved using an iterative Jacobi method combined with a scheduled relaxation.

3.3 2D and GPU Implementation

The described numerical scheme can be easily adapted to the 2D system (1). In this case, the computational domain is decomposed into subsets with a simple geometry, called cells or finite volumes. We use one common arrangement of the variables, known as the Arakawa C-grid (see [5] for the details).

The first step of the algorithm adapts well to GPUs architectures as is shown in [7]. Moreover, the compactness of the numerical stencil and the easy parallelization of the Jacobi method makes that the second step can also be *easily* implemented on GPUs.

4 Numerical Tests

4.1 Periodic Waves Breaking over a Submerged Bar

The experiment of plunging breaking periodic waves over a submerged bar by Beji and Battjes [1] is considered here. The numerical test is performed in a one-dimensional channel with a trapezoidal obstacle submerged. Waves in the free surface are measured in seven point stations S_0, $S_1, \ldots, S_6$ (See [1] for the details).

Figure 2 shows the time evolution of the free surface at points $S_1, \ldots, S_6$. The comparison with experimental data emphasizes the need to consider a dispersive model to faithfully capture the shape of the waves near the continental slope. Both amplitude and frequency of the waves are captured on all wave gauges successfully.

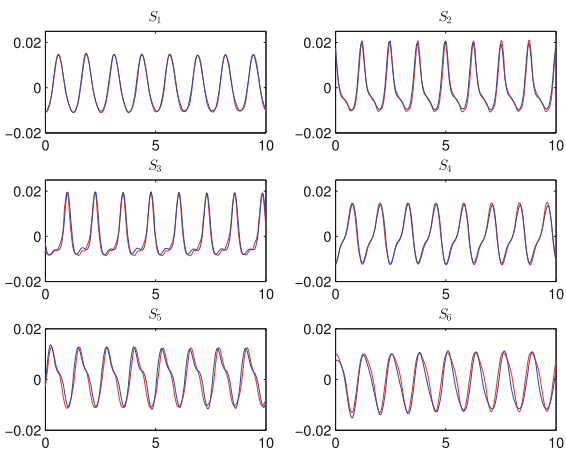
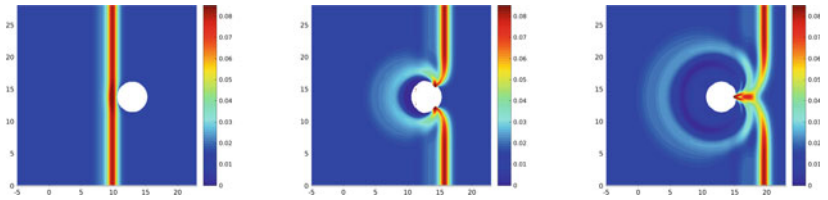


Fig. 2 Comparison of data time series (*red*) and numerical (*blue*) at wave gauges $S_1,\ S_2,\ S_3,\ S_4,\ S_5,\ S_6$



(a) $T = 5\ s$ (b) $T = 8\ s$ (c) $T = 10\ s$

Fig. 3 Computed free surface at different times

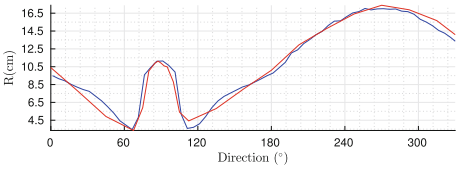


Fig. 4 Maximum run-up measured (*red*) and simulated (*blue*)

4.2 Solitary Wave on a Conical Island

We compare now to the experimental data obtained at the Coastal and Hydraulic Laboratory, Engineer Research and Development Center of the U.S. Army Corps of Engineers ([3]). The laboratory experiment consists in an idealized representation of Babi Island, in the Flores Sea, in Indonesia.

A directional wave-maker is used to produce planar solitary waves of specified crest lengths and heights (See [3] for the details).

Numerical simulation shows two wave fronts splitting in front of the island and collide behind it (See Fig. 3). Comparison with measured and computed water level at gauges WG_1, WG_2, WG_3, WG_4 shows good results, as well as comparison between computed run-up and laboratory measurement (Fig. 4).

4.3 Circular Dam-Break

We consider a 2D problem consisting in a circular dam-break in the $[-5, 5] \times [-5, 5]$ domain (See [5] for the details).

The goal of this numerical test, is to compare the execution times in seconds for the SWE and non-hydrostatic GPU codes for different mesh sizes. Simulations are carried out in the time interval $[0, 1]$. CFL parameter is set to 0.9 and open boundary conditions are considered. Due to maximum page limit of this contribution, the corresponding figures could not be included but we refer to [5] for a complete description. We include here just the execution times for both codes, shown in Table 1. As we see, the additional computation cost of the non-hydrostatic model with the algorithm described here is only 2.4 times that of a SWE code.

Table 1 Execution times in sec for SWE and NH GPU implementations

Number of volumes	Runtime (s)	
	SWE	Non-hydrostatic
250×250	0.64	0.64
500×500	2.29	5.79
750×750	7.17	17.33
1000×1000	16.75	40.47
1250×1250	33.88	79.67
1500×1500	56.38	136.12

5 Conclusions

A non-hydrostatic model has been considered in order to incorporate dispersive effects in the propagation of waves in a homogeneous, inviscid and incompressible fluid.

The numerical scheme employed, combines a finite volume path-conservative scheme for the underlying hyperbolic system and finite differences for the discretization of non-hydrostatic terms and a GPU implementation is carried out.

It can be stated that the scheme presented here is efficient and can model dispersive effects with a moderate computational cost. Computational times for the non-hydrostatic code are no higher than 2.4 SWE times for refined meshes. To our knowledge, the approach presented here is one of the most efficient from the computational point of view.

Acknowledgements This research has been supported by the Spanish Government through the Research projects MTM2015-70490-C2-1-R, MTM2015-70490-C2-2-R.

References

1. Beji, S., Battjes, J.: Experimental investigation of wave propagation over a bar. Coast. Eng. **19**, 151–162 (1993)
2. Boussinesq, J.: Théorie des ondes et des remous qui se propagent le long dun canal rectangulaire horizontal, en communiquant au liquide contenu dans ce canal des vitesses sensiblement pareilles de la surface au fond. J. Mathmatiques Pures Appliquèes **17**, 55–108 (1872)
3. Briggs., M., Synolakis, C., Harkins, G., Green, D.: Laboratory experiments of tsunami runup on a circular island. Pure Appl. Geophys. 144(3): 569–593 (1995)
4. Bristeau, M.O., Sainte-Marie, J.: Derivation of a non-hydrostatic shallow water model; comparison with saint-venant and Boussinesq systems. Discrete Continuous Dyn. Syst. Ser. B **10**(4), 733–759 (2008)
5. Castro, M., Escalante, C., Morales de Luna, T.: Non-hydrostatic pressure shallow flows: GPU implementation using finite-volume and finite-difference scheme. Preprint (2017)
6. Castro, M., Fernández-Nieto, E.: A class of computationally fast first order finite volume solvers: PVM methods. SIAM J. Sci. Comput. **34**(4), 173–196 (2012)

7. Castro, M.J., Ortega, S., de la Asunción, M., Mantas, J.M., Gallardo, J.M.: GPU computing for shallow water flow simulation based on finite volume schemes. CR Mécanique **339**(2–3), 165–184 (2011)
8. Casulli, V.: A semi-implicit finite difference method for non-hydrostatic free surface flows. Numer. Meth. Fluids **30**(4), 425–440 (1999)
9. Green, A., Naghdi, P.: A derivation of equations for wave propagation in water of variable depth. Fluid Mech. **78**, 237–246 (1976)
10. Madsen, P., Sorensen, O.: A new form of the Boussinesq equations with improved linear dispersion characteristics. part 2: A slowing varying bathymetry. Coast. Eng. **18**, 183–204 (1992)
11. Nwogu, O.: An alternative form of the Boussinesq equations for nearshore wave propagation. Waterway Port Coastal Ocean Eng. **119**, 618–638 (1994)
12. Parés, C., Castro, M.J.: On the well-balance property of Roe's method for nonconservative hyperbolic systems. Applications to shallow-water systems. ESAIM. Math. Model. Numer. Anal. **38**(5), 821–852 (2004)
13. Peregrine, D.: Long waves on a beach. Fluid Mech. **27**(4), 815–827 (1967)
14. Roeber, V., Cheung, K.F., Kobayashi, M.H.: Shock-capturing Boussinesq-type model for nearshore wave processes. Coast. Eng. **57**, 407–423 (2010)
15. Witting, J.: A unified model for the evolution nonlinear water waves. J. Comput. Phys. **56**(2), 203–236 (1984)
16. Yamazaki, Y., Kowalik, Z., Cheung, K.: Depth-integrated, non-hydrostatic model for wave breaking and run-up. Numer. Meth. Fluids **61**, 473–497 (2008)

A Flux Splitting Method for the Baer-Nunziato Equations of Compressible Two-Phase Flow

Svetlana Tokareva and Eleuterio Toro

Abstract We extend the Toro-Vázquez flux vector splitting approach (TV), originally proposed for the ideal 1D Euler equations in [11], to the Baer-Nunziato equations of compressible two-phase flow. Following the TV approach we identify corresponding advection and pressure operators and assess the TV flux splitting in the setting of finite volume and path-conservative methods in terms of accuracy and efficiency.

Keywords Compressible multiphase flow · Non-conservative systems · Flux splitting

MSC (2010): 65M08 · 76T99

1 Introduction

The Baer-Nunziato model was first proposed in [1] in the context of granular energetic combustible materials embedded in gaseous combustion products. A distinctive feature of the Baer-Nunziato model is the admission of two velocity vectors and two pressures. The equations are hyperbolic, except for some well identified situations, and the complete mathematical structure of the 1D system, as well as split 3D system, is available [6, 8]. The homogeneous one-dimensional Baer-Nunziato equations are a time-dependent system of seven partial differential equations:

S. Tokareva (✉)
Institute of Mathematics, University of Zurich, Winterthurerstrasse 190,
8057 Zurich, Switzerland
e-mail: svetlana.tokareva@math.uzh.ch

E. Toro
Laboratory of Applied Mathematics, DICAM, University of Trento, via Mesiano, 77,
38123 Trento, Italy
e-mail: eleuterio.toro@unitn.it

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII - Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_14

$$\partial_t \mathbf{Q} + \partial_x \mathbf{F}(\mathbf{Q}) + \mathbf{T}(\mathbf{Q})\partial_x \bar{\alpha} = \mathbf{0} \tag{1}$$

with

$$\mathbf{Q} = \begin{bmatrix} \bar{\alpha} \\ \bar{\alpha}\bar{\rho} \\ \bar{\alpha}\bar{\rho}\bar{u} \\ \bar{\alpha}\bar{\rho}\bar{E} \\ \alpha\rho \\ \alpha\rho u \\ \alpha\rho E \end{bmatrix}, \quad \mathbf{F}(\mathbf{Q}) = \begin{bmatrix} 0 \\ \bar{\alpha}\bar{\rho}\bar{u} \\ \bar{\alpha}\left(\bar{\rho}\bar{u}^2 + \bar{p}\right) \\ \bar{\alpha}\bar{u}\left(\bar{\rho}\bar{E} + \bar{p}\right) \\ \alpha\rho u \\ \alpha\left(\rho u^2 + p\right) \\ \alpha u\left(\rho E + p\right) \end{bmatrix}, \quad \mathbf{T}(\mathbf{Q}) = \begin{bmatrix} \bar{u} \\ 0 \\ -p \\ -p\bar{u} \\ 0 \\ p \\ p\bar{u} \end{bmatrix}.$$

Here ρ, u, p, E are gas density, velocity, pressure and specific total energy, and $\bar{\rho}$, $\bar{u}$, $\bar{p}$, $\bar{E}$ are the corresponding variables for the solid phase; α and $\bar{\alpha}$ are volume fractions. The specific total energies of the phases are expressed as $E = e + \frac{1}{2}u^2$ and $\bar{E} = \bar{e} + \frac{1}{2}\bar{u}^2$, where e and $\bar{e}$ are specific internal energies. System (1) requires additional closure relations involving density, internal energy and pressure of each phase. Such relations are provided by the equations of state (EOS). An ideal EOS for the gas phase and a stiffened EOS for the solid phase are frequently used, namely $p = (\gamma - 1)\rho e$, $\bar{p} = (\bar{\gamma} - 1)\bar{\rho}\bar{e} - \bar{\gamma}\bar{P}_0$, where γ and $\bar{\gamma}$ are the specific heat ratios of the gas and solid phases, respectively, and $\bar{P}_0$ is a known constant. The volume fractions are related through the saturation condition: $\bar{\alpha} + \alpha = 1$.

2 TV Flux Splitting Method for the Baer-Nunziato Equations

Consider the homogeneous one-dimensional Baer-Nunziato equations (1). We follow the Toro-Vázquez (TV) flux splitting approach [11] taking into account that the equations of interest here do not have a conservation-law form. First, we identify the conservative part and express the conservative flux as the sum of advection and pressure fluxes as follows:

$$\mathbf{F}(\mathbf{Q}) = \begin{bmatrix} 0 \\ \bar{\alpha}\bar{\rho}\bar{u} \\ \bar{\alpha}\left(\bar{\rho}\bar{u}^2 + \bar{p}\right) \\ \bar{\alpha}\bar{u}\left(\frac{1}{2}\bar{\rho}\bar{u}^2 + \bar{\rho}\bar{e} + \bar{p}\right) \\ \alpha\rho u \\ \alpha\left(\rho u^2 + p\right) \\ \alpha u\left(\frac{1}{2}\rho u^2 + \rho e + p\right) \end{bmatrix} = \begin{bmatrix} 0 \\ \bar{\alpha}\bar{\rho}\bar{u} \\ \bar{\alpha}\bar{\rho}\bar{u}^2 \\ \frac{1}{2}\bar{\alpha}\bar{\rho}\bar{u}^3 \\ \alpha\rho u \\ \alpha\rho u^2 \\ \frac{1}{2}\alpha\rho u^3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \bar{\alpha}\bar{p} \\ \bar{\alpha}\bar{u}\left(\bar{\rho}\bar{e} + \bar{p}\right) \\ 0 \\ \alpha p \\ \alpha u\left(\rho e + p\right) \end{bmatrix}, \tag{2}$$

with the respective advection and pressure fluxes defined as

$$\mathbf{A}(\mathbf{Q}) = \left[0, \bar{\alpha}\bar{\rho}\bar{u}, \bar{\alpha}\bar{\rho}\bar{u}^2, \tfrac{1}{2}\bar{\alpha}\bar{\rho}\bar{u}^3, \alpha\rho u, \alpha\rho u^2, \tfrac{1}{2}\alpha\rho u^3\right]^T, \tag{3}$$

$$\mathbf{P}(\mathbf{Q}) = \left[0, 0, \bar{\alpha}\bar{p}, \bar{\alpha}\bar{u}\left(\bar{\rho}\bar{e} + \bar{p}\right), 0, \alpha p, \alpha u\left(\rho e + p\right)\right]^T. \tag{4}$$

Following [10, 11], we consider two systems, the advection system (A-system) and the pressure system (P-system), noting however that here the pressure system is augmented by the nonconservative term present in the Baer-Nunziato equations. Thus, the two systems are

$$\begin{aligned} \partial_t \mathbf{Q} + \partial_x \mathbf{A}(\mathbf{Q}) &= \mathbf{0}, \quad \text{(advection system, conservative)} \\ \partial_t \mathbf{Q} + \partial_x \mathbf{P}(\mathbf{Q}) + \mathbf{T}(\mathbf{Q})\partial_x \bar{\alpha} &= \mathbf{0}. \quad \text{(pressure system, non-conservative)} \end{aligned} \tag{5}$$

The TV flux splitting approach consists of approximating the numerical fluxes for the pressure system and advection system separately and constructing the numerical fluxes for the full system based on these.

The construction of the numerical flux corresponding to the advection system is straightforward and follows directly from [11]. We now turn our attention to the non-conservative pressure system by considering the associated Riemann problem written in primitive variables

$$\partial_t \mathbf{V} + \mathbf{B}(\mathbf{V})\partial_x \mathbf{V} = \mathbf{0}, \quad \mathbf{V}(x,0) = \begin{cases} \mathbf{V}_L, & \text{if } x < 0, \\ \mathbf{V}_R, & \text{if } x > 0, \end{cases} \tag{6}$$

with

$$\mathbf{V} = \begin{bmatrix} \bar{\alpha} \\ \bar{\rho} \\ \bar{u} \\ \bar{p} \\ \rho \\ u \\ p \end{bmatrix}, \quad \mathbf{B}(\mathbf{V}) = \begin{bmatrix} \bar{u} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{\bar{\rho}\bar{u}}{\bar{\alpha}} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{\Delta p}{\bar{\alpha}\bar{\rho}} & 0 & 0 & \frac{1}{\bar{\rho}} & 0 & 0 & 0 \\ \frac{\bar{u}}{\bar{\alpha}\bar{e}_{\bar{p}}}\left(\bar{e}_{\bar{p}}\bar{\rho} + \bar{e}\right) & \frac{\bar{u}(\bar{\rho}\bar{e}_{\bar{p}} + \bar{e})}{\bar{\rho}\bar{e}_{\bar{p}}} & \frac{\bar{h}}{\bar{e}_{\bar{p}}} & \bar{u} & 0 & 0 & 0 \\ \frac{\rho\bar{u}}{\alpha} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{\rho} \\ \frac{1}{\alpha e_p}\left(-\bar{u}e_\rho\rho - \frac{p\Delta u}{\rho} - ue\right) & 0 & 0 & 0 & \frac{u(\rho e_\rho + e)}{\rho e_p} & \frac{h}{e_p} & u \end{bmatrix}$$

where $h = e + p/\rho$ and $\bar{h} = \bar{e} + \bar{p}/\bar{\rho}$ are specific enthalpies of the gas and solid phase, respectively, and $\Delta p = p - \bar{p}$, $\Delta u = u - \bar{u}$. The eigenvalues of the matrix $\mathbf{B}(\mathbf{V})$ are

$$\lambda_1 = \frac{1}{2}(u - A), \quad \lambda_2 = 0, \quad \lambda_3 = \frac{1}{2}(u + A), \tag{7}$$

$$\lambda_4 = \frac{1}{2}(\bar{u} - \bar{A}), \quad \lambda_5 = 0, \quad \lambda_6 = \frac{1}{2}(\bar{u} + \bar{A}), \quad \lambda_7 = \bar{u}, \tag{8}$$

where $A = \sqrt{u^2 + \frac{4h}{\rho e_p}}$, $\bar{A} = \sqrt{\bar{u}^2 + \frac{4\bar{h}}{\bar{\rho}\bar{e}_{\bar{p}}}}$. The corresponding linearly independent right eigenvectors can be found in [9].

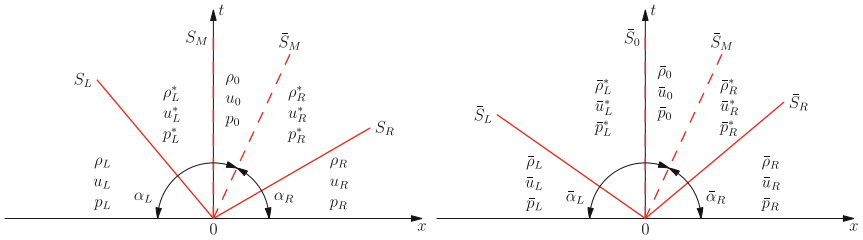


Fig. 1 Intermediate states for the gas (*left*) and solid (*right*) phase

A typical characteristic structure of the solution of the Riemann problem for the pressure system is shown in Fig. 1, where S_L, S_M, S_R and $\bar{S}_L$, S_0, $\bar{S}_M$, $\bar{S}_R$ denote the speeds of the characteristics of the gas and solid phase, respectively. The case illustrated in Fig. 1 corresponds to the right subsonic wave configuration, i.e. $S_L < \bar{S}_M = \bar{u} < S_R$, when $\bar{u} > 0$. In [9] it is shown that $\rho_L^* = \rho_L$, $\rho_R^* = \rho_R$, $p_0 = p_L^*$ for the gas phase and $\bar{\rho}_L^* = \bar{\rho}_L$, $\bar{\rho}_R^* = \bar{\rho}_R$, $\bar{p}_0 = \bar{p}_L^*$, $\bar{u}_0 = \bar{u}_R^*$ for the solid phase. Since $\lambda_1 <= 0 < \lambda_3$ and $\lambda_4 < 0 < \lambda_6$, the Godunov state in the subsonic wave configuration will be completely defined by the sign of $\lambda_7 = \bar{u}$ resulting in significant CPU time savings in the sampling procedure. We also note that due to the above mentioned conditions no entropy fix will be needed for linearized fluxes.

Having computed the intermediate states, we sample the solution of the Riemann problem at $x = 0$ to define the Godunov state $\mathbf{V}_{i+1/2}$ and construct the conservative numerical fluxes for the pressure and advection systems as follows:

$$\mathbf{P}_{i+1/2} = \begin{bmatrix} 0 \\ 0 \\ \bar{\alpha}_{i+1/2}\bar{p}_{i+1/2} \\ \bar{\alpha}_{i+1/2}\bar{u}_{i+1/2}\left(\bar{\rho}_{i+1/2}\bar{e}_{i+1/2} + \bar{p}_{i+1/2}\right) \\ 0 \\ \alpha_{i+1/2}p_{i+1/2} \\ \alpha_{i+1/2}u_{i+1/2}\left(\rho_{i+1/2}e_{i+1/2} + p_{i+1/2}\right) \end{bmatrix}, \tag{9}$$

$$\mathbf{A}_{i+1/2} = \bar{\alpha}_{i+1/2}\bar{u}_{i+1/2}\begin{bmatrix} 0 \\ \bar{\rho} \\ \bar{\rho}\bar{u} \\ \frac{1}{2}\bar{\rho}\bar{u}^2 \\ 0 \\ 0 \\ 0 \end{bmatrix}_k + \alpha_{i+1/2}u_{i+1/2}\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \rho \\ \rho u \\ \frac{1}{2}\rho u^2 \end{bmatrix}_l, \tag{10}$$

where we take

$$k = \begin{cases} i, & \text{if } \bar{u}_{i+1/2} \geq 0, \\ i+1, & \text{if } \bar{u}_{i+1/2} < 0, \end{cases} \quad \text{and} \quad l = \begin{cases} i, & \text{if } u_{i+1/2} \geq 0, \\ i+1, & \text{if } u_{i+1/2} < 0 \end{cases}$$

The complete flux for the conservative term is given by $\mathbf{F}_{i+1/2} = \mathbf{A}_{i+1/2} + \mathbf{P}_{i+1/2}$.

Finally, we use the following approximation of the non-conservative terms at the cell interface $x_{i+1/2}$ proposed in [7]:

$$\mathbf{T}_{i+1/2} = \begin{bmatrix} \bar{u}_{i+1/2}(\bar{\alpha}_{i+1} - \bar{\alpha}_i) \\ 0 \\ -(\bar{p}^*_{R,i+1/2}\bar{\alpha}_{i+1} - \bar{p}^*_{L,i+1/2}\bar{\alpha}_i) \\ -\bar{u}_{i+1/2}(\bar{p}^*_{R,i+1/2}\bar{\alpha}_{i+1} - \bar{p}^*_{L,i+1/2}\bar{\alpha}_i) \\ 0 \\ \bar{p}^*_{R,i+1/2}\bar{\alpha}_{i+1} - \bar{p}^*_{L,i+1/2}\bar{\alpha}_i \\ \bar{u}_{i+1/2}(\bar{p}^*_{R,i+1/2}\bar{\alpha}_{i+1} - \bar{p}^*_{L,i+1/2}\bar{\alpha}_i) \end{bmatrix}. \tag{11}$$

The numerical flux constructed in the previous section can be used directly in the finite volume scheme which in 1D takes the form [7]

$$\mathbf{Q}_i^{n+1} = \mathbf{Q}_i^n - \frac{\Delta t^n}{\Delta x_i}\Big(\mathbf{H}^-_{i+1/2} - \mathbf{H}^+_{i-1/2}\Big), \tag{12}$$

where $\mathbf{H}^-_{i+1/2}$ and $\mathbf{H}^+_{i+1/2}$ are defined by

$$\mathbf{H}^-_{i+1/2} = \begin{cases} \mathbf{F}_{i+1/2} + \mathbf{T}_{i+1/2}, & \text{if } \bar{u}_{i+1/2} \leq 0, \\ \mathbf{F}_{i+1/2}, & \text{if } \bar{u}_{i+1/2} > 0, \end{cases} \tag{13}$$

$$\mathbf{H}^+_{i+1/2} = \begin{cases} \mathbf{F}_{i+1/2}, & \text{if } \bar{u}_{i+1/2} \leq 0, \\ \mathbf{F}_{i+1/2} - \mathbf{T}_{i+1/2}, & \text{if } \bar{u}_{i+1/2} > 0. \end{cases} \tag{14}$$

A first-order path-conservative scheme is given by

$$\mathbf{Q}_i^{n+1} = \mathbf{Q}_i^n - \frac{\Delta t^n}{\Delta x_i}\Big(\mathbf{D}^+_{i-1/2} + \mathbf{D}^-_{i+1/2}\Big), \tag{15}$$

where

$$\mathbf{D}^+_{i-1/2} = \int_0^1 \mathbf{M}\big(\varphi^+_{i-1/2}(s, \mathbf{Q}^n_{i-1/2}, \mathbf{Q}^n_i)\big)\frac{\partial \varphi^+_{i-1/2}}{\partial s}\, ds - \mathbf{A}_{i-1/2},$$

$$\mathbf{D}^-_{i+1/2} = \int_0^1 \mathbf{M}\big(\varphi^-_{i+1/2}(s, \mathbf{Q}^n_i, \mathbf{Q}^n_{i+1/2})\big)\frac{\partial \varphi^-_{i+1/2}}{\partial s}\, ds + \mathbf{A}_{i+1/2},$$

with $\mathbf{M}(\mathbf{Q}) = \frac{\partial \mathbf{P}}{\partial \mathbf{Q}} + \hat{\mathbf{T}}(\mathbf{Q})$ and $\hat{\mathbf{T}} = [\mathbf{T}, \mathbf{0}, \ldots, \mathbf{0}]$, using the canonical paths

$$\varphi^{+}_{i-1/2}(s, \mathbf{Q}^n_{i-1/2}, \mathbf{Q}^n_i) = \mathbf{Q}^n_{i-1/2} + s(\mathbf{Q}^n_i - \mathbf{Q}^n_{i-1/2}),$$
$$\varphi^{-}_{i+1/2}(s, \mathbf{Q}^n_i, \mathbf{Q}^n_{i+1/2}) = \mathbf{Q}^n_i + s(\mathbf{Q}^n_{i+1/2} - \mathbf{Q}^n_i)\ ,$$

with $\mathbf{Q}^n_{i\pm1/2}$ being the Godunov state at the corresponding cell interface.

3 Numerical Results and Efficiency Study

In this section, we test the performance of the TV numerical flux implemented in the first-order finite volume or path-conservative frameworks using various approximate Riemann solvers for the associated P-system. We consider the Riemann problem, introduced in [7], which includes large variations of initial data and non-ideal EOS for the solid phase. The initial data consists of two constant states separated by a discontinuity at $x = 0.5$; the initial data are listed in Table 1 and the EOS parameters are the following: $\gamma = 1.35$, $\bar{\gamma} = 3$ and $\bar{P}_0 = 3400$. Transmissive boundary conditions are imposed at $x = 0$ and $x = 1$. We use the following estimation for the time step: $\Delta t^n = C_{\mathrm{CFL}} \Delta x / S^n_{\max}$, where C_{CFL} is prescribed and the expression for $S^n_{\max}$ is given by $S^n_{\max} = \max_i \{|u^n_i| + a^n_i, |\bar{u}^n_i| + \bar{a}^n_i\},\ i = 1 \ldots N$, and a^i_i and $\bar{a}^i_i$ are the sound speeds of the gas and solid phase, respectively.

The results of the computations using first-order finite volume and path-conservative schemes with various Riemann solvers for the P-system are shown in Fig. 2. For the efficiency study of the TV flux splitting method with various Riemann solvers for the P-system we performed computations on a sequence of meshes using first order flux splitting schemes as well as the finite-volume scheme with HLL and HLLC Riemann solvers for the full Baer-Nunziato system. Figure 3 is an efficiency plot, namely an Error versus CPU time plot. The curve "HLLC full" corresponds to the finite-volume scheme with the HLLC-type Riemann solver for the Baer-Nunziato equations from [8], the curve "HLL full" is the numerical solution obtained by the nonconservative HLL-PVM method applied to the complete unsplit Baer-Nunziato system according to [3], while other curves correspond to versions of the TV flux splitting scheme depending on the Riemann solver used for the pressure system: "HLL-TV" denotes the HLL-PVM Riemann solver of [3], "HLLEM-TV" denotes the HLLEM solver [5], "NumRoe-TV" corresponds to the numerical Roe approach of [2] and, finally, "LinRS-TV" illustrates the results from the linearized

Table 1 Initial data

$\bar{\alpha}_L$	$\bar{\rho}_L$	$\bar{u}_L$	$\bar{p}_L$	$\bar{\alpha}_R$	$\bar{\rho}_R$	$\bar{u}_R$	$\bar{p}_R$	α_L	ρ_L	u_L	p_L	α_R	ρ_R	u_R	p_R
0.2	1900.0	0.0	10.0	0.9	1950.0	0.0	1000.0	0.8	2.0	0.0	3.0	0.1	1.0	0.0	1.0

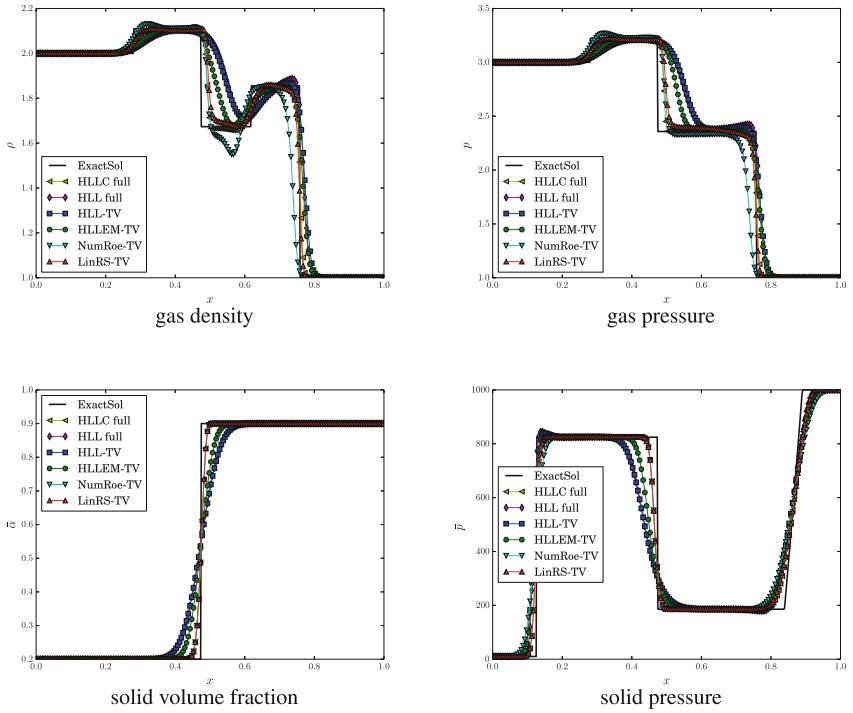
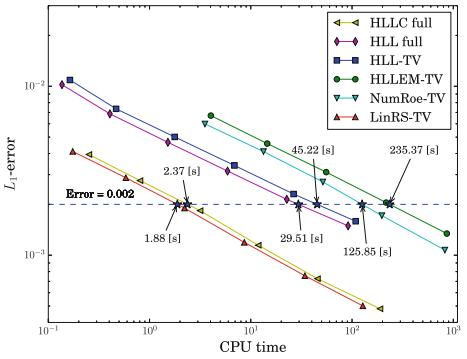


Fig. 2 Numerical (symbol) and exact solution (line) at $t = 0.15$

Fig. 3 Efficiency plot



Riemann solver of [9] applied to the pressure system. In practice, we have observed that the maximum CFL coefficient which guarantees stable results for a range of test problems depends on the Riemann solver used for the P-system [9]. Therefore, for the linearized Riemann solver and full HLL solver we set $C_{\mathrm{CFL}} = 0.9$, for HLL-PVM and HLLEM Riemann solvers $C_{\mathrm{CFL}} = 0.8$ and for the numerical Roe approach $C_{\mathrm{CFL}} = 0.6$. Each curve displays six points of the form (Error, CPU time), corresponding to six meshes. Errors were computed in the L_1-norm for the variable $\bar{\rho}$. In Fig. 3 we choose $Error = 0.002$. We see that the TV flux splitting method with the linearized Riemann solver for the P-system is the most efficient; it takes only 1.88 s of CPU time to attain the chosen error. This solver is slightly more efficient than the HLLC scheme of [8] which takes 2.37 s to reach the indicated error; moreover, the implementation of the HLLC Riemann solver for the Baer-Nunziato equations and the solution sampling for the numerical flux computation are more complicated than in the linearized Riemann solver for the P-system. The next most efficient method is the HLL scheme applied directly to the full Baer-Nunziato system (no flux vector splitting); it is however 15.7 times more expensive than the present TV splitting with the linearized Riemann solver for the P-system. Very close to the HLL-full is the TV splitting with the HLL flux for the P-system, followed by the TV splitting with numerical Roe flux for the P-system. The most inefficient scheme turns out to be the TV splitting with HLLEM scheme. However, it should be noted that these last two schemes are rather general and can easily be implemented to solve more general hyperbolic systems than the one considered in this paper. See also [4] for other comparisons of different schemes for the Baer-Nunziato model.

The attraction of the TV splitting is the simplicity with which one can construct the numerical flux for the full scheme. Such simplicity is two fold, first the work is centred on a reduced system, namely pressure system; then for such system one can devise very simple numerical schemes for the associated pressure numerical flux. The end result is a very simple and efficient scheme for the full system, without sacrificing robustness and accuracy. As soon as one devises complicated and expensive methods for the pressure system, the attraction of the TV flux vector splitting approach is lost, as demonstrated by our efficiency study.

References

1. Baer, M.R., Nunziato, J.W.: A two-phase mixture theory for the deflagration-to-detonation transition (DDT) in reactive granular materials. J. Multiphase Flow **12**, 861–889 (1986)
2. Castro, C.E., Toro, E.F.: Roe-type Riemann solvers for general hyperbolic systems. Int. J. Numer. Meth. Fluids **75**, 467–486 (2014)
3. Castro Díaz, M.J., Fernández-Nieto, E.D.: A class of computationally fast first order finite volume solvers: PVM methods. SIAM J. Sci. Comput. **34**, A2173–A2196 (2012)
4. Coquel, F., Hérard, J.M., Saleh, K.: A positive and entropy-satisfying finite volume scheme for the BaerNunziato model. J. Comput. Phys. **330**, 401–435 (2017)
5. Dumbser, M., Balsara, D.S.: A new efficient formulation of the HLLEM Riemann solver for general conservative and non-conservative hyperbolic systems. J. Comput. Phys. **304**, 275–319 (2016)

6. Embid, P., Baer, M.: Mathematical analysis of a two-phase continuum mixture theory. Continuum Mech. Thermodyn. **4**, 279–312 (1992)
7. Schwendeman, D.W., Wahle, C.W., Kapila, A.K.: The Riemann problem and a high-resolution Godunov method for a model of compressible two-phase flow. J. Comput. Phys. **212**, 490–526 (2006)
8. Tokareva, S.A., Toro, E.F.: HLLC-type Riemann solver for the Baer-Nunziato equations of compressible two-phase flow. J. Comput. Phys. **229**, 3573–3604 (2010)
9. Tokareva, S.A., Toro, E.F.: A flux splitting method for the baer-nunziato equations of compressible two-phase flow. J. Comput. Phys. **323**, 45–74 (2016)
10. Toro, E.F., Castro, C.E., Lee, B.J.: A novel numerical flux for the 3D Euler equations with general equation of state. J. Comput. Phys. **303**, 80–94 (2015)
11. Toro, E.F., Vázquez-Cendón, M.E.: Flux splitting schemes for the Euler equations. Comput. Fluids **70**, 1–12 (2012)

GPU Accelerated Finite Volume Methods for Three-Dimensional Shallow Water Flows

Mohamed Boubekeur, Fayssal Benkhaldoun and Mohammed Seaid

Abstract This paper presents a newly developed finite volume method to simulate three-dimensional free-surface flows on GPU-equipped supercomputer. The model consists of a class of multi-layered shallow water equations with exchange terms between layers and the finite volume method uses a predictor-corrector procedure. These techniques are devised to be computationally efficient and well-suitable for hardwares of multi-core CPUs with many core GPU accelerators. An extensible multi-threading programming API is used as a common kernel language that allows runtime selection of different computing devices (GPU and CPU, CUDA and OpenMP). Numerical results are presented for a circular dam-break problem.

Keywords Shallow water flows · Finite volume methods · GPU computing

1 Introduction

Recently a set of multi-layer shallow water equations has been proposed in [1] to model vertical effects and mass exchanges in one-dimensional free-surface flows. Authors in [3] proposed a simple finite volume method to solve this class of multi-layer shallow water models. The method belongs to predictor-corrector solvers and avoids the solution of Riemann problems to reconstruct numerical fluxes. In the predictor stage, the multi-layered shallow water equations are rewritten in a non-conservative form and the intermediate solutions are calculated using the modified

M. Boubekeur (✉) · F. Benkhaldoun
LAGA, Université Paris 13, Sorbonne Paris Cité, 99 Av J.B. Clement,
93430 Villetaneuse, France
e-mail: boubekeur@math.univ-paris13.fr

F. Benkhaldoun
e-mail: fayssal@math.univ-paris13.fr

M. Seaid
School of Engineering and Computing Sciences, University of Durham, Durham, UK
e-mail: m.seaid@durham.ac.uk

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_15

method of characteristics. In the corrector stage, the numerical fluxes are reconstructed from the intermediate solutions in the first stage and used in the conservative form of the multi-layered shallow water equations. The proposed method is simple to implement, and easy to parallelize, and allows to be faster than conventional finite volume methods, see for instance [2].

A cost effective way to obtain higher performance and reduce time of simulations consists in using Graphics Processor Units (GPU). The popularity of using these devices is growing to accelerate computationally intensive tasks, see for example [4] for a GPU implementation of finite volume methods for single-layered shallow water flows. GPU card presents a massively parallel architecture which includes hundreds of processing units optimized for performing floating point operations and multi-threaded execution. These architectures allow to obtain higher performance than standard CPU at a very affordable price. The objective of the current work is twofold: on one hand we extend our finite volume method to the two-dimensional multi-layered shallow water system and on the other hand we implement these techniques on GPU to speed up the computational process.

2 Multi-layered Shallow Water Equations

In the current work we consider the two-dimensional version of the multi-layered shallow water equations proposed in [1]. The system is reformulated in a conservative form as

$$\frac{\partial \mathbf{W}}{\partial t} + \frac{\partial \mathbf{F}(\mathbf{W})}{\partial x} + \frac{\partial \mathbf{G}(\mathbf{W})}{\partial y} = \mathbf{Q}(\mathbf{W}), \tag{1}$$

where $\mathbf{W}$ is the vector of conserved variable, $\mathbf{F}$ and $\mathbf{G}$ the vectors of flux functions, and $\mathbf{Q}$ are the vector of source terms

$$\mathbf{W} = \begin{pmatrix} H \\ Hu_1 \\ Hv_1 \\ Hu_2 \\ Hv_2 \\ \vdots \\ Hu_{M-1} \\ Hv_{M-1} \\ Hu_M \\ Hv_M \end{pmatrix}, \quad \mathbf{F}(\mathbf{W}) = \begin{pmatrix} \sum_{\alpha=1}^{n} l_\alpha Hu_\alpha \\ l_\alpha Hu_1^2 + \frac{1}{2} g l_\alpha H^2 \\ l_\alpha Hu_1 v_1 \\ l_\alpha Hu_2^2 + \frac{1}{2} g l_\alpha H^2 \\ l_\alpha Hu_2 v_2 \\ \vdots \\ l_\alpha Hu_{M-1}^2 + \frac{1}{2} g l_\alpha H^2 \\ l_\alpha Hu_{M-1} v_{M-1} \\ l_\alpha Hu_M^2 + \frac{1}{2} g l_\alpha H^2 \\ l_\alpha Hu_M v_M \end{pmatrix}, \quad \mathbf{G}(\mathbf{W}) = \begin{pmatrix} \sum_{\alpha=1}^{n} l_\alpha Hv_\alpha \\ l_\alpha Hu_1 v_1 \\ l_\alpha Hv_1^2 + \frac{1}{2} g l_\alpha H^2 \\ l_\alpha Hu_2 v_2 \\ l_\alpha Hv_2^2 + \frac{1}{2} g l_\alpha H^2 \\ \vdots \\ l_\alpha Hu_{M-1} v_{M-1} \\ l_\alpha Hv_{M-1}^2 + \frac{1}{2} g l_\alpha H^2 \\ l_\alpha Hu_M v_M \\ l_\alpha Hv_M^2 + \frac{1}{2} g l_\alpha H^2 \end{pmatrix},$$

$$\mathbf{Q}(\mathbf{W}) = \begin{pmatrix} 0 \\ \frac{1}{l_1}\left(u_{\frac{3}{2}}\xi^u_{\frac{3}{2}} - \frac{\zeta_b}{\rho} + 2\nu\frac{u_2-u_1}{(l_2+l_1)H}\right) \\ \frac{1}{l_1}\left(v_{\frac{3}{2}}\xi^v_{\frac{3}{2}} - \frac{\eta_b}{\rho} + 2\nu\frac{v_2-v_1}{(l_2+l_1)H}\right) \\ \frac{1}{l_2}\left(u_{\frac{5}{2}}\xi^u_{\frac{5}{2}} - u_{\frac{3}{2}}\xi^u_{\frac{3}{2}} + 2\nu\frac{u_3-u_2}{(l_3+l_2)H} - 2\nu\frac{u_2-u_1}{(l_2+l_1)H}\right) \\ \frac{1}{l_2}\left(v_{\frac{5}{2}}\xi^v_{\frac{5}{2}} - v_{\frac{3}{2}}\xi^v_{\frac{3}{2}} + 2\nu\frac{v_3-v_2}{(l_3+l_2)H} - 2\nu\frac{v_2-v_1}{(l_2+l_1)H}\right) \\ \vdots \\ \frac{1}{l_{M-1}}\left(u_{M-\frac{1}{2}}\xi^u_{M-\frac{1}{2}} - u_{M-\frac{3}{2}}\xi^u_{M-\frac{3}{2}} + 2\nu\frac{u_M-u_{M-1}}{(l_M+l_{M-1})H} - 2\nu\frac{u_{M-1}-u_{M-2}}{(l_{M-1}+l_{M-2})H}\right) \\ \frac{1}{l_{M-1}}\left(v_{M-\frac{1}{2}}\xi^v_{M-\frac{1}{2}} - v_{M-\frac{3}{2}}\xi^v_{M-\frac{3}{2}} + 2\nu\frac{v_M-v_{M-1}}{(l_M+l_{M-1})H} - 2\nu\frac{v_{M-1}-v_{M-2}}{(l_{M-1}+l_{M-2})H}\right) \\ \frac{1}{l_M}\left(-u_{M-\frac{1}{2}}\xi^u_{M-\frac{1}{2}} + \frac{\zeta_w}{\rho} - 2\nu\frac{u_M-u_{M-1}}{(l_M+l_{M-1})H}\right) \\ \frac{1}{l_M}\left(-v_{M-\frac{1}{2}}\xi^v_{M-\frac{1}{2}} + \frac{\eta_w}{\rho} - 2\nu\frac{v_M-v_{M-1}}{(l_M+l_{M-1})H}\right) \end{pmatrix},$$

where (u_α, v_α) is the local water velocity for the αth layer, ν the eddy viscosity, g the gravitational acceleration, ρ the water density, ζ_b and η_b are the bed shear stress, and ζ_w and η_w are the shear of the blowing wind. Here, H denotes the water height of the whole flow system and l_α denotes the relative size of the αth layer with

$$l_\alpha > 0, \qquad \sum_{\alpha=1}^{M} l_\alpha = 1.$$

The mass exchange term $\xi^u_{\alpha+\frac{1}{2}}$ is defined as

$$\xi^u_{\alpha+\frac{1}{2}} = \sum_{\beta=1}^{\alpha}\left(\frac{\partial(l_\beta H u_\beta)}{\partial x} - l_\beta\sum_{\gamma=1}^{M}\frac{\partial(l_\gamma H u_\gamma)}{\partial x}\right),$$

and the interface velocity is computed by a simple upwinding following the sign of the mass exchange term as

$$u_{\alpha+\frac{1}{2}} = \begin{cases} u_\alpha, & \text{if } \xi^u_{\alpha+\frac{1}{2}} \geqslant 0, \\ u_{\alpha+1}, & \text{if } \xi^u_{\alpha+\frac{1}{2}} < 0. \end{cases}$$

Similarly, the mass exchange term $\xi^v_{\alpha+\frac{1}{2}}$ is defined as

$$\xi^v_{\alpha+\frac{1}{2}} = \sum_{\beta=1}^{\alpha}\left(\frac{\partial(l_\beta H v_\beta)}{\partial y} - l_\beta\sum_{\gamma=1}^{M}\frac{\partial(l_\gamma H v_\gamma)}{\partial y}\right),$$

and the interface velocity

$$v_{\alpha+\frac{1}{2}} = \begin{cases} v_\alpha, & \text{if } \xi^v_{\alpha+\frac{1}{2}} \geqslant 0, \\ v_{\alpha+1}, & \text{if } \xi^v_{\alpha+\frac{1}{2}} < 0. \end{cases}$$

It should also be pointed out that the layers defined in the model do not refer to physical interfaces between non-miscible fluids but to a meshless discretization of the flow domain. Hence, the possibility of water exchange between the layers is accounted for in the model. The great interest of this strategy is to preserve an accurate description of the velocity profile but to deal with a two-dimensional fluid model and thus to avoid the drawback of remeshing a three-dimensional moving domain for which the free-surface may present very sharp profiles such as dam-break problems and hydraulic jumps.

3 GPU Accelerated Finite Volume Characteristics Solver

For the space discretization of the system (1) we cover the spatial domain with cells $C_{ij} = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \times [y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}]$ centered at (x_i, y_j) with uniform sizes Δx and Δy for simplicity in the presentation only. We also use the notations

$$\mathbf{W}_{i\pm\frac{1}{2},j}(t) = \mathbf{W}(t, x_{i\pm\frac{1}{2}}, y_j), \qquad \mathbf{W}_{i,j\pm\frac{1}{2}}(t) = \mathbf{W}(t, x_i, y_{j\pm\frac{1}{2}}),$$

$$\text{and} \qquad \mathbf{W}_{i,j}(t) = \frac{1}{\Delta x}\frac{1}{\Delta y}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}\int_{y_{j-\frac{1}{2}}}^{y_{i+\frac{1}{2}}} \mathbf{W}(t, x, y)dydx,$$

to denote the point-values and the approximate cell-average of the variable $\mathbf{W}$ at the gridpoint $(t, x_{i\pm\frac{1}{2}}, y_j)$, $(t, x_i, y_{j\pm\frac{1}{2}})$, and (t, x_i, y_j), respectively. Integrating the Eq. (1) with respect to space over the control volume $C_{i,j}$, we obtain the following semi-discrete system

$$\frac{d\mathbf{W}_{i,j}}{dt} + \frac{\mathbf{F}_{i+1/2,j} - \mathbf{F}_{i-1/2,j}}{\Delta x} + \frac{\mathbf{G}_{i,j+1/2} - \mathbf{G}_{i,j-1/2}}{\Delta y} = \mathbf{Q}_{i,j}, \tag{2}$$

where $\mathbf{F}_{i\pm1/2,j} = \mathbf{F}(\mathbf{W}_{i\pm1/2,j})$ and $\mathbf{G}_{i,j\pm1/2} = \mathbf{G}(\mathbf{W}_{i,j\pm1/2})$ are the numerical fluxes at the cell interfaces $x = x_{i\pm1/2}$ and $y = y_{i\pm1/2}$, respectively. In (2), $\mathbf{Q}_{i,j}$ is a consistent discretization of the source term $\mathbf{Q}$ in (1). The spatial discretization of Eq. (2) is resumed when a numerical construction of the fluxes $\mathbf{F}_{i\pm1/2,j}$ and $\mathbf{G}_{i,j\pm1/2}$ is chosen. In general, this construction requires a solution of Riemann problems at the interfaces $x_{i\pm1/2}$ and $y_{i\pm1/2}$. From a computational viewpoint, this procedure is very demanding and may restrict the application of the method to shallow water equations for which Riemann solutions are not available.

In the current study we consider the finite volume characteristics method proposed in [3] for the numerical solution of one-dimensional counterpart of the system (1). The method computes the intermediate solutions $\mathbf{W}_{i\pm1/2,\,j}$ and $\mathbf{W}_{i,\,j\pm1/2}$ by reformulating the system (1) in a non-conservative form and apply the modified method of characteristics. This step in the method is referred to by predictor stage whereas the solution is recovered from the corrector stage (2). Details on the implementation of finite volume characteristics method can be found in [3] and will not be repeated here. Time integration of the semi-discrete system can be achieved by using any explicit scheme such as Runge-Kutta methods. For instance, a first-order explicit Euler scheme applied to (2) yields

$$\mathbf{W}_{i,j}^{n+1} = \mathbf{W}_{i,j}^{n} - \frac{\Delta t}{\Delta x}\left(\mathbf{F}_{i+1/2,\,j}^{n} - \mathbf{F}_{i-1/2,\,j}^{n}\right) + \frac{\Delta t}{\Delta y}\left(\mathbf{G}_{i,\,j+1/2}^{n} - \mathbf{G}_{i,\,j-1/2}^{n}\right) + \Delta t\mathbf{Q}_{i,j}^{n}, \tag{3}$$

where the time interval is divided into N subintervals $[t_n, t_{n+1}]$ with length $\Delta t = t_{n+1} - t_n$ for $n = 0, 1, \ldots, N$ and W^n denotes the value of a generic function W at time t_n. Because the time integration scheme is explicit, the time step Δt has to satisfy a stability condition of the form

$$\Delta t = C\frac{\min\left(\Delta x, \Delta y\right)}{\max\left(\lambda, \mu\right)},$$

where C is the Courant number to be chosen less than unity, λ and μ are the maximum of eigenvalues associated to the single-layer model defined as

$$\lambda = \max_{\alpha=1,\ldots,M}\left(\left|u_\alpha + \sqrt{gH}\right|, \left|u_\alpha\right|, \left|u_\alpha - \sqrt{gH}\right|\right),$$
$$\mu = \max_{\alpha=1,\ldots,M}\left(\left|v_\alpha + \sqrt{gH}\right|, \left|v_\alpha\right|, \left|v_\alpha - \sqrt{gH}\right|\right).$$

It well established that in the CUDA framework, both the CPU and the GPU maintain their own memory. However, it is possible to copy data from CPU memory to GPU memory and vice versa without any computational difficulties. The GPU is constituted by a set of multiprocessors for which each of these multiprocessors has a number of processors and each processor executes the same instructions but it operates on different data. Using GPU, a kernel is executed by many threads which are organized forming a grid of thread blocks that run logically in parallel. All blocks and threads have spatial indices, so that the spatial position of each thread could be identified in the program and each thread block runs in a single multiprocessor. Here, four CUDA kernels were implemented: one to compute the mass exchange, one to compute the characteristics in the predictor stage, one to compute the fluxes, and the last one to update the solution in the corrector stage. In our implementation, this architecture is used to implement some kernel functions to compute the solution $\mathbf{W}^{n+1}$ at each time step. The combination of the finite volume method for space discretization and the explicit Euler scheme for time integration offers a straightforward

parallel execution by considering that each thread represents a control volume. To minimize the execution time, the data exchange between the CPU and GPU memories is also limited. Here, the data exchange is used only for initialization and to export results. It should be stressed that the memory exchange between GPU and CPU is negligible, it is used only to initialize and to extract the final solution. The OpenGL library uses GPU to GPU transfer and it is nearly 100 GB/s. The CUDA performance of the code for the mesh with 500×500 grid-points and 20 layers is $718.65\, GFLOPS$. For each time step, the method is carried out using the following two steps:

- Predictor step: Three kernel functions are implemented, the first one computes the mass exchanges terms, the second computes the characteristics, and the last one computes the numerical fluxes $\mathbf{F}_{i\pm1/2,j}$ and $\mathbf{G}_{i,j\pm1/2}$. Using these kernels, we compute the intermediate solutions $\mathbf{W}_{i\pm1/2,j}$ and $\mathbf{W}_{i,j\pm1/2}$
- Corrector step: A kernel is implemented to update the solution $\mathbf{W}_{i,j}^{n+1}$ using the time stepping (3)
- Post-processing: We use $\mathbf{W}_{i,j}^{n+1}$ to draw the water height using the OpenGL library.

Note that the OpenGL library is integrated in the CUDA platform which allows displaying the simulated results in real-time without communication between CPU and GPU memories. All the simulations use a double numerical precision.

4 Numerical Results

We solve the test example of dam-break problem in three-dimensional free-surface flows. The single-layer version of this example has been investigated in [2] to study cyclone/anticyclone asymmetry in nonlinear geostrophic adjustment. Hence, we adapt the same parameters as those used in [2] and the multi-layer system (1) is solved in the spatial domain $[-10, 10] \times [-10, 10]$ subject to Neumann boundary conditions. The initial conditions are

$$H(0, x, y) = 1 + \frac{1}{4}\left(1 - \tanh\left(\frac{\sqrt{ax^2 + by^2} - 1}{c}\right)\right), \quad u_\alpha(0, x, y) = v_\alpha(0, x, y) = 0,$$

where $a = 52$, $b = 25$ and $c = 0.1$. In our simulations $g = 1$, the eddy viscosity $\nu = 0.01$ and the courant number $C = 0.75$. In all our simulations we used for the GPU card a NVIDA Quadro K5100M with 8 multiprocessors and each processor is constituted by 192 processors. The times of execution are also compared to those obtained using CPU simulation which are performed using an OpenMP implementation. The CPU is an Intel i7 with 8 cores cadenced to 3 GHz.

In Fig. 1 we present the water free-surface and velocity field obtained at times $t = 1$, 3 and 5 s. As can be seen a bore has formed and the water drains from the deepest region as a rarefaction wave progresses outwards. The flow in that region becomes supercritical. It is also clear that the numerical solution preserves rotational

symmetry in a perfect way and the problem is solved correctly by our finite volume method.

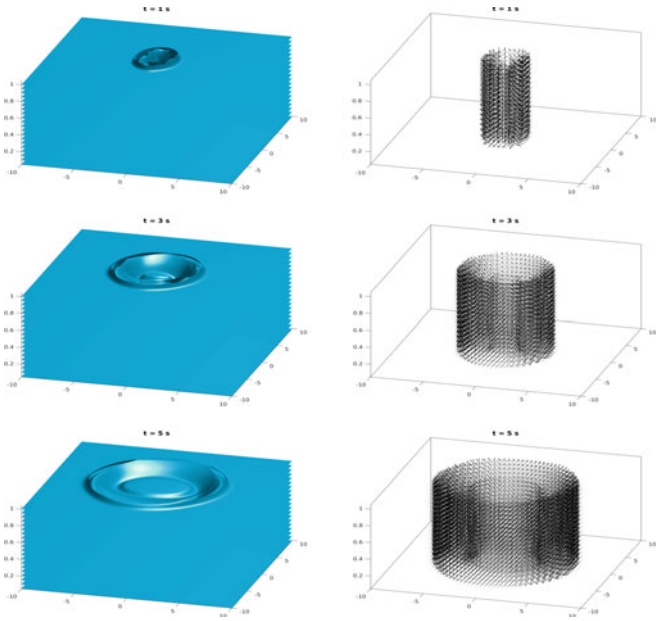


Fig. 1 Water heights (*left*) and velocity fields (*right*) obtained for dam-break problem with 20 layers at three different instants using the GPU simulation

Table 1 Execution times in seconds obtained using 5, 10 and 20 layers on different meshes with 50 × 50, 100 × 100, 200 × 200 and 500 × 500 gridpoints

	5 layers			
	50 × 50	100 × 100	200 × 200	500 × 500
CPU 1 core	0.92	7.62	65.45	1046.14
CPU 8 cores	0.54	3.65	24.76	355.31
GPU	0.47	1.54	6.98	77.26
	10 layers			
	50 × 50	100 × 100	200 × 200	500 × 500
CPU 1 core	1.81	15.68	134.40	2202.90
CPU 8 cores	0.96	6.70	45.25	738.15
GPU	0.77	2.82	13.29	157.0
	20 layers			
	50 × 50	100 × 100	200 × 200	500 × 500
CPU 1 core	3.43	31.94	273.74	4349.97
CPU 8 cores	1.87	12.90	96.85	1496.98
GPU	1.42	5.71	28.71	360.50

Next we compare the results obtained for this test problem on the GPU to those obtained using the CPU for 1 core and 8 cores without the post-processing step. Table 1 summarizes the execution times for both implementations on CPU and GPU platforms using different numbers of layers and meshes. These execution times do not include the OpenGL. Here we solve the multi-layer system for 5, 10 and 20 layers using meshes with 50×50, 100×100, 200×200 and 500×500 gridpoints, and the results are reported at the simulation time $t = 10\ s$. It is evident from the results in Table 1 that a linear increase in the execution times results from any increase in the number of layers in the system for all implementations and meshes. However, in all considered cases the GPU simulation is the fasted. Computing the mass exchange terms is not totally parallel because of the dependency between the layers.

5 Conclusion

A GPU accelerated finite volume method is presented for solving three-dimensional free-surface flows using the multi-layered shallow water equations. The method combines the modified method of characteristics and the finite volume method in a predictor and corrector procedures. The method is simple, accurate and avoids resolution of Riemann problems in its reconstruction of numerical fluxes. The method is entirely ported to such a system using CUDA-driver to make fully possible use of the GPU acceleration capability on large-scale parallel computations. The high-resolution simulations with large number of layers and gridpoints have been demonstrated for a circular dam-break problem. The simulation time is compared to CPU implementation for 1 core and 8 cores by varying the number of layers and mesh size. We noted that the GPU simulations are the fastest for all cases and for the same precision.

Acknowledgements The authors would like to thank Prof. S. Sari for valuable discussions about the implementation of finite volume characteristics method.

References

1. Audusse, E., Benkhaldoun, F., Sari, S., Seaid, M., Tassi, P.: A fast finite volume solver for multi-layered shallow water flows with mass exchange. J. Comput. Phys. **272**, 23–45 (2014)
2. Benkhaldoun, F., Sari, S., Seaid, M.: Projection finite volume method for shallow water flows. Math. Comput. Simul. **118**, 87–101 (2015)
3. Benkhaldoun, F., Seaid, M.: A simple finite volume method for the shallow water equations. J. Comput. Appl. Math. **234**, 58–72 (2010)
4. Castro, M., Ortega, S., Asunción, M., Mantas, J., Gallardo, J.: GPU computing for shallow water flow simulation based on finite volume schemes. C R Mécanique **339**, 165–184 (2011)

Projective Integration for Nonlinear BGK Kinetic Equations

Ward Melis, Thomas Rey and Giovanni Samaey

Abstract We present a high-order, fully explicit, asymptotic-preserving projective integration scheme for the nonlinear BGK equation. The method first takes a few small (inner) steps with a simple, explicit method (such as direct forward Euler) to damp out the stiff components of the solution. Then, the time derivative is estimated and used in an (outer) Runge–Kutta method of arbitrary order. Based on the spectrum of the linearized BGK operator, we deduce that, with an appropriate choice of inner step size, the time step restriction on the outer time step as well as the number of inner time steps is independent of the stiffness of the BGK source term. We illustrate the method with numerical results in one and two spatial dimensions.

Keywords Projective integration · BGK · Asymptotic-preserving · WENO

MSC (2010): 82B40 · 76P05 · 65M08 · 65L06

1 Introduction

The Boltzmann equation constitutes the cornerstone of kinetic theory. It describes the evolution of the one-particle mass distribution function $f^\varepsilon(\mathbf{x}, \mathbf{v}, t) \in \mathbb{R}^+$ as:

$$\partial_t f^\varepsilon + \mathbf{v} \cdot \nabla_{\mathbf{x}} f^\varepsilon = \frac{1}{\varepsilon} \mathcal{Q}\left(f^\varepsilon\right)(\mathbf{v}), \tag{1}$$

W. Melis · G. Samaey
NUMA (Numerical Analysis and Applied Mathematics) Dept. Computer Science,
KU Leuven, Celestijnenlaan 200A, 3001 Leuven, Belgium
e-mail: ward.melis@cs.kuleuven.be

G. Samaey
e-mail: giovanni.samaey@cs.kuleuven.be

T. Rey (✉)
Laboratoire Paul Painlevé, Université de Lille, Cité Scientifique,
59655 Villeneuve D'Ascq, France
e-mail: thomas.rey@math.univ-lille1.fr

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_16

where $t \geq 0$ represents time, and $(\mathbf{x}, \mathbf{v}) \subset \mathbb{R}^{D_x \times D_v}$ are the D_x-dimensional particle positions and D_v-dimensional particle velocities. In Eq. (1), the dimensionless constant $\varepsilon > 0$ determines the regime of the gas flow, for which we roughly identify the hydrodynamic regime ($\varepsilon \leq 10^{-4}$), the transitional regime ($\varepsilon \in [10^{-4}, 10^{-1}]$), and the kinetic regime ($\varepsilon \geq 10^{-1}$). Furthermore, the left hand side of (1) corresponds to a linear transport operator that comprises the convection of particles in space, whereas the right hand side contains the Boltzmann collision operator that entails velocity changes due to particle collisions. However, due to its high-dimensional and complicated structure, the Boltzmann collision operator is often replaced by simpler collision models that capture most essential features of the former. The most well-known such model is the BGK model [1], which models collisions as a linear relaxation towards thermodynamic equilibrium, and is given by:

$$\partial_t f^\varepsilon + \mathbf{v} \cdot \nabla_{\mathbf{x}} f^\varepsilon = \frac{1}{\varepsilon}(\mathscr{M}_{\mathbf{v}}(f^\varepsilon) - f^\varepsilon), \tag{2}$$

in which $\mathscr{M}_{\mathbf{v}}(f^\varepsilon)$ denotes the local Maxwellian distribution, which, for a D_v-dimensional velocity space, is given by:

$$\mathscr{M}_{\mathbf{v}}(f^\varepsilon) = \frac{\rho}{(2\pi T)^{D_v/2}} \exp\left(-\frac{|\mathbf{v} - \bar{\mathbf{v}}|^2}{2T}\right) := \mathscr{M}_{\mathbf{v}}^{\rho, \bar{\mathbf{v}}, T}. \tag{3}$$

The Maxwellian distribution contains the velocity moments of the distribution function f^ε, which are calculated as:

$$\rho = \int_{\mathbb{R}^{D_v}} f^\varepsilon d\mathbf{v}, \qquad \bar{\mathbf{v}}^d = \frac{1}{\rho} \int_{\mathbb{R}^{D_v}} v^d f^\varepsilon d\mathbf{v}, \qquad T = \frac{1}{D_v \rho} \int_{\mathbb{R}^{D_v}} |\mathbf{v} - \bar{\mathbf{v}}|^2 f^\varepsilon d\mathbf{v}, \tag{4}$$

where $\rho \in \mathbb{R}^+$, $\bar{\mathbf{v}} = \left(\bar{\mathbf{v}}^d\right)_{d=1}^{D_v} \in \mathbb{R}^{D_v}$ and $T \in \mathbb{R}^+$ are the density, macroscopic velocity and temperature, respectively, which all depend on space $\mathbf{x}$ and time t. Then, in the limit $\varepsilon \to 0$, the solution to Eq. (2) converges towards $\mathscr{M}_{\mathbf{v}}^{\rho, \bar{\mathbf{v}}, T}$, whose moments in (4) are solution to the compressible Euler system:

$$\begin{cases} \partial_t \rho + \operatorname{div}_{\mathbf{x}}(\rho\, \bar{\mathbf{v}}) = 0, \\ \partial_t (\rho\, \bar{\mathbf{v}}) + \operatorname{div}_{\mathbf{x}} (\rho\, \bar{\mathbf{v}} \otimes \bar{\mathbf{v}} + \rho\, T\, \mathbf{I}) = \mathbf{0}, \\ \partial_t E + \operatorname{div}_{\mathbf{x}} (\bar{\mathbf{v}} (E + \rho\, T)) = 0, \end{cases} \tag{5}$$

in which E is the second moment of f^ε, namely its total energy.

In this paper, we construct a fully explicit, asymptotic-preserving, arbitrary order time integration method for the stiff Eq. (2). The asymptotic-preserving property [6] implies that, in the limit when ε tends to zero, an ε-independent time step constraint, of the form $\Delta t = O(\Delta x)$, can be used, in agreement with the classical hyperbolic CFL constraint for the limiting fluid Eq. (5). To achieve this, we will use a projective integration method, which was introduced in [5] and first applied to kinetic equations

in [7]. For a comprehensive review of numerical schemes for collisional kinetic equations such as Eq. (1), we refer to [4]. Although it is known that an implicit treatment of (2) can be implemented explicitly [4], the order in time is usually restricted to 2. Therefore, the main advantage of the proposed method is its arbitrary order in time.

The remainder of this paper is structured as follows. We describe the projective integration method in more detail in Sect. 2, after which we discuss (in Sect. 3) the spectral properties of the linearized BGK operator, which are needed to ensure stability of the method. Some numerical experiments are done in Sect. 4.

2 Projective Integration

Projective integration [5, 7] combines a few small time steps with a naive (*inner*) timestepping method (here, a direct forward Euler discretization) with a much larger (*projective, outer*) time step. The idea is sketched in Fig. 1.

Inner integrators. We discretize Eq. (2) on a uniform, constant in time, periodic spatial mesh with spacing Δx, consisting of I mesh points $x_i = i\Delta x$, $1 \leq i \leq I$, with $I\Delta x = 1$, and a uniform time mesh with time step δt and discrete time instants $t^k = k\delta t$. Furthermore, we discretize velocity space by choosing J discrete components denoted by $\mathbf{v}_j$. The numerical solution on this mesh is denoted by $f^k_{i,j}$, where we have dropped the superscript ε on discretized quantities. We then obtain a semidiscrete system of ODEs of the form:

$$\dot{\mathbf{f}} = \mathrm{D}_t(\mathbf{f}), \qquad \mathrm{D}_t(\mathbf{f}) = -\mathrm{D}_{\boldsymbol{x},\boldsymbol{v}}(\mathbf{f}) + \frac{1}{\varepsilon}\left(\mathscr{M}_{\boldsymbol{v}}\left(\mathbf{f}\right) - \mathbf{f}\right), \tag{6}$$

where $\mathrm{D}_{\boldsymbol{x},\boldsymbol{v}}(\cdot)$ represents a suitable discretization of the convective derivative $\mathbf{v} \cdot \nabla_{\mathbf{x}}$ (for instance, using upwind differences), and $\mathbf{f}$ is a vector of size $I \cdot J$.

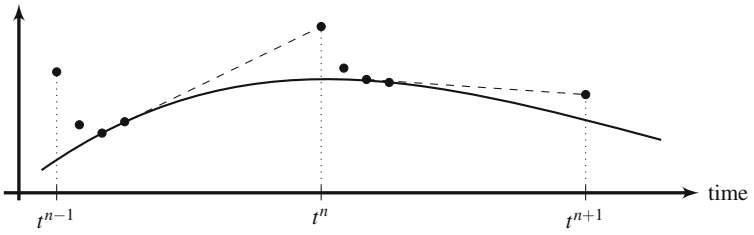


Fig. 1 Sketch of projective integration. At each time, an explicit method is applied over a number of small time steps (*black dots*) so as to stably integrate the fast modes. As soon as these modes are sufficiently damped the solution is extrapolated using a much larger time step (*dashed lines*)

As inner integrator, we choose the (explicit) forward Euler method with time step δt, for which we will, later on, use the shorthand notation:

$$\mathbf{f}^{k+1} = S_{\delta t}(\mathbf{f}^k) = \mathbf{f}^k + \delta t \mathrm{D}_t(\mathbf{f}^k), \qquad k = 0, 1, \ldots. \tag{7}$$

Outer integrators. In system (6), the small parameter ε leads to the classical time step restriction of the form $\delta t = O(\varepsilon)$ for the inner integrator. However, as ε goes to 0, we obtain the limiting system (5) for which a standard finite volume/forward Euler method only needs to satisfy a stability restriction of the form $\Delta t \leq C\Delta x$, with C a constant that depends on the specific choice of the scheme.

In [7], it was proposed to use a projective integration method to accelerate such a brute-force integration; the idea, originating from [5], is the following. Starting from a computed numerical solution $\mathbf{f}^n$ at time $t^n = n\Delta t$, one first takes $K+1$ *inner* steps of size δt using (7), denoted as $\mathbf{f}^{n,k+1}$, in which the superscripts (n,k) denote the numerical solution at time $t^{n,k} = n\Delta t + k\delta t$. The aim is to obtain a discrete derivative to be used in the *outer* step to compute $\mathbf{f}^{n+1} = \mathbf{f}^{n+1,0}$ via extrapolation in time:

$$\mathbf{f}^{n+1} = \mathbf{f}^{n,K+1} + (\Delta t - (K+1)\delta t)\frac{\mathbf{f}^{n,K+1} - \mathbf{f}^{n,K}}{\delta t}. \tag{8}$$

Higher-order projective Runge–Kutta (PRK) methods can be constructed by replacing each time derivative evaluation $\mathbf{k}_s$ in a classical Runge–Kutta method by $K+1$ steps of an inner integrator as follows:

$$s = 1: \quad \begin{cases} \mathbf{f}^{n,k+1} &= \mathbf{f}^{n,k} + \delta t \mathrm{D}_t(\mathbf{f}^{n,k}), \qquad 0 \leq k \leq K \\ \mathbf{k}_1 &= \dfrac{\mathbf{f}^{n,K+1} - \mathbf{f}^{n,K}}{\delta t} \end{cases} \tag{9}$$

$$2 \leq s \leq S: \quad \begin{cases} \mathbf{f}_s^{n+c_s,0} &= \mathbf{f}^{n,K+1} + (c_s\Delta t - (K+1)\delta t)\sum_{l=1}^{s-1}\dfrac{a_{s,l}}{c_s}\mathbf{k}_l, \\ \mathbf{f}_s^{n+c_s,k+1} &= \mathbf{f}_s^{n+c_s,k} + \delta t \mathrm{D}_t(\mathbf{f}_s^{n+c_s,k}), \qquad 0 \leq k \leq K \\ \mathbf{k}_s &= \dfrac{\mathbf{f}_s^{n+c_s,K+1} - \mathbf{f}_s^{n+c_s,K}}{\delta t} \end{cases} \tag{10}$$

$$\mathbf{f}^{n+1} = \mathbf{f}^{n,K+1} + (\Delta t - (K+1)\delta t)\sum_{s=1}^{S} b_s \mathbf{k}_s. \tag{11}$$

To ensure consistency, the Runge–Kutta matrix $\mathbf{a} = (a_{s,i})_{s,i=1}^{S}$, weights $\mathbf{b} = (b_s)_{s=1}^{S}$, and nodes $\mathbf{c} = (c_s)_{s=1}^{S}$ satisfy the conditions $0 \leq b_s \leq 1$ and $0 \leq c_s \leq 1$, as well as:

$$\sum_{s=1}^{S} b_s = 1, \qquad \sum_{i=1}^{S-1} a_{s,i} = c_s, \quad 1 \leq s \leq S. \tag{12}$$

3 Spectral Properties

To choose the method parameters (the size of the small and large time steps δt and Δt, as well as the number K of small steps), one needs to analyze the spectrum of the collision operator. In [8], this was done in the hyperbolic scaling for a system with a linear Maxwellian that serves as a relaxation of a nonlinear hyperbolic conservation law.

By linearizing the Maxwellian (3) around the global Maxwellian distribution $\mathscr{M}_{\mathbf{v}}^{\rho^\infty,\bar{\mathbf{v}}^\infty,T^\infty} = \mathscr{M}_{\mathbf{v}}^{1,0,1}$, it is shown in [3, p.206] that the resulting linearized equilibrium can be written as:

$$\mathscr{M}_{\text{lin}}(f^\varepsilon)(\mathbf{x},\mathbf{v},t) = \sum_{k=0}^{D_v+1} \Psi_k(\mathbf{v})(\Psi_k, f^\varepsilon)(\mathbf{x},t), \tag{13}$$

in which the scalar product is defined by:

$$(g,h) = \int_{\mathbb{R}^{D_v}} g(\mathbf{v})\overline{h(\mathbf{v})}\frac{1}{(2\pi)^{D_v/2}}\exp\left(\frac{-|\mathbf{v}|^2}{2}\right)d\mathbf{v}. \tag{14}$$

Furthermore, the orthonormal set of basis functions $\Psi_k(\mathbf{v})$ in (13) are obtained from a straightforward application of the Gram-Schmidt process to the D_v+1 collision invariants $(1,\mathbf{v},|\mathbf{v}|^2)$, yielding:

$$\big(\Psi_0(\mathbf{v}),\ldots,\Psi_{D_v+1}(\mathbf{v})\big) = \left(1, v^1, ..., v^{D_v}, \frac{|\mathbf{v}|^2 - D_v}{2^{D_v/2}}\right). \tag{15}$$

Using the linearized Maxwellian (13), the linearized version of the full BGK equation (2) reads:

$$\partial_t f^\varepsilon + \mathbf{v}\cdot\nabla_{\mathbf{x}} f^\varepsilon = -\frac{1}{\varepsilon}(\mathscr{I} - \Pi_{\text{BGK}})f^\varepsilon, \tag{16}$$

where $\mathscr{I}$ denotes the identity operator and Π_{BGK} is the following rank-(D_v+2) projection operator:

$$\Pi_{\text{BGK}} f^\varepsilon = \sum_{k=0}^{D_v+1} \Psi_k(\mathbf{v})(\Psi_k, f^\varepsilon). \tag{17}$$

This shows that the structure of the linearized Maxwellian (13) and the linearized BGK projection operator (17) are almost identical to those in [8]. We can actually view these linear kinetic models as a special simplified case of the linearized BGK equation. Therefore, it is expected that the construction of stable, asymptotic-preserving projective integration methods for the full BGK equation (2) is practically identical to that in [8]. In particular, the conclusion is that, when choosing $\delta t = \varepsilon$, one is able to choose $\Delta t = O(\Delta x)$ and K independent of ε, resulting in a scheme with computational cost independent of ε.

4 Numerical Experiments

BGK in 1D. As a first experiment, we focus on the nonlinear BGK equation (2) in 1D. We consider a Sod-like test case for $x \in [0, 1]$ consisting of an initial centered Riemann problem with the following left and right state values:

$$\left(\rho_L, \bar{\mathbf{v}}_L, T_L\right) = (1, 0, 1), \qquad \left(\rho_R, \bar{\mathbf{v}}_R, T_R\right) = (0.125, 0, 0.25). \tag{18}$$

The initial distribution $f^\varepsilon(x, v, 0)$ is then chosen as the Maxwellian (3) corresponding to the above initial macroscopic variables. We impose outflow boundary conditions and perform simulations for $t \in [0, 0.15]$. As velocity space, we take the interval $[-8, 8]$, which we discretize on a uniform grid using $J = 80$ velocity nodes. In all simulations, space is discretized using the WENO3 spatial discretization with $\Delta x = 0.01$. Below, we compare solutions for three gas flow regimes: $\varepsilon = 10^{-1}$ (kinetic regime), $\varepsilon = 10^{-2}$ (transitional regime) and $\varepsilon = 10^{-5}$ (fluid regime).

In the kinetic ($\varepsilon = 10^{-1}$) and transitional ($\varepsilon = 10^{-2}$) regimes, we compute the numerical solution using the fourth order Runge–Kutta (RK4) time discretization with time step $\delta t = 0.1\Delta x$. In the fluid regime ($\varepsilon = 10^{-5}$), direct integration schemes such as RK4 become too expensive due to a severe time step restriction, which is required to ensure stability of the method. Exploiting that the spectrum of the linearized BGK equation is close to that of the linear kinetic models used in [8], see Sect. 3, we construct a projective integration method to accelerate time integration in the fluid regime. As inner integrator, we select the forward Euler time discretization with $\delta t = \varepsilon$. As outer integrator, we choose the fourth-order projective Runge–Kutta (PRK4) method, using $K = 2$ inner steps and an outer step of size $\Delta t = 0.4\Delta x$.

The results are shown in Fig. 2, where we display the density ρ, macroscopic velocity $\bar{\mathbf{v}}$ and temperature T as given in (4) at $t = 0.15$. In addition, we plot the heat flux q, which, in a general D_v-dimensional setting, is a vector $\mathbf{q} = \left(q^d\right)_{d=1}^{D_v}$ with components given by:

$$q^d = \frac{1}{2}\int_{\mathbb{R}^{D_v}} |\mathbf{c}|^2 \, c^d \, f^\varepsilon \, d\mathbf{v}, \tag{19}$$

in which $\mathbf{c} = \left(c^d\right)_{d=1}^{D_v} = \mathbf{v} - \bar{\mathbf{v}}$ is the peculiar velocity. The different regimes are shown by blue (kinetic), purple (transitional) and green (fluid) dots. The red line in each plot denotes the limiting ($\varepsilon \to 0$) solution of each macroscopic variable, which all converge to the solution of the compressible Euler equations (5) with ideal gas law $P = \rho T$ and heat flux $q = 0$. From this, we observe that the BGK solution is increasingly dissipative for increasing values of ε since the rate with which f^ε converges to its equilibrium $\mathscr{M}_v\left(f^\varepsilon\right)$ becomes slower. In contrast, for sufficiently small ε, relaxation to thermodynamic equilibrium occurs practically instantaneous and the Euler equations (5) yield a valid description. Since this is a hyperbolic system, it allows for the development of sharp discontinuous and shock waves which are clearly seen in the numerical solution.

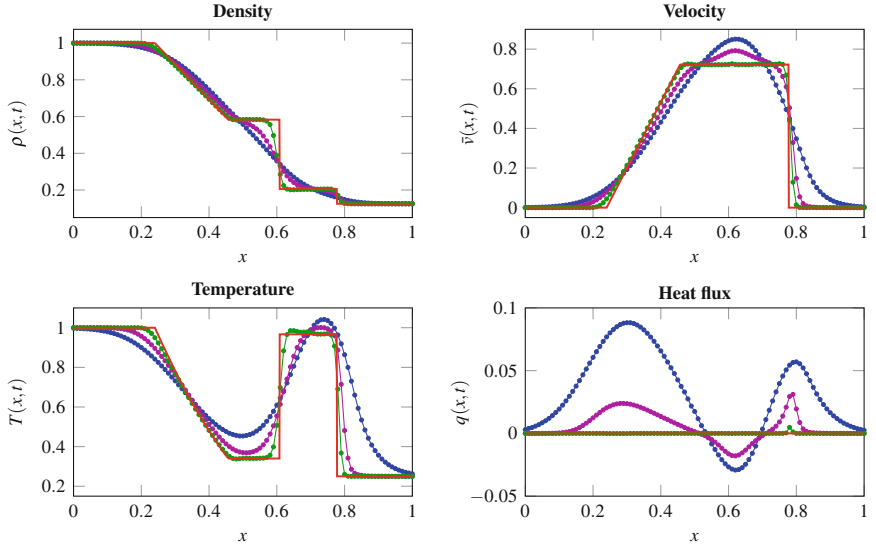


Fig. 2 Numerical solution of the BGK equation in 1D at $t = 0.15$ for a Sod-like shock test (18) using WENO3 with $\Delta x = 0.01$. RK4 is used for $\varepsilon = 10^{-1}$ (*blue dots*) and $\varepsilon = 10^{-2}$ (*purple dots*). The PRK4 method is used for $\varepsilon = 10^{-5}$ (*green dots*). *Red line* hydrodynamic limit ($\varepsilon \to 0$)

Shock-bubble interaction in 2D. Here, we consider the BGK equation in 2D and we investigate the interaction between a moving shock wave and a stationary smooth bubble, which was proposed in [9], see also [2]. This problem consists of a shock wave positioned at $x = -1$ in a spatial domain $\mathbf{x} = (x, y) \in [-2, 3] \times [-1, 1]$ traveling with Mach number $Ma = 2$ into an equilibrium flow region. Over the shock wave, the following left ($x \leq -1$) and right ($x > -1$) state values are imposed [2]:

$$\left(\rho_L, \bar{\mathbf{v}}_L^x, \bar{\mathbf{v}}_L^y, T_L\right) = \left(\frac{16}{7}, \sqrt{\frac{5}{3}}\frac{7}{16}, 0, \frac{133}{64}\right), \qquad \left(\rho_R, \bar{\mathbf{v}}_R, T_R\right) = (1, \mathbf{0}, 1)\,. \quad (20)$$

Due to this initial profile, the shock wave will propagate rightwards into the flow region at rest ($x > -1$). Moreover, in this equilibrium region, a smooth Gaussian density bubble centered at $\mathbf{x}_0 = (0.5, 0)$ is placed, given by:

$$\rho(\mathbf{x}, 0) = 1 + 1.5 \exp\left(-16\,|\mathbf{x} - \mathbf{x}_0|^2\right). \quad (21)$$

Then, the initial distribution $f^\varepsilon(\mathbf{x}, \mathbf{v}, 0)$ is chosen as the Maxwellian (3) corresponding to the initial macroscopic variables in (20)–(21). We impose outflow and periodic boundary conditions along the x- and y-directions, respectively, and we perform simulations for $t \in [0, 0.8]$. As velocity space, we take the domain $[-10, 10]^2$, which we discretize on a uniform grid using $J_x = J_y = 30$. We discretize space using the

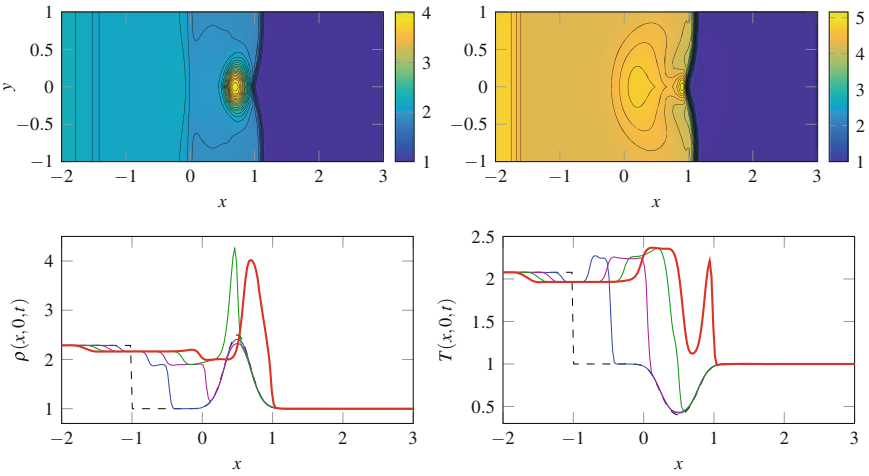


Fig. 3 Shock-bubble interaction. *Top* contour plot of density (*left*) and pressure (*right*). *Bottom* density (*left*) and temperature (*right*) along $y = 0$ at $t = 0$ (*black dashed*), $t = 0.2$ (*blue*), $t = 0.4$ (*purple*), $t = 0.6$ (*green*) and $t = 0.8$ (*red*)

WENO2 spatial discretization with $I_x = 200$ and $I_y = 25$. Furthermore, we consider a fluid regime by taking $\varepsilon = 10^{-5}$.

We construct a PRK4 method with FE as inner integrator to speed up simulation in time. The inner time step is fixed as $\delta t = \varepsilon$ and we use $K = 2$ inner steps in each outer integrator iteration. The outer time step is chosen as $\Delta t = 0.4\Delta x$. To compare our results with those in [9], where the smallest value of ε is chosen as $\varepsilon = 10^{-2}$, we regard the one-dimensional evolution of density and temperature along the axis $y = 0$. For $t \in \{0, 0.2, 0.4, 0.6, 0.8\}$, we plot these intersections in Fig. 3. We conclude that we obtain the same solution structure at $t = 0.8$ as in [9]. However, our results are sharper and less dissipative supposedly due to the particular small value of ε (10^{-5} vs. 10^{-2}). In contrast to [2], we nicely capture the swift changes in the temperature profile for $x \in [0.5, 1]$ at $t = 0.8$.

References

1. Bhatnagar, P.L., Gross, E.P., Krook, M.: A model for collision processes in gases. I. Small amplitude processes in charged and neutral one-component systems. Phys. Rev. **94**(3) (1954)
2. Cai, Z., Li, R.: Numerical regularized moment method of arbitrary order for boltzmann-BGK equation. SIAM J. Sci. Comput. **32**(5), 2875–2907 (2010)
3. Cercignani, C.: The Boltzmann Equation and Its Applications. Springer Science & Business Media, Heidelberg (1988)
4. Dimarco, G., Pareschi, L.: Numerical methods for kinetic equations. Acta Numer. **23**, 369–520 (2014)

5. Gear, C.W., Kevrekidis, I.G.: Projective methods for stiff differential equations: problems with gaps in their eigenvalue spectrum. SIAM J. Sci. Comput. **24**(4), 1091–1106 (2003)
6. Jin, S.: Efficient asymptotic-preserving (AP) schemes for some multiscale kinetic equations. SIAM J. Sci. Comput. **21**(2), 441–454 (1999)
7. Lafitte, P., Samaey, G.: Asymptotic-preserving projective integration schemes for kinetic equations in the diffusion limit. SIAM J. Sci. Comput. **34**(2), A579–A602 (2012)
8. Lafitte, P., Melis, W., Samaey, G.: A high-order relaxation method with projective integration for solving nonlinear systems of hyperbolic conservation laws (2016). Submitted
9. Torrilhon, M.: Two-dimensional bulk microflow simulations based on regularized Grad's 13-moment equations. Multi. Model. Simul. **5**(3), 695–728 (2006)

Asymptotic Preserving Property of a Semi-implicit Method

Lei Zhang, Jean-Michel Ghidaglia and Anela Kumbaro

Abstract This work focuses on the study of the asymptotic preserving property of a semi-implicit method. The semi-implicit method, which is a pressure-based method, has been successfully used to simulate two-phase flows in numerous industrial applications. This method is used in our studies due to the fact that pressure-based methods generally perform well at low Mach numbers. The semi-implicit method is applied to the homogeneous equilibrium model (HEM) in this work to simulate two-phase flows. We show that the semi-implicit method is asymptotic preserving, i.e. the discretization for a compressible model tends to a consistent discretization for the related incompressible model at the low Mach number limit. Finally, test cases are performed to show that the numerical method is able to deal with low Mach number flows, as well as flows with a wide range of Mach numbers.

Keywords Semi-implicit method · Asymptotic preserving · Homogeneous equilibrium model · Low mach number

1 Introduction

In numerous situations, there are low Mach number regions within a globally compressible flow, for instance in a nozzle with a large variation in cross-sectional area, in two-phase flows due to the mixture of liquid and gas, etc. Consequently it is important to develop a numerical method capable of treating both the compressible (local

L. Zhang (✉) · J.-M. Ghidaglia
CMLA, ENS Paris-Saclay, CNRS, Université Paris-Saclay,
61 avenue du président Wilson, 94230 Cachan, France
e-mail: lzhang@cmla.ens-cachan.fr; tongjizhanglei@gmail.com

J.-M. Ghidaglia
e-mail: ghidaglia@free.fr

A. Kumbaro
CEA-Saclay DEN, DM2S, STMF, LMEC, F91191 Gif-sur-yvette Cedex, France
e-mail: anela.kumbaro@cea.fr

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_17

Mach number of order O(1)) and the incompressible regime (very small local Mach number), which is referred to as all-speed scheme in the literature.

Several authors have used the notion of asymptotic preserving as a guideline to design such appropriate numerical methods [4–6] with the following definition [4]: for a physical model $\mathscr{M}^\varepsilon$ with a perturbation parameter ε (in our context, ε is related to the Mach number), $\mathscr{M}^0$ represents the limit of $\mathscr{M}^\varepsilon$ when $\varepsilon \longrightarrow 0$; given $\mathscr{M}^\varepsilon_\Delta$ a discretization of the physical model $\mathscr{M}^\varepsilon$, is said to be asymptotic preserving if its limit as $\varepsilon \longrightarrow 0$ is a consistent discretization of the model $\mathscr{M}^0$, moreover the stability condition should be independent of the parameter ε.

For density-based methods, the CFL time step restriction becomes extremely stringent for small Mach numbers due to the large discrepancy between the sound speed and characteristic flow velocity. Therefore a pressure-based method, called semi-implicit method [9, 10, 12], is adopted in our studies to simulate two-phase flows using the homogeneous equilibrium model (HEM). In this work, we show that the semi-implicit method is asymptotic preserving.

This paper is organized as follows. In Sect. 2, the homogeneous equilibrium model (HEM) is introduced, as well as the semi-implicit method used to solve this model. Section 3 gives the low Mach limit of the HEM and shows that the semi-implicit method is asymptotic preserving. Section 4 presents two test cases to illustrate that the semi-implicit method is able to deal with low Mach number flows, as well as flows with a large range of Mach numbers.

2 HEM and Its Semi-implicit Discretization

The homogeneous equilibrium model (HEM), which assumes a dynamic and thermal equilibrium between the two phases of a fluid [13], is used in our study. Although it is the simplest model for two phase flows it has been used by several authors [14, 15] to simulate applications of industrial interest. The scaled HEM used in our work is given as follows [16]

$$(\tilde{\rho})_t + \tilde{\nabla} \cdot (\tilde{\rho}\tilde{\mathbf{u}}) = 0, \tag{1}$$

$$\tilde{\rho}(\tilde{\mathbf{u}})_t + \tilde{\nabla} \cdot (\tilde{\rho}\tilde{\mathbf{u}} \otimes \tilde{\mathbf{u}}) - \tilde{\mathbf{u}}\tilde{\nabla} \cdot (\tilde{\rho}\tilde{\mathbf{u}}) + \tilde{\nabla}\tilde{P}/M^2 = 0, \tag{2}$$

$$(\tilde{\rho}\tilde{h})_t + \tilde{\nabla} \cdot (\tilde{\rho}\tilde{e}\tilde{\mathbf{u}}) + \tilde{P}\tilde{\nabla} \cdot \tilde{\mathbf{u}} - (\tilde{P})_t = 0, \tag{3}$$

which can be obtained by rescaling variables $(\rho,\ \mathbf{u},\ \mathbf{x},\ P,\ t,\ h,\ e)$ in the original HEM with respect to the reference parameters $(\rho_{\text{ref}},\ u_{\text{ref}},\ x_{\text{ref}},\ P_{\text{ref}},\ \frac{x_{\text{ref}}}{u_{\text{ref}}},\ \frac{P_{\text{ref}}}{\rho_{\text{ref}}},\ \frac{P_{\text{ref}}}{\rho_{\text{ref}}})$ [4, 5], for example $\tilde{\rho} = \rho/\rho_{\text{ref}}$. The parameter $M = \sqrt{\rho_{\text{ref}}\, u_{\text{ref}}^2/P_{\text{ref}}}$ represents the global Mach number which characterizes a fluid flow. In fact this number is known in the literature as the Euler number. However in the case of a perfect gas, since $\gamma\ P_{\text{ref}} = \rho_{\text{ref}}\, c_{\text{ref}}^2$, we have $\sqrt{\rho_{\text{ref}}\, u_{\text{ref}}^2/P_{\text{ref}}} = \sqrt{\gamma}\, u_{\text{ref}}/c_{\text{ref}}$ (where γ is the specific heat ratio, and c_{ref} is a reference speed of sound), and we call this number the global Mach number.

In the remainder of the paper, the tildes will be omitted for simplicity of notation, in addition we note $\varepsilon = M^2$. The specific enthalpy h is used in energy equation (3), because in the initial stages of development, we have worked with a water table with pressure P and specific enthalpy h as independent thermodynamic variables. One can notice that the HEM has the same form as the Euler equations, except that the physical quantities are interpreted differently, i.e. ρ, h, e are respectively the density, the specific enthalpy, and the specific internal energy of a mixture, $\mathbf{u}$ is the common velocity of the two phases, P is the pressure. In addition we have the following thermodynamic relation $h = e + P/\rho$.

The resolution of the HEM is the same as for the Euler equations, except that it is necessary to determine whether a fluid is a mixture or single phase (liquid or gas) in order to apply different equations of state. This can be done by deducing the gas mass fraction [3] using the following formula once the pressure P and the specific enthalpy h are obtained

$$X = \left(h - h_l^{\text{sat}}(P)\right) / \left(h_v^{\text{sat}}(P) - h_l^{\text{sat}}(P)\right), \tag{4}$$

where h_v^{sat} and h_l^{sat}, which are functions of pressure P, are respectively the saturation specific enthalpy for the gas and for the liquid. If $0 < X < 1$, i.e. the fluid is a mixture of gas and liquid, the density is given by

$$1/\rho = X/\rho_v^{\text{sat}}(P) + (1 - X)/\rho_l^{\text{sat}}(P), \tag{5}$$

where ρ_v^{sat} and ρ_l^{sat} are respectively the saturation density for the gas and for the liquid, and they are only dependent on the pressure P. Otherwise the fluid is single phase, and we are in fact resolving the Euler equations. As the HEM and its discretization is the same as the Euler equations, these two terms will be used interchangeably in the following.

Using the semi-implicit method [9, 10] for the scaled system (1)–(3) over a control volume K yields

$$\frac{\rho_K^{n+1} - \rho_K^n}{\Delta t} + \frac{\sum_f (\rho)_f^n (\mathbf{u})_f^{n+1} \cdot \mathbf{S}_f}{V_K} = 0, \tag{6}$$

$$\rho_K^n \frac{\mathbf{u}_K^{n+1} - \mathbf{u}_K^n}{\Delta t} + \frac{1}{V_K} \sum_f (\rho \mathbf{u})_f^n (\mathbf{u})_f^n \cdot \mathbf{S}_f - \frac{\mathbf{u}_K^n}{V_K} \sum_f (\rho)_f^n (\mathbf{u})_f^n \cdot \mathbf{S}_f + \frac{1}{\varepsilon} \nabla P_K^{n+1} = 0, \tag{7}$$

$$\frac{(\rho h)_K^{n+1} - (\rho h)_K^n}{\Delta t} + \frac{\sum_f (\rho e)_f^n (\mathbf{u})_f^{n+1} \cdot \mathbf{S}_f}{V_K} + \frac{P_K^n}{V_K} \sum_f (\mathbf{u})_f^{n+1} \cdot \mathbf{S}_f - \frac{P_K^{n+1} - P_K^n}{\Delta t} = 0, \tag{8}$$

where V_K is the volume of cell K, $\mathbf{S}_f$ is the area vector on face f pointing outward from cell K. In the semi-implicit method, the pressure gradient term in momentum equation (7) and the fluid velocity at a cell face in scalar equations (6) and (8) are

evaluated implicitly, whereas all other terms are determined explicitly. The pressure gradient is evaluated using the Green–Gauss reconstruction [2]. In addition, as in [9], Rhie and Chow's interpolation [11] is used to determine the velocity at the cell face between the two cells K and L at time instant t^{n+1}:

$$\mathbf{u}_f^{n+1} = \frac{1}{2}(\mathbf{u}_K^{n+1} + \mathbf{u}_L^{n+1}) - \frac{1}{2\varepsilon}\left(\frac{\Delta t}{\rho_K^n} + \frac{\Delta t}{\rho_L^n}\right)\frac{P_L^{n+1} - P_K^{n+1}}{|\mathbf{dr}_{KL}|}\mathbf{n}_f + \frac{1}{2\varepsilon}\frac{\Delta t}{\rho_K^n}\nabla P_K^{n+1} + \frac{1}{2\varepsilon}\frac{\Delta t}{\rho_K^n}\nabla P_L^{n+1}, \tag{9}$$

where $\mathbf{dr}_{KL}$ is the vector joining the center of cell K to the center of cell L and $\mathbf{n}_f$ is the normal vector on face f pointing outward from cell K to cell L. This interpolation method is used to prevent the well-known checker-board problem [7] encountered on a mesh with co-located variables (all variables are located at the same position in a cell). This formulation is used to calculate the velocity at cell face for all the discretized Eqs. (6)–(8), however the convective term $(\rho\mathbf{u})_f^n$ is determined using the upwind scheme depending on the sign of $\mathbf{u}_f^n$. The resolution strategy for the system consists of obtaining a $N \times N$ (N is the number of cells in a mesh) linear system containing the pressure as an unknown variable from the above discretization, and one can refer to [9, 16] for details. It should be noticed that this numerical method is not conservative, and consequently is not able to capture exactly the shock for compressible flows. In [16], a conservative version of the numerical method was developed.

3 Asymptotic Preserving Property

The low Mach limit (where $\bar{P}$ is a constant) of the Euler equations is [4]:

$$\rho_0(\mathbf{u}_0)_t + \nabla\cdot(\rho_0\mathbf{u}_0\otimes\mathbf{u}_0) - \mathbf{u}_0\nabla\cdot(\rho_0\mathbf{u}_0) + \nabla P_1 = 0, \tag{10}$$

$$P_0 = \bar{P}, \tag{11}$$

$$\nabla\cdot\mathbf{u}_0 = 0, \tag{12}$$

$$(\rho_0)_t + \nabla\cdot(\rho_0\mathbf{u}_0) = 0, \tag{13}$$

which can be obtained by performing the asymptotic development for the physical quantities

$$(\cdot) = (\cdot)_0 + \varepsilon(\cdot)_1 + \cdots, \tag{14}$$

where $(\cdot)$ represents the pressure P, velocity $\mathbf{u}$, specific internal energy e, specific enthalpy h and density ρ.

Now we can show that the semi-implicit discretization (6)–(8) of the scaled system (1)–(3) tends to a discretization of the low Mach limit (10)–(13) as $\varepsilon \longrightarrow 0$. As for the continuous problem, we can postulate for physical quantities at the center of cell K that $(\cdot)_K^m = (\cdot)_{0,K}^m + \varepsilon(\cdot)_{1,K}^m$, where $m \in \{n,\ n+1\}$, and for the fluid velocity at the cell face $(\mathbf{u})_f^m = (\mathbf{u})_{0,f}^m + \varepsilon(\mathbf{u})_{1,f}^m$. We suppose that at time instant t^n: $P_{0,K}^n = \bar{P}$ (a constant), thus the upwind part used to calculate $(\mathbf{u})_{0,f}^n$ of order $O\left(\frac{1}{\varepsilon}\right)$ disappears (see Eq. (9)). Then the discretized momentum Eq. (7) leads to (the equality of terms of order $O\left(\frac{1}{\varepsilon}\right)$ and with appropriate boundary conditions (see [16] for more details)):

$$P_{0,K}^{n+1} = \bar{P}. \tag{15}$$

Therefore with a well prepared initial condition $P_{0,K}^0 = \bar{P}$, the pressure remains uniform in the domain. The equality of terms of order $O(1)$ in mass equation (6), momentum equation (7) and energy equation (8) leads to respectively

$$\frac{\rho_{0,K}^{n+1} - \rho_{0,K}^n}{\Delta t} + \frac{\sum_f (\rho)_{0,f}^n (\mathbf{u})_{0,f}^{n+1} \cdot \mathbf{S}_f}{V_K} = 0, \tag{16}$$

$$\rho_{0,K}^n \frac{\mathbf{u}_{0,K}^{n+1} - \mathbf{u}_{0,K}^n}{\Delta t} + \frac{\sum_f (\rho\mathbf{u})_{0,f}^n (\mathbf{u})_{0,f}^n \cdot \mathbf{S}_f}{V_K} - \mathbf{u}_{0,K}^n \frac{\sum_f (\rho)_{0,f}^n (\mathbf{u})_{0,f}^n \cdot \mathbf{S}_f}{V_K} + \nabla P_{1,K}^{n+1} = 0, \tag{17}$$

$$\frac{(\rho e)_{0,K}^{n+1} - (\rho e)_{0,K}^n}{\Delta t} + \frac{\sum_f (\rho e)_{0,f}^n (\mathbf{u})_{0,f}^{n+1} \cdot \mathbf{S}_f}{V_K} + P_{0,K}^n \frac{\sum_f (\mathbf{u})_{0,f}^{n+1} \cdot \mathbf{S}_f}{V_K} = 0. \tag{18}$$

It can be seen that the discretizations (16) and (17) are respectively a consistent discretization of Eqs. (13) and (10). It remains to be shown that a consistent discretization of Eq. (12) can be found from the discretized energy equation (18). Inspired by the work of [4], the linearisation of ρe with respect to ρ and P leads to (the term concerning the derivative of ρe with respect to P is zero as the pressure is constant)

$$\frac{(\rho e)_{0,K}^{n+1} - (\rho e)_{0,K}^n}{\Delta t} = \frac{1}{\Delta t} \left(\frac{\partial(\rho_0 e_0)}{\partial \rho_0} \right)_{\bar{P}} (\rho_{0,K}^n,\ \bar{P})(\rho_{0,K}^{n+1} - \rho_{0,K}^n), \tag{19}$$

$$\frac{\sum_f (\rho e)_{0,f}^n (\mathbf{u})_{0,f}^{n+1} \cdot \mathbf{S}_f}{V_K} = \left(\frac{\partial(\rho_0 e_0)}{\partial \rho_0} \right)_{\bar{P}} (\rho_{0,K}^n,\ \bar{P}) \frac{\sum_f (\rho_{0,f}^n - \rho_{0,K}^n)(\mathbf{u})_{0,f}^{n+1} \cdot \mathbf{S}_f}{V_K} + (\rho e)_{0,K}^n \frac{\sum_f (\mathbf{u})_{0,f}^{n+1} \cdot \mathbf{S}_f}{V_K}, \tag{20}$$

which can be combined with the discretized mass equation (16) to obtain

$$\left[\bar{P} + (\rho e)_{0,K}^n - \left(\frac{\partial(\rho_0 e_0)}{\partial \rho_0} \right)_{\bar{P}} (\rho_{0,K}^n,\ \bar{P}) \rho_{0,K}^n \right] \frac{\sum_f (\mathbf{u})_{0,f}^{n+1} \cdot \mathbf{S}_f}{V_K} = 0. \tag{21}$$

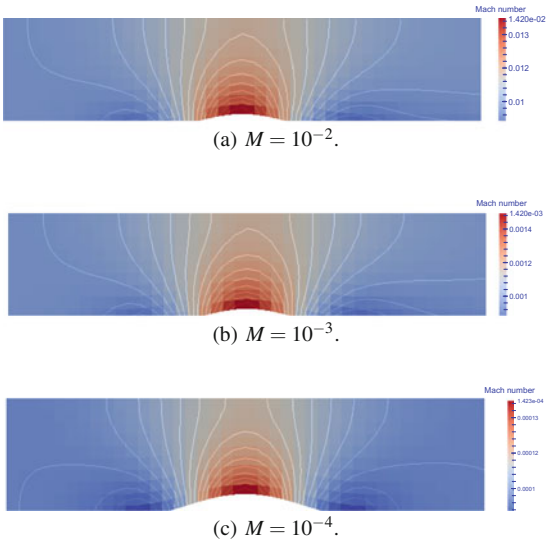
The coefficient $\bar{P} + (\rho e)^n_{0,K} - \left(\frac{\partial(\rho_0 e_0)}{\partial \rho_0}\right)_{\bar{P}} (\rho^n_{0,K}, \bar{P})\rho^n_{0,K}$ is generally different to 0 (e.g. for stiffened gas [16]), thus we have $\sum_f (\mathbf{u})^{n+1}_{0,f} \cdot \mathbf{S}_f = 0$, which is effectively a consistent discretization for Eq. (12).

Furthermore, it was shown in [16] that the stability condition for the semi-implicit method with co-located variables is the CFL condition limited by the fluid velocity, and hence is independent of Mach number. In conclusion, the semi-implicit method for the HEM is asymptotic preserving.

4 Numerical Results

4.1 Single Phase Flow in a Channel with Bump

This test case consists of a single phase flow (Euler equations) at low Mach numbers in a channel, which is also studied in [1, 3, 8]. The computational domain is a 4 m long and 1 m high rectangle with a geometric perturbation in the lower wall. The initial conditions are: $p = 10^5$ Pa, $\mathbf{u} = (u_x, 0)$ m/s, $h = 25 \times 10^3$ J/kg, with $u_x = 1,\ 0.1,\ 0.01$ to have the corresponding Mach numbers $M = 10^{-2}, 10^{-3}, 10^{-4}$ (the perfect gas equation of state with $\gamma = 1.4$ is used). At the inlet, the velocity and the specific enthalpy are imposed, whereas the pressure is given at the outlet, with these values applied throughout the domain as the initial conditions. A slip condition is specified for the walls. The objective of this test case is to study the behaviour of the numerical method at low Mach numbers. The Mach number contours for different fluid velocities at the inlet are presented in Fig. 1. As expected [1], symmetry of the curves with respect to the geometry is obtained.



(a) $M = 10^{-2}$.

(b) $M = 10^{-3}$.

(c) $M = 10^{-4}$.

Fig. 1 Mach number contours for different fluid velocities at the inlet

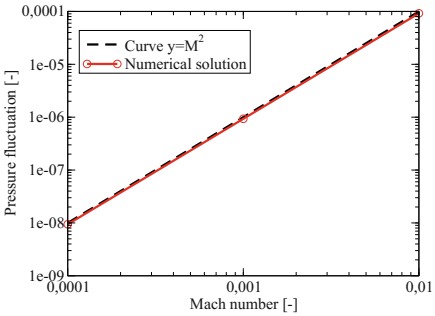
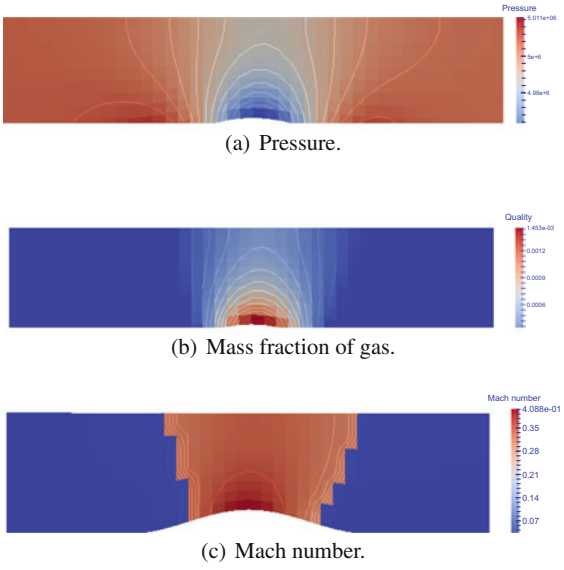


Fig. 2 The pressure variation as a function of Mach number

We also evaluate the pressure variation against Mach number, as indicated in Fig. 2. We can remark that the pressure variation is of order $O(M^2)$, which agrees well with the continuous model (see Eq. (14)). The pressure variation is defined by $P_{\text{var}} = (P_{\text{max}} - P_{\text{min}}) / P_{\text{max}}$, with P_{max} the pressure maximum, and P_{min} the pressure minimum.

4.2 *Two-Phase Flow in a Channel with Bump*

This test case [3] is similar to the test in the previous section, however at the inlet the fluid is a liquid that is close to saturation. Therefore, as the pressure drops around the geometric perturbation (Fig. 3a), a small amount of liquid evaporates, as



(a) Pressure.

(b) Mass fraction of gas.

(c) Mach number.

Fig. 3 Numerical results for two-phase flow in a channel with bump

indicated in Fig. 3b. Due to the large discrepancy between the sound speed in a liquid and in a mixture, we can observe a dramatic Mach number change across the phase transition lines, as illustrated in Fig. 3c. These numerical results agree qualitatively with those obtained in [3]. This test case shows that the semi-implicit method is able to simulate flows with a large range of Mach numbers.

References

1. Bijl, H., Weseling, P.: A unified method for computing incompressible and compressible flows in boundary-fitted coordinates. J. Comput. Phys. (1998)
2. Christon, M.A.: The consistency of pressure-gradient approximations used in multi-dimensional shock hydrodynamics. Int. J. Numer. Methods Fluids (2010)
3. Clerc, S.: Numerical simulation of the homogeneous equilibrium model for two-phase flows. J. Comput. Phys. (2000)
4. Cordier, F., Degond, P., Kumbaro, A.: An asymptotic-preserving all-speed scheme for the Euler and Navier–Stokes equations. J. Comput. Phys. (2012)
5. Degond, P., Tang, M.: All speed scheme for the low mach number limit of the isentropic Euler equations. Commun. Comput. Phys. (2011)
6. Dimarco, G., Loubère, R., Vignal, M.H.: Study of a new asymptotic preserving scheme for the Euler system in the low mach number limit. <hal-01297238> (2016)
7. Ferziger, J.H., Peric, M.: Computational Methods for Fluid Dynamics. Springer, Heidelberg (2002)
8. Van der Heul, D., Vuik, C., Wesseling, P.: A conservative pressure-correction method for flow at all speeds. Comput. Fluids (2003)
9. Jeong, J.J., et al.: Developement and preliminary assessment of a three-dimensional thermal hydraulics code, CUPID. Nucl. Eng. Technol. (2010)
10. Liles, D.R., Reed, W.H.: A semi-implicit method for two-phase fluid dynamics. J. Comput. Phys. (1978)
11. Rhie, C.M., Chow, W.L.: Numerical study of the turbulent flow past an airfoil with trailing edge separation. AIAA (1983)
12. Shieh, A.S., et al.: RELAP5/MOD3 Code Manuel, Validation of Numerical Techniques in RELAP5/MOD3 (1994)
13. Stewart, H.B., Wendroff, B.: Two-phase flow: models and methods. J. Comput. Phys. (1984)
14. Toumi, I.: A weak formulation of Roe's approximate Riemann solver. J. Comput. Phys. (1992)
15. De Vuyst, F., Ghidaglia, J.M., Coq, G.L.: On the numerical simulation of multiphase water flows with changes of phase and strong gradients using the homogeneous equilibrium model. Int. J. Finite Vol. (2005)
16. Zhang, L.: Modélisation, analyse et simulation d'écoulements en thermohydraulique par modèles 6 équations. Ph.D. thesis, ENS Paris-Saclay (in preparation)

A Finite-Volume Discretization of Viscoelastic Saint-Venant Equations for FENE-P Fluids

Sébastien Boyaval

Abstract Saint-Venant equations can be generalized to account for a viscoelastic rheology in shallow flows. A Finite-Volume discretization for the 1D Saint-Venant system generalized to Upper-Convected Maxwell (UCM) fluids was proposed in Bouchut and Boyaval (M3AS 23(08): 1479–1526, 2013, [6]), which preserved a physically-natural stability property (i.e. free-energy dissipation) of the full system. It invoked a relaxation scheme of Suliciu type for the numerical computation of approximate solutions to Riemann problems. Here, the approach is extended to the 1D Saint-Venant system generalized to the finitely-extensible nonlinear elastic fluids of Peterlin (FENE-P). We are currently not able to ensure all stability conditions a priori, but the scheme is fully computable. And, using numerical simulations, it may help understand the famous High-Weissenberg number problem (HWNP) well-known in computational rheology.

Keywords Saint-venant equations · Fene-p viscoelastic fluids · Finite-volume · Simple Riemann solver · Suliciu relaxation scheme

MSC (2010): 65M08 · 65N08 · 35Q30

1 Introduction

Saint-Venant equations standardly model shallow free-surface gravity flows and can be generalized to account for the viscoelastic rheology of non-Newtonian fluids [7], Upper-Convected Maxwell (UCM) fluids in particular [6]. Here, we consider a generalized Saint-Venant (gSV) system for *finitely-extensible nonlinear elastic* fluids

S. Boyaval (✉)
Laboratoire D'hydraulique Saint-Venant (Ecole des Ponts ParisTech – EDF R&D – CEREMA), Université Paris-Est, EDF'lab 6 Quai Watier, 78401 Chatou Cedex, France
e-mail: sebastien.boyaval@enpc.fr

S. Boyaval
INRIA MATHERIALS, Paris, France

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_18

using Peterlin closure (i.e. FENE-P fluids [3]) in Cartesian coordinates

$$\partial_t h + \partial_x(hu) = 0 \tag{1}$$

$$\partial_t(hu) + \partial_x\left(hu^2 + gh^2/2 + hN\right) = 0 \tag{2}$$

$$\lambda\left(\partial_t\sigma_{xx} + u\partial_x\sigma_{xx} + 2(\zeta-1)\sigma_{xx}\partial_x u\right) = 1 - \sigma_{xx}/(1-(\sigma_{zz}+\sigma_{xx})/\ell) \tag{3}$$

$$\lambda\left(\partial_t\sigma_{zz} + u\partial_x\sigma_{zz} + 2(1-\zeta)\sigma_{zz}\partial_x u\right) = 1 - \sigma_{zz}/(1-(\sigma_{zz}+\sigma_{xx})/\ell) \tag{4}$$

for 1D $\mathbf{e}_y$-translation invariant flow along $\mathbf{e}_x$ under a uniform gravity field $-g\mathbf{e}_z$ with

- mean flow depth $h(t,x) > 0$ (in case of a non-rugous flat bottom),
- mean flow velocity $u(t,x)$ (for *uniform* cross sections), plus
- a normal-stress difference $N = G(\sigma_{zz}-\sigma_{xx})/(1-(\sigma_{zz}+\sigma_{xx})/\ell)$ given by the two conformation variables $\sigma_{zz}, \sigma_{xx} > 0$ and *constrained by* $0 < \sigma_{zz}+\sigma_{xx} < \ell$.

The nonlinear formula $N(\sigma_{zz},\sigma_{xx})$ accounts for finite-extensibility effects of matter in the elastic response, which are not present in UCM fluids [3]. These are controlled by the parameter $\ell > 0$, and one formally recovers UCM fluids [6] (with linear response) when $\ell \to \infty$. Besides, the relaxation time $\lambda \geq 0$ and the elasticity modulus $G > 0$ bear the same meaning as for UCM fluids. In particular, when $\lambda, G^{-1} \to 0$ and $G\lambda < \infty$, (1), (2), (3) and (4) formally reduces to the viscous Saint-Venant system with viscosity $\nu \equiv 2\lambda G$. Last, note that (3) and (4) invoke the quite general Gordon–Schowalter derivatives with a slip parameter $\zeta \in [0, \frac{1}{2}]$ constrained by the hyperbolicity of the system (1), (2), (3) and (4). (This follows after an easy computation similar to [9]).

In this work, we discuss a Finite-Volume method to solve (numerically) the Cauchy problem for the nonlinear hyperbolic 1D system (1), (2), (3) and (4). Standardly, we need to consider *weak* solutions (in fact, to (6), (7), (8) and (9), see below) plus *admissibility* constraints that are physically-meaningful dissipation rules formalizing the thermodynamics second principle close to an equilibrium [10]. Here, we consider the *inequality* associated with the companion conservation law for the *free-energy*

$$F = h\left(\frac{u^2}{2} + \frac{gh}{2} - \frac{G}{2(1-\zeta)}\left(\ell\log\left((\ell-(\sigma_{xx}+\sigma_{zz}))/(\ell-2)\right) + \log(\sigma_{xx}\sigma_{zz})\right)\right)$$

that is, on denoting the impulse by $P = gh^2/2 + hN$,

$$-\frac{Gh}{2(1-\zeta)\lambda}\left(\sigma_{xx}^{-1}\left(1-\frac{\sigma_{xx}}{1-(\sigma_{zz}+\sigma_{xx})/\ell}\right)^2 + \sigma_{zz}^{-1}\left(1-\frac{\sigma_{zz}}{1-(\sigma_{zz}+\sigma_{xx})/\ell}\right)^2\right)$$
$$=: D \geq \partial_t F + \partial_x\left(u(F+P)\right) \tag{5}$$

where the left-hand-side is obviously non-positive on the admissibility domain

$$\mathscr{U}^\ell := \{0 < h, 0 < \sigma_{xx}, 0 < \sigma_{zz}, \sigma_{xx}+\sigma_{zz} < \ell\}.$$

Note that we do not consider the vacuum state $h = 0$ as admissible here, see [9].

2 Finite-Volume Discretization of FENE-P/Saint-Venant

Piecewise-constant approximate solutions to the Cauchy problem on $(t, x) \in [0, T) \times \mathbb{R}$ for the gSV system can be defined by a Finite-Volume (FV) method. With a view to preserving $\mathscr{U}^\ell$ and the dissipation (5) after discretization, we choose $q = (h, hu, h\sigma_{xx}, h\sigma_{zz})$ as discretization variable. Indeed, the free-energy functional F is *convex* on the convex domain $\mathscr{U}^\ell \ni q$ (this follows after an easy computation from [5, Lemma 1.3]) while it is not convex in the variable $(h, hu, h\Pi, h\Sigma)$ whatever smooth invertible functions ϖ, ς are used for the reformulation of gSV

$$\partial_t h + \partial_x(hu) = 0 \tag{6}$$

$$\partial_t(hu) + \partial_x\left(hu^2 + \frac{gh^2}{2} + hN\right) = 0 \tag{7}$$

$$\partial_t(h\Pi) + \partial_x(hu\Pi) = \frac{h^{3-2\zeta}\varpi'(\sigma_{xx}h^{2(1-\zeta)})}{\lambda}\left(1 - \frac{\sigma_{xx}}{1 - \frac{\sigma_{zz}+\sigma_{xx}}{\ell}}\right) \tag{8}$$

$$\partial_t(h\Sigma) + \partial_x(hu\Sigma) = \frac{h^{2\zeta-1}\varsigma'(\sigma_{zz}h^{2(\zeta-1)})}{\lambda}\left(1 - \frac{\sigma_{zz}}{1 - \frac{\sigma_{zz}+\sigma_{xx}}{\ell}}\right) \tag{9}$$

with $\Pi = \varpi(\sigma_{xx}h^{2(1-\zeta)})$, $\Sigma = \varsigma(\sigma_{zz}h^{2(\zeta-1)})$ (computations are similar to [6, Appendix]). In the sequel, we therefore discretize a quasilinear system with source

$$\partial_t q + A(q)\partial_x q = S(q)\,, \tag{10}$$

but thanks to (6), (7), (8) and (9) and the dissipation rule (5), there is no ambiguity (for those discontinuous solutions built using a Riemann solver at least, see [2, 9, 12]).

2.1 Time Splitting Method

In a cell $(x_{i-1/2}, x_{i+1/2})$, $i \in \mathbb{Z}$, with volume $\Delta x_i = x_{i+1/2} - x_{i-1/2} > 0$ and center $x_i = (x_{i-1/2} + x_{i+1/2})/2$, we approximate q solution to (10) on $\mathbb{R}_{\geq 0} \times \mathbb{R} \ni (t, x)$ by

$$q_i^{n+1} \approx \frac{1}{\Delta x_i}\int_{x_{i-1/2}}^{x_{i+1/2}} q(t, x)dx\,,\ i \in \mathbb{Z}, t \in (t^n, t^{n+1}]$$

on a time grid $0 = t^0 < t^1 < \cdots < t^n < t^{n+1} < \cdots < t^N = T$ where $\Delta t^n = |t^{n+1} - t^n|$ will be chosen small enough compared with $\Delta x = \sup_{i\in\mathbb{Z}} \Delta x_i < \infty$ to ensure stability.

More precisely, we have in mind the numerical approximation of q as solution to a (well-posed) Cauchy problem for (10) on $\mathbb{R}_{\geq 0} \times \mathbb{R}$ given some ini-

tial condition $q(t \to 0^+) = q^0 \in L^\infty(\mathbb{R})$. So we start from approximations $q_i^0 \approx \frac{1}{\Delta x_i}\int_{x_{i-1/2}}^{x_{i+1/2}} q^0(x)dx$, $i \in \mathbb{Z}$, and define the cell values q_i^n for each $n = 1, \ldots, N$ by a (two-step) time scheme:
(i) first, an approximate solution on $[t^n, t^{n+1})$ to the *homogeneous* gSV system (i.e. without the source term S) is computed by an explicit three-point scheme

$$q_i^{n+1/2} = q_i^n - \frac{\Delta t^n}{\Delta x_i}\left(F_l(q_i^n, q_{i+1}^n) - F_r(q_{i-1}^n, q_i^n)\right), \tag{11}$$

(ii) next, an approximate solution to the full gSV system on $(t^n, t^{n+1}]$ is computed

$$q_i^{n+1} = q_i^{n+1/2} + \Delta t^n S(q_i^{n+1}). \tag{12}$$

Then, we require the scheme

$$q_i^{n+1} = q_i^n - \frac{\Delta t^n}{\Delta x_i}\left(F_l(q_i^n, q_{i+1}^n) - F_r(q_{i-1}^n, q_i^n)\right) + \Delta t^n S(q_i^{n+1}) \tag{13}$$

to be consistent with the conservative formulation (6), (7), (8) and (9) in the sense of [5, 2.1], i.e. the numerical fluxes F_l, F_r are required to satisfy $F_{l,h} = F_{r,h} := F_h$, $F_{l,hu} = F_{r,hu} := F_{hu}$ …with $F_h(q,q) = hu|_q$, $F_{hu}(q,q) = (hu^2 + gh^2/2 + hN)|_q$ etc. To that aim, we shall define F_l and F_r using a *simple* approximate Riemann solver [11] for (6), (7), (8) and (9). Moreover, with a view to preserving $\mathscr{U}^\ell$ and a discrete version of (5)

$$F(q_i^{n+1/2}) - F(q_i^n) + \frac{\Delta t^n}{\Delta x_i}\left(G(q_i^n, q_{i+1}^n) - G(q_{i-1}^n, q_i^n))\right) \le 0 \tag{14}$$

for a numerical free-energy flux function consistent with $G(q,q) = u(F+P)|_q$ in (5), we shall discuss the relaxation technique introduced by Suliciu as simple Riemann solver in the sequel. For analogous systems equipped with an entropy that is convex in the discretization variable similarly to F, discretizations that satisfy an entropy inequality could be constructed with that Riemann solver in the past [4–6].

In the end, the scheme (13) computed from (11) and (12) satisfies:

Proposition 1 *If* (14) *holds, then a discrete free-energy dissipation holds*

$$F(q_i^{n+1}) - F(q_i^n) + \frac{\Delta t^n}{\Delta x_i}\left(G(q_i^n, q_{i+1}^n) - G(q_{i-1}^n, q_i^n))\right) \le \Delta t^n D(q_i^{n+1}). \tag{15}$$

Proof On noting $h_i^{n+1/2} = h_i^{n+1}$, $u_i^{n+1/2} = u_i^{n+1}$ it suffices to show that

$$\lambda\left(\sigma_{xx,i}^{n+1} - \sigma_{xx,i}^n\right)/\Delta t^n = 1 - \sigma_{xx,i}^{n+1}/(1 - (\sigma_{zz,i}^{n+1} + \sigma_{xx,i}^{n+1})/\ell)$$
$$\lambda\left(\sigma_{zz,i}^{n+1} - \sigma_{zz,i}^n\right)/\Delta t^n = 1 - \sigma_{zz,i}^{n+1}/(1 - (\sigma_{zz,i}^{n+1} + \sigma_{xx,i}^{n+1})/\ell)$$

imply $F(q_i^{n+1}) - F(q_i^{n+1/2}) \leq \Delta t^n D(q_i^{n+1}) \leq 0$. Now, this is a consequence of the convexity of $F|_{h,u}$ in $(\sigma_{xx}, \sigma_{zz})$ and $\nabla_{(\sigma_{xx},\sigma_{zz})} F|_{h,u} \cdot S = D$.

2.2 Suliciu Relaxation of the Riemann Problem Without Source

For all time ranges $t \in [t^n, t^{n+1})$, $n = 0 \ldots N-1$, we now define at each interface $x_{i+\frac{1}{2}}$, $i \in \mathbb{Z}$, between cells i and $i+1$ consistent numerical flux functions F_l and F_r

$$\begin{aligned} F_l(q_l, q_r) &= F_0(q_l) - \int_{-\infty}^{0} \Big(R(\xi, q_l, q_r) - q_l \Big) d\xi, \\ F_r(q_l, q_r) &= F_0(q_r) + \int_{0}^{\infty} \Big(R(\xi, q_l, q_r) - q_r \Big) d\xi, \end{aligned} \tag{16}$$

invoking an approximate solution $R\left((x - x_{i+1/2})/(t - t^n), q_i^n, q_{i+1}^n\right)$ to the Riemann problem for (10) with initial condition $q_i^n 1_{x<0} + 1_{x>0} q_{i+1}^n$ at $t = t^n$, and *any* F_0.

In this work, we propose as approximate solution that given by Suliciu relaxation

$$R(\xi, q_l, q_r) = L\mathscr{R}(\xi, \mathscr{Q}_l, \mathscr{Q}_r) \ , \tag{17}$$

i.e. the projection (operator L) onto $q \equiv (h, hu, h\sigma_{xx}, h\sigma_{zz})$ of the *exact* solution $\mathscr{R}(\xi, \mathscr{Q}_l, \mathscr{Q}_r)$ to the Riemann problem for a system with relaxed pressure

$$\left\{ \begin{aligned} \partial_t h + \partial_x(hu) &= 0 \\ \partial_t(hu) + \partial_x(hu^2 + \pi) &= 0 \\ \partial_t(\sigma_{xx} h^{2(1-\zeta)}) + u\partial_x(\sigma_{xx} h^{2(1-\zeta)}) &= 0 \\ \partial_t(\sigma_{zz} h^{2(\zeta-1)}) + u\partial_x(\sigma_{zz} h^{2(\zeta-1)}) &= 0 \\ \partial_t(h\pi) + \partial_x(hu\pi + uc^2) &= 0 \\ \partial_t\left(h(u^2/2 + \hat{e})\right) + \partial_x\left(hu(u^2/2 + \hat{e}) + u\pi\right) &= 0 \\ \partial_t c + u\partial_x c &= 0 \end{aligned} \right. \tag{18}$$

and initial condition given by $(o = l, r)$

$$\mathscr{Q}_o = \left(h_o, (hu)_o, h_o^{1-2\zeta}(h\sigma_{xx})_o, h_o^{2\zeta-3}(h\sigma_{zz})_o, h_o P(q_o), (hu)_o^2/2h_o + e(q_o), c_o\right), \tag{19}$$

where $c_o(q_l, q_r)$ are chosen so as to ensure stability, that is the dissipation rule (14) here (see below). The Riemann solver R is consistent under the CFL condition

$$\Delta t^n \leq \frac{1}{2} \inf_{i \in \mathbb{Z}} \frac{1}{\Delta x_i} \max\left(u_i^n - c_l(q_i^n, q_{i+1}^n)/h_i^n, u_i^n + c_r(q_i^n, q_{i+1}^n)/h_{i+1}^n\right) \ , \tag{20}$$

and has an analytic expression as a function of q_o and c_o since the hyperbolic system (18) fully decomposes into linearly degenerate eigenfields (see formulas in [5, 6]).

It remains to specify a choice of functions c_l, c_r preserving $\mathscr{U}^\ell$ and ensuring (14).

Although it is not clear whether our construction allows one to approximate solutions on any time ranges $t \in [0, T)$, since the series $\sum_n \Delta t^n$ may be bounded uniformly for all space-grid choice ($\sup_i |u_i^n|$ may grow unboundedly as $n \to \infty$), specifying such c_l, c_r defines a fully computable scheme. In particular, the nonlinear system (12) at step (ii) is quadratic, with at least one admissible solution for any Δt^n that is analytically computable.

Note however a difficulty here for FENE-P fluids with c_l, c_r. Suliciu relaxation approach (18) was retained at step (i) because the solver often allows one to preserve invariant domains like $\mathscr{U}^\ell$ *and* a dissipation rule (14) through well-chosen c_l, c_r, see e.g. [4–6]. Indeed, on noting the exact Riemann solution to (18), to get (14) on choosing $G(q_l, q_r) = u\Big(h\big(\frac{u^2}{2} + \hat{e}\big) + \pi\Big)|_{\mathscr{R}(0,q_l,q_r)}$, it is enough that $\forall q_l, q_r \in \mathscr{U}^\ell$

$$q_\xi := L\mathscr{R}(\xi, \mathscr{Q}_l, \mathscr{Q}_r) \in \mathscr{U}^\ell \text{ and } h_\xi^2 \partial_h|_{h^{2-2\zeta}\sigma_{xx}, h^{2\zeta-2}\sigma_{zz}} P(q_\xi) \leq c_\xi^2\,,\ \forall \xi \in \mathbb{R} \quad (21)$$

where $c_\xi = c_l(q_l, q_r)$ if $\xi < u^*$ and $c_\xi = c_r(q_l, q_r)$ if $\xi > u^*$ with $u^* := \dfrac{c_l u_l + \pi_l + c_r u_r - \pi_r}{c_l + c_r}$.

One can easily propose c_l, c_r satisfying the first condition in (21), i.e.

$$\frac{1}{h_l^*} = \frac{1}{h_l}\left(1 + \frac{c_r(u_r - u_l) + \pi_l - \pi_r}{(c_l/h_l)(c_l + c_r)}\right) > 0 \quad (22)$$

$$\frac{1}{h_r^*} = \frac{1}{h_r}\left(1 + \frac{c_l(u_r - u_l) + \pi_r - \pi_l}{(c_r/h_r)(c_l + c_r)}\right) > 0 \quad (23)$$

as usual for Saint-Venant systems, plus the admissibility conditions ($o = l/r$)

$$(h_o^*)^{2(1-\zeta)}(h_o)^{2(\zeta-1)}\sigma_{zz,o} + (h_o^*)^{2(\zeta-1)}(h_o)^{2(1-\zeta)}\sigma_{xx,o} < \ell \quad (24)$$

for any $\sigma_{zz,o}, \sigma_{xx,o} > 0$ satisfying $\sigma_{zz,o} + \sigma_{xx,o} < \ell$ (see below). But the second condition is usually treated for a monotone function $\phi_o : h \to h\sqrt{\partial_h|_{h_o^{2-2\zeta}\sigma_{xx,o}, h_o^{2\zeta-2}\sigma_{zz,o}} P}$. Unfortunately, a lengthy (but easy) computation shows that the latter is not monotone here, so the standard method to choose c_l, c_r a priori does not apply.

2.3 Choice of Relaxation Parameter

We treat the first part of (21) as usual and define $c_o = \max(h_o a_o, \tilde{c}_o)$, where $a_o := \sqrt{\partial_h P(q_o)}$ and $o = l/r$, such that the functions $\tilde{c}_o(q_l, q_r)$ ensure (22), (23) and (24).

First, let us inspect (22) and (23) classically following [8, Sect. 3.3]. Denoting $a_l Y_l = (u_l - u_r)_+ + \frac{(\pi_r - \pi_l)_+}{h_l a_l + h_r a_r} \geq 0$, $a_r Y_r = (u_l - u_r)_+ + \frac{(\pi_l - \pi_r)_+}{h_l a_l + h_r a_r} \geq 0$ so $\frac{1}{h_o^*} \geq \frac{1 - h_o a_o Y_o / c_o}{h_o}$, it then holds $(h_o^*)^{-1} \geq (h_o)^{-1} y_o > 0$ with $y_o := 1 - \frac{Y_o}{1 + \alpha_o Y_o} \in (\frac{\alpha_o - 1}{\alpha_o}, 1]$ provided one chooses $\tilde{c}_o > 0$ such that $c_o \geq h_o a_o (1 + \alpha_o Y_o)$ for $\alpha_o > 1$. This yields $h_o^* \in (0, h_o / y_o]$ and thus (22) and (23) in particular.

On the other hand, let us now inspect (24), which rewrites with $h_o^* > 0$

$$w_o A_o + w_o^{-1} B_o < 1 \Leftrightarrow 2 A_o w_o \in \left(1 - \sqrt{1 - 4 A_o B_o}, 1 + \sqrt{1 - 4 A_o B_o}\right) \subset \mathbb{R}_{>0}. \tag{25}$$

Here $w_o = (h_o^* / h_o)^{2(1-\zeta)}$, $A_o = \sigma_{zz,o} / \ell$, $B_o = \sigma_{xx,o} / \ell$ are positive such that $A_o + B_o < 1$ (hence $A_o B_o \leq A_o (1 - A_o) \leq \frac{1}{4}$) and $2(1 - \zeta) \in [1, 2]$. The upper-bound in (25) is satisfied with $\alpha_o = (w_o^+)^{\frac{1}{2(1-\zeta)}} / ((w_o^+)^{\frac{1}{2(1-\zeta)}} - 1) > 1$, on noting

$$(w_o^+)^{\frac{1}{2(1-\zeta)}} := \left((1 + \sqrt{1 - 4 A_o B_o})/(2 A_o)\right)^{\frac{1}{2(1-\zeta)}} \geq \frac{\alpha_o}{\alpha_o - 1} \geq 1 / y_o \geq h_o^* / h_o \,. \tag{26}$$

It remains to ensure the lower bound in (25). Obviously, $w_o^- := \frac{1 - \sqrt{1 - 4 A_o B_o}}{2 A_o} < 1$ so one only needs to inspect the case $h_o^* \leq h_o$. Now, with $a_l W_l = (u_r - u_l)_+ + \frac{(\pi_l - \pi_r)_+}{h_l a_l + h_r a_r} \geq 0$, $a_r W_r = (u_r - u_l)_+ + \frac{(\pi_r - \pi_l)_+}{h_l a_l + h_r a_r} \geq 0$, if $c_o \geq h_o a_o W_o ((w_o^-)^{-\frac{1}{2(1-\zeta)}} - 1)^{-1}$ then holds

$$(w_o^-)^{\frac{1}{2(1-\zeta)}} \leq (1 + a_o h_o W_o / c_o)^{-1} \leq h_o^* / h_o \,.$$

At the end, we claim the following choices

$$c_l = h_l \max\left(a_l + \alpha_l \left((u_l - u_r)_+ + \frac{(\pi_r - \pi_l)_+}{h_l a_l + h_r a_r}\right), \beta_l \left((u_r - u_l)_+ + \frac{(\pi_l - \pi_r)_+}{h_l a_l + h_r a_r}\right)\right) \tag{27}$$

$$c_r = h_r \max\left(a_r + \alpha_r \left((u_l - u_r)_+ + \frac{(\pi_l - \pi_r)_+}{h_l a_l + h_r a_r}\right), \beta_r \left((u_r - u_l)_+ + \frac{(\pi_r - \pi_l)_+}{h_l a_l + h_r a_r}\right)\right) \tag{28}$$

satisfy simultaneously (22), (23) and (24) in a compatible way with $a_o = \sqrt{\partial_h P(q_o)}$, $\alpha_o = \max(2, (w_o^+)^{\frac{1}{2(1-\zeta)}} / ((w_o^+)^{\frac{1}{2(1-\zeta)}} - 1))$, $\beta_o = (w_o^-)^{\frac{1}{2(1-\zeta)}} / (1 - (w_o^-)^{\frac{1}{2(1-\zeta)}}))$, $w_o^- = \frac{\ell - \sqrt{\ell - 4\sigma_{zz,o}\sigma_{xx,o}}}{2\sigma_{zz,o}}$, $w_o^+ = \frac{\ell + \sqrt{\ell - 4\sigma_{zz,o}\sigma_{xx,o}}}{2\sigma_{zz,o}}$, for $o = l/r$. Moreover, note that we have chosen α_o such that all subcharacteristic conditions (21) are satisfied in the $\ell \to \infty$ limit, hence also the free-energy dissipation (15). Indeed, ϕ_o is monotone in the $\ell \to \infty$ limit and one can then apply the standard method to choose c_l, c_r [6].

3 Conclusion

We have proposed a fully computable FV discretization for a 1D nonlinear hyperbolic system modelling viscoelastic shallow flows with a new ingredient in comparison with [6]: a (physical) bound on the elastic behaviour $\sigma_{xx} + \sigma_{zz}$.

But the free-energy dissipation criterium proposed herein, for consistency of approximating sequences with admissible solutions, cannot be ensured a priori by classical relaxation techniques.

In fact, numerical experiments even seem to indicate that the dissipation may be difficult to satisfy for Riemann problems with sufficiently large initial data, small ℓ and fine meshes, even when the scheme does not blow up.

A careful analysis of such specific situations may help understand the famous High-Weissenberg problem, well-known in the field of computational rheology [1, 13]. This remains to be done in the future.

References

1. Barrett, J.W., Boyaval, S.: Existence and approximation of a (regularized) Oldroyd-B model. M3AS **21**(9), 1783–1837 (2011)
2. Berthon, C., Coquel, F., LeFloch, P.G.: Why many theories of shock waves are necessary: kinetic relations for non-conservative systems. Proc. R. Soc. Edinb. Sect. A Math. **142**, 1–37 (2012)
3. Bird, R., Dotson, P., Johnson, N.: Polymer solution rheology based on a finitely extensible beadspring chain model. J. Non Newton. Fluid Mech. **7**, 213–235 (1980)
4. Bouchut, F.: Entropy satisfying flux vector splittings and kinetic BGK models. Numer. Math. **94**, 623–672 (2003)
5. Bouchut, F.: Nonlinear stability of finite volume methods for hyperbolic conservation laws and well-balanced schemes for sources. Frontiers in Mathematics. Birkhäuser Verlag, Basel (2004)
6. Bouchut, F., Boyaval, S.: A new model for shallow viscoelastic fluids. M3AS **23**(08), 1479–1526 (2013)
7. Bouchut, F., Boyaval, S.: Unified derivation of thin-layer reduced models for shallow free-surface gravity flows of viscous fluids. Eur. J. Mech. B Fluids Part 1 **55**, 116–131 (2016)
8. Bouchut, F., Klingenberg, C., Waagan, K.: A multiwave approximate Riemann solver for ideal MHD based on relaxation II: numerical implementation with 3 and 5 waves. Numer. Math. **115**(4), 647–679 (2010)
9. Boyaval, S.: Johnson-Segalman – Saint-Venant equations for viscoelastic shallow flows in the pure elastic limit. Proceedings in Mathematics and Statistics (PROMS) of the *International Conference on Hyperbolic Problems: Theory, Numeric and Applications in Aachen* 2016, (2017), ArXiv e-prints (2016)
10. Dafermos, C.M.: Hyperbolic conservation laws in continuum physics, vol. GM 325. Springer, Berlin (2000)
11. Harten, A., Lax, P.D., van Leer, B.: On upstream differencing and Godunov-type schemes for hyperbolic conservation laws. SIAM Rev. **25**(1), 35–61 (1983)
12. LeFloch, P.G.: Hyperbolic Systems of Conservation Laws: The Theory of Classical and Non-classical Shock Waves. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel (2002)
13. Owens, R.G., Philips, T.N.: Computational Rheology Imperial. College Press/World Scientific, London (2002)

Palindromic Discontinuous Galerkin Method

David Coulette, Emmanuel Franck, Philippe Helluy, Michel Mehrenberger and Laurent Navoret

Abstract We present a high-order scheme for approximating kinetic equations with stiff relaxation. The construction is based on a high-order, implicit, upwind Discontinuous Galerkin formulation of the transport equations. In practice, because of the triangular structure of the implicit system, the computations are explicit. High order in time is achieved thanks to a palindromic composition method. The whole method is asymptotic-preserving with respect to the stiff relaxation and remains stable even with large CFL numbers.

Keywords Lattice boltzmann · Discontinuous galerkin · Implicit · Composition method · High order · Stiff relaxation.

MSC (2010): 65L04 · 65M99

1 Introduction

The Lattice Boltzmann Method (LBM) is a general method for solving systems of conservation laws [5]. The LBM relies on a kinetic representation of the system of conservation laws by a small set of transport equations coupled through a stiff relaxation source term. The kinetic model is solved with a splitting method, in which the transport and relaxation steps are treated separately. Usually, the transport is exactly solved by the characteristic method.

The main drawback of the LBM is that it requires regular grids and that the time step Δt is imposed by the grid step Δx. In this paper, we thus prefer to solve the transport equation with a Discontinuous Galerkin (DG) method. We extend the DGLBM [12] in several directions. The first improvement is to apply an implicit DG method instead of an explicit one for solving the transport equations. This can be done at almost no additional cost. Indeed, with an upwind numerical flux, the

D. Coulette · E. Franck · P. Helluy (✉) · M. Mehrenberger · L. Navoret
IRMA Strasbourg, Inria Tonus, Strasbourg, France
e-mail: helluy@unistra.fr

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_19

linear system of the implicit DG method is triangular and, in the end, can be solved explicitly. In this way, we obtain stable methods even with high CFL numbers. This kind of ideas can be found for instance in [3, 6].

The second improvement is to construct a high order time integrator that remains accurate even for infinitely fast relaxation, thanks to a composition method [11].

We validate our approach on a few one-dimensional test cases.

2 A Vectorial Kinetic Model

We consider the following kinetic equation

$$\partial_t \mathbf{f} + \sum_{k=1}^{D} \mathbf{V}^k \partial_k \mathbf{f} = \frac{1}{\tau}(\mathbf{f}^{eq}(\mathbf{f}) - \mathbf{f}). \tag{1}$$

The unknown is a vectorial distribution function $\mathbf{f}(\mathbf{x}, t) \in \mathbb{R}^n$ depending on the space variable $\mathbf{x} \in \mathbb{R}^D$ and time $t \in \mathbb{R}$. The relaxation time τ is a small positive constant. The constant matrices $\mathbf{V}^k$, $1 \leq k \leq d$ are diagonal. In other words (1) is a set of transport equations at constant velocities coupled through a stiff BGK relaxation.

Generally this kinetic model is an approximation of an underlying hyperbolic system of conservation laws. The macroscopic conservative variables $\mathbf{w}(\mathbf{x}, t) \in \mathbb{R}^m$ are obtained through a linear transformation

$$\mathbf{w} = \mathbf{P}\mathbf{f}$$

where $\mathbf{P}$ is a $m \times n$ constant matrix. Generally the number of conservative variables is smaller than the number of kinetic data: $m < n$. The equilibrium distribution $\mathbf{f}^{eq}(\mathbf{f})$ is such that

$$\mathbf{P}\mathbf{f} = \mathbf{P}\mathbf{f}^{eq}(\mathbf{f}),$$

and

$$\mathbf{w} = \mathbf{P}\mathbf{f} = \mathbf{P}\mathbf{g} \Rightarrow \mathbf{f}^{eq}(\mathbf{f}) = \mathbf{f}^{eq}(\mathbf{g}). \tag{2}$$

When $\tau \to 0$, the kinetic equations provide an approximation of the system of conservation laws

$$\partial_t \mathbf{w} + \sum_{k=1}^{D} \partial_k \mathbf{q}^k(\mathbf{w}) = 0,$$

where the flux is given by

$$\mathbf{q}^k(\mathbf{w}) = \mathbf{P}\mathbf{V}^k\mathbf{f}^{eq}(\mathbf{f}), \quad \mathbf{w} = \mathbf{P}\mathbf{f}.$$

The flux is indeed a function of $\mathbf{w}$ only because of (2). For more details, we refer to [1, 4, 8]. In the following, without loss of generality, we shall only consider the one-dimensional case $D = 1$.

3 Implicit Discontinous Galerkin (DG) Method

In this section, we briefly recall how to approximate the transport equation by a DG method of order d, based on Lagrange polynomials. We wish to approximate the solution $f(x,t)$ of a scalar transport equation in the case $v > 0$ (the case $v < 0$ is obtained in a similar way)

$$\partial_t f + v\partial_x f = 0.$$

The space variable $x \in [a,b]$. We split the interval $[a,b]$ into N_x cells of size $h = (b-a)/N_x$. In each cell C_ℓ, $\ell = 0 \dots N_x - 1$, we consider the $d+1$ Gauss–Lobatto points $x_{\ell,i}$, $i = 0 \dots d$, associated to quadrature weights $\omega_{\ell,i}$. The DG basis function $\varphi_{\ell,i}$ has its support in cell ℓ and in addition satisfies the interpolation property $\varphi_{\ell,i}(x_{\ell,j}) = \delta_{ij}$. The transported function $f(x,t)$ is then approximated by an expansion on the DG basis

$$f(x,t) \simeq f_h(x,t) = \sum_{j=0}^{d} f_{\ell,j}(t)\varphi_{\ell,j}(x), \quad x \in C_\ell.$$

We can also identify f_h with a vector of $\mathbb{R}^N$, $N = N_x(d+1)+2$, $\mathbf{f}_h = (f_0(t), f_{0,0}(t), f_{0,1}(t), \dots, f_{N_x-1,d}(t), f_{N-1}(t))$. It is useful to consider the boundary values $f_0 = f_{-1,d}$ and $f_{N_x,0} = f_{N-1}$ as artificial unknowns. For simplicity, we will also assume that the boundary conditions do not depend on time. Now we apply the DG formulation to f_h: for all cell C_ℓ and all test function $\varphi_{\ell,i}$

$$\int_{C_\ell} (\partial_t f_h + v\partial_x f_h)\varphi_{\ell,i} + v\left(f_{\ell,0} - f_{\ell-1,d}\right)\varphi_{\ell,i}(x_{\ell,0}) = 0. \tag{3}$$

Let us point out the upwind nature of the formulation (3): when $v > 0$, for computing the values inside cell C_ℓ we only need the knowledge of the values inside cell $C_{\ell-1}$, or the left boundary condition. Therefore, after applying a Gauss-Lobatto quadrature to (3), we obtain a set of linear differential equations

$$\partial_t \mathbf{f}_h + \mathbf{L}_h \mathbf{f}_h = 0, \tag{4}$$

where $\mathbf{L}_h$ is a lower block-triangular matrix (with a good numbering of the unknowns). The diagonal blocks are of size $(d+1) \times (d+1)$. If the velocity $v < 0$, the structure is similar. Therefore, in the following, we adopt the same notation $\mathbf{L}_h$ for the scalar or vectorial DG transport operator.

4 High Order Time Integration

We can also define an approximation $\mathbf{N}_h$ of the collision operator $\mathbf{N}$ defined by $\mathbf{Nf} = (\mathbf{f}^{eq}(\mathbf{f}) - \mathbf{f})/\tau$. The DGLBM approximation of (1) finally reads

$$\partial_t \mathbf{f}_h = \mathbf{L}_h \mathbf{f}_h + \mathbf{N}_h \mathbf{f}_h.$$

We use the following Crank–Nicolson second order time integrator for the transport equation:

$$\exp(\Delta t \mathbf{L}_h) \simeq T_2(\Delta t) := (\mathbf{I} + \frac{\Delta t}{2}\mathbf{L}_h)(\mathbf{I} - \frac{\Delta t}{2}\mathbf{L}_h)^{-1}. \tag{5}$$

Similarly, for the collision integrator, we use

$$\exp(\Delta t \mathbf{N}_h)\mathbf{f}_h \simeq C_2(\Delta t)\mathbf{f}_h = \frac{(2\tau - \Delta t)\mathbf{f}_h}{2\tau + \Delta t} + \frac{2\Delta t \mathbf{f}_h^{eq}(\mathbf{f}_h)}{2\tau + \Delta t}. \tag{6}$$

When $\tau \to 0$, it becomes

$$C_2(\Delta t)\mathbf{f}_h = 2\mathbf{f}_h^{eq}(\mathbf{f}_h) - \mathbf{f}_h. \tag{7}$$

An integrator $M_2(\Delta t)$ is *time-symmetric* if it satisfies

$$M_2(-\Delta t) = M_2(\Delta t)^{-1}, \quad M_2(0) = Id. \tag{8}$$

This property implies that M_2 is necessarily a second order approximation of the exact integrator [9, 11]. As explained in [7], T_2 and C_2 are time-symmetric when $\tau > 0$. But, because of (7), C_2 is no more symmetric for $\tau = 0$. Therefore, the usual Strang splitting operator is not time-symmetric either. We rather consider the following splitting method, which is time-symmetric and remains second order accurate even for infinitely fast relaxation:

$$M_2(\Delta t) = T_2(\frac{\Delta t}{4})C_2(\frac{\Delta t}{2})T_2(\frac{\Delta t}{2})C_2(\frac{\Delta t}{2})T_2(\frac{\Delta t}{4}).$$

By palindromic compositions of the second order method M_2 it is then very easy to achieve any even order of accuracy. See [7, 9, 11]. A general palindromic scheme with $s + 1$ steps has the form

$$M_p(\Delta t) = M_2(\gamma_0 \Delta t) M_2(\gamma_1 \Delta t) \dots M_2(\gamma_s \Delta t), \tag{9}$$

where the γ_i's are real numbers such that

$$\gamma_i = \gamma_{s-i}, \quad 0 \le i \le s.$$

For $p = 4$ and $s = 4$ we have for example the fourth-order Suzuki scheme (see [9, 11, 13])

$$\gamma_0 = \gamma_1 = \gamma_3 = \gamma_4 = \frac{1}{4 - 4^{1/3}}, \quad \gamma_2 = -\frac{4^{1/3}}{4 - 4^{1/3}}. \tag{10}$$

This scheme requires five stages and one negative time step. For $p = 6$ and $s = 8$, we have also the sixth-order Kahan-Li scheme [10] given by:

$$\begin{aligned}
\gamma_0 = \gamma_8 &= 0.39216144400731413927565533003 8\ldots \\
\gamma_1 = \gamma_7 &= 0.33259913678935943860427212532 5\ldots \\
\gamma_2 = \gamma_6 &= -0.70624617255763935980984533722 27\ldots \\
\gamma_3 = \gamma_5 &= 0.08221359629355080023044270533 41\ldots \\
\gamma_4 &= 0.79854399093482996339895035304 8\ldots
\end{aligned} \tag{11}$$

The methods (10) and (11) require to apply the elementary collision or transport bricks C_2 and T_2 with negative time steps $-\Delta t < 0$.

The exact transport operator is perfectly reversible. If we were using an exact characteristic solver, negative time steps would not cause any problem. However, the DG approximation introduces a slight dissipation due to upwinding. In order to ensure stability, we have thus to replace $T_2(-\Delta t)$ with a more stable operator. This can be done by observing that solving $\partial_t f + v\partial_x f = 0$ for negative time $t < 0$ is equivalent to solve $\partial_{t'} f - v\partial_x f = 0$ for $t' = -t > 0$. Therefore we simply apply the DG solver $T_2'(\Delta t)$ with opposite velocities instead of $T_2(-\Delta t)$.

The numerical collision operator C_2 is reversible when $\tau \to 0$. Actually, it does not depend on Δt anymore (see (7)). In this stage, negative time steps do not cause any difficulty, at least when $\tau \ll \Delta t$.

5 Numerical Results

In this section we consider an isothermal compressible flow of density ρ and velocity u. The sound speed is fixed to $c = 0.6$. The conservative system is given by $m = 2$ and $w = (\rho, \rho u)$, $q(w) = (\rho u, \rho u^2 + c^2\rho)$. The kinetic model is given by $n = 4$ and

$$\mathbf{V} = \mathrm{diag}(-\lambda, \lambda, -\lambda, \lambda), \quad \mathbf{P} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix},$$

$$f^{eq}_{2k-1} = \frac{w_k}{2} - \frac{q(w)_k}{2\lambda}, \quad f^{eq}_{2k} = \frac{w_k}{2} + \frac{q(w)_k}{2\lambda}, \quad k = 1, 2.$$

The lattice velocity λ has to satisfy the sub-characteristic condition $\lambda > |u| + c$.

5.1 Smooth Solution

For the first validation of the method we consider a test case with a smooth solution, in the fluid limit $\tau = 0$. The initial condition is given by

$$\rho(x, 0) = 1 + e^{-30x^2}, \quad u(x, 0) = 0.$$

The sound speed is set to $c = 0.6$ and the lattice velocity to $\lambda = 2$. We define the CFL number $\beta = \lambda \Delta t / \delta$, where δ is the minimal distance between two Gauss-Lobatto points in the mesh. First, the CFL number is fixed to $\beta = 5$. We consider a sufficiently large computational domain $[a, b] = [-2, 2]$ and a sufficiently short final time $t_{\max} = 0.4$ so that the boundary conditions play no role. The reference solution $\mathbf{f}(\cdot, t_{\max})$ is computed numerically with a very fine mesh. In the DG solver the polynomial order in x is fixed to $d = 5$.

On Fig. 1 (left picture) we give the results of the convergence study for the smooth solution. We consider the L^2 error.

We make the same experiment with $\beta = 50$. The convergence study for the Suzuki and Kahan-Li schemes is also presented on Fig. 1 (right picture). At high CFL, not only the scheme remains stable, but the high accuracy is also preserved.

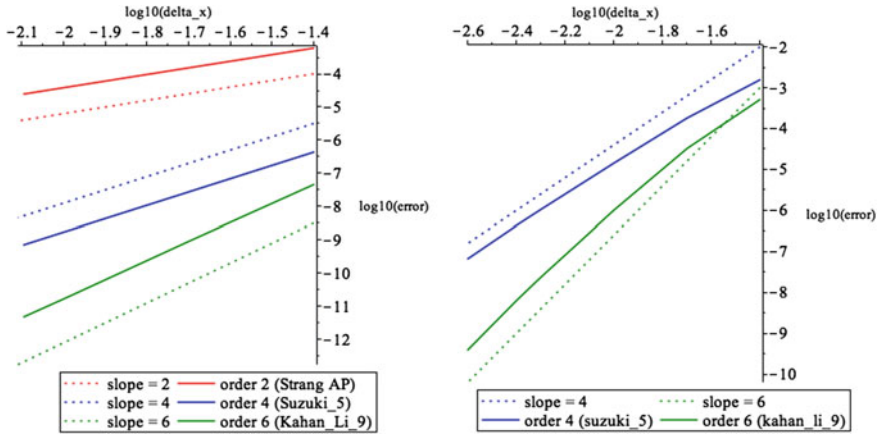


Fig. 1 Convergence study for several palindromic methods, order 2 (*red*), 4 (*blue*) and 6 (*green*). The *dotted lines* are reference lines with slopes 2, 4 and 6 respectively. *Left* CFL number $\beta = 5$. *Right* CFL number $\beta = 50$

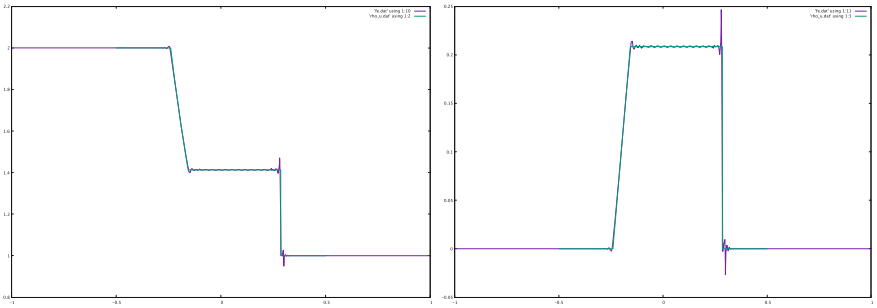


Fig. 2 Riemann problem with $\tau = 0$. Comparison of the exact solution (*green curve*), and the numerical sixth-order solution (*purple curve*). *Left* density. *Right* velocity

5.2 Behavior for Discontinuous Solutions

We have also experimented the scheme for discontinuous solutions. Of course, in this case the effective order of the method cannot be higher than one and we expect Gibbs oscillations near the discontinuities. On the interval $[a, b] = [-1, 1]$, we consider a Riemann problem with the following initial condition

$$\rho(x, 0) = \begin{cases} 2 & \text{if } x < 0, \\ 1 & \text{otherwise.} \end{cases}, \quad u(x, 0) = 0.$$

We consider numerical results in the fluid limit $\tau = 0$. On Fig. 2 we compare the sixth-order numerical solution with the exact one at $t = t_{\max} = 0.4$ for a CFL number $\beta = 3$ and $N_x = 100$ cells. We observe that the high order scheme is able to capture a precise rarefaction wave and the correct position of the shock wave. We observe oscillations emanating from the points where the solution is not smooth (shocks and boundaries of the rarefaction wave), as expected.

6 Conclusion

In this paper we have described a new numerical method, the Palindromic Discontinuous Galerkin Method, for solving kinetic equations with stiff relaxation. The method has the following features:

- The transport solver is based on an implicit, high-order, upwind Discontinuous Galerkin method. Thanks to the upwind flux, the linear system to be solved at each time step is triangular.

- Time integration is high-order, based on a general palindromic composition method. We have tested it up to order 6. The method is low-storage.
- The scheme remains stable and accurate at high CFL numbers and for infinitely fast relaxation.

We are currently working on the extension of the method to higher dimensions and to optimizations of the implementation on hybrid computers (preliminary results can be found in [2]). We are also studying the possibility of more general boundary conditions. For practical applications it will be important to add to the method a limiter strategy for controlling oscillations in shock waves.

References

1. Aregba-Driollet, D., Natalini, R.: Discrete kinetic schemes for multidimensional systems of conservation laws. SIAM J. Numer. Anal. **37**(6), 1973–2004 (2000)
2. Badwaik, J., Boileau, M., Coulette, D., Franck, E., Helluy, P., Mendoza, L., Oberlin, H.: Task-based parallelization of an implicit kinetic scheme. arXiv preprint arXiv:1702.00169 (2017)
3. Bey, J., Wittum, G.: Downwind numbering: robust multigrid for convection-diffusion problems. Appl. Numer. Math. **23**(1), 177–192 (1997)
4. Bouchut, F.: Construction of BGK models with a family of kinetic entropies for a given system of conservation laws. J. Stat. Phys. **95**(1–2), 113–170 (1999)
5. Chen, S., Doolen, G.D.: Lattice Boltzmann method for fluid flows. Annu. Rev. Fluid Mech. **30**(1), 329–364 (1998)
6. Coquel, F., Nguyen, Q.L., Postel, M., Tran, Q.H.: Large time step positivity-preserving method for multiphase flows. In: Hyperbolic Problems: Theory, Numerics, Applications, pp. 849–856. Springer, Heidelberg (2008)
7. Coulette, D., Franck, E., Helluy, P., Mehrenberger, M., Navoret, L.: Palindromic discontinuous Galerkin method for kinetic equations with stiff relaxation. arXiv preprint arXiv:1612.09422 (2016)
8. Graille, B.: Approximation of mono-dimensional hyperbolic systems: a lattice Boltzmann scheme as a relaxation method. J. Comput. Phy. **266**, 74–88 (2014)
9. Hairer, E., Lubich, C., Wanner, G.: Geometric numerical integration: structure-preserving algorithms for ordinary differential equations, vol. 31. Springer Science & Business Media (2006)
10. Kahan, W., Li, R.C.: Composition constants for raising the orders of unconventional schemes for ordinary differential equations. Math. Comput. Am. Math. Soc. **66**(219), 1089–1099 (1997)
11. McLachlan, R.I., Quispel, G.R.W.: Splitting methods. Acta Numer. **11**, 341–434 (2002)
12. Shi, X., Lin, J., Yu, Z.: Discontinuous Galerkin spectral element lattice Boltzmann method on triangular element. Int. J. Numer. Methods Fluids **42**(11), 1249–1261 (2003)
13. Suzuki, M.: Fractal decomposition of exponential operators with applications to many-body theories and Monte Carlo simulations. Phys. Lett. A **146**(6), 319–323 (1990)

IMEX Finite Volume Methods for Cloud Simulation

M. Lukáčová-Medvid'ová, J. Rosemeier, P. Spichtinger and B. Wiebe

Abstract We present new implicit-explicit (IMEX) finite volume schemes for numerical simulation of cloud dynamics. We use weakly compressible equations to describe fluid dynamics and a system of advection-diffusion-reaction equations to model cloud dynamics. In order to efficiently resolve slow dynamics we split the whole nonlinear system in a stiff linear part governing the acoustic and gravitational waves as well as diffusive effects and a non-stiff nonlinear part that models nonlinear advection effects. We use a stiffly accurate second order IMEX scheme for time discretization to approximate the stiff linear operator implicitly and the non-stiff nonlinear operator explicitly. Fast microscale cloud physics is approximated by small scale subtractions.

Keywords Weakly compressible flows · Euler equations · Navier-Stokes equations · Low mach number · IMEX schemes · Cloud physics · Multiphase system

MSC (2010): 65M08 · 65N08 · 35Q30 · 35L65

M. Lukáčová-Medvid'ová · B. Wiebe (✉)
Institute of Mathematics, Johannes Gutenberg-University Mainz, Staudingerweg 9,
55 128 Mainz, Germany
e-mail: b.wiebe@uni-mainz.de

M. Lukáčová-Medvid'ová
e-mail: lukacova@uni-mainz.de

J. Rosemeier · P. Spichtinger
Institute of Atmospheric Physics, Johannes Gutenberg-University Mainz, Becherweg 21,
55 128 Mainz, Germany
e-mail: rosemeie@uni-mainz.de

P. Spichtinger
e-mail: spichtin@uni-mainz.de

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_20

1 Mathematical Model

In this paper we present a new operator splitting finite volume method for weakly compressible flows including cloud dynamics. The mathematical model consists of the Navier–Stokes equations describing weakly compressible fluid flow including viscous and friction effects. Further atmospheric factors like the Coriolis force and turbulence are not considered in this paper. In order to model microscale cloud physics we add evolution equations for water vapor, cloud water and rain. Phase change between these phases is modeled by an advection-diffusion-reaction system. Note that the total mass of the dry air remains constant, whereas the momentum and energy are not conserved, but satisfy the balance laws.

Let $\bar{p}$, $\bar{\rho}$, $\bar{\mathbf{u}}(=0)$, $\bar{\theta}$, $\overline{\rho\theta}$ express the pressure, density, velocity, potential temperature and energy for a dry background state, which is in the hydrostatic equilibrium

$$\partial_{x_3}\bar{p} = -\bar{\rho}g,$$

where $g = 9.81\,(\mathrm{m/s^2})$ is the gravitational acceleration. Furthermore let p', ρ', $\mathbf{u}'$, θ', $(\rho\theta)'$ stand for the corresponding perturbations of the background states. Thus, we have $p = \bar{p} + p'$, $\rho = \bar{\rho} + \rho'$, $\theta = \bar{\theta} + \theta'$, and $(\rho\theta) = \bar{\rho}\bar{\theta} + \bar{\rho}\theta' + \rho'\bar{\theta} + \rho'\theta' \equiv \overline{\rho\theta} + (\rho\theta)'$. Since the background velocity $\bar{\mathbf{u}} = 0$, it holds $\mathbf{u} \equiv \mathbf{u}'$ and we will omit the prime symbol hereinafter.

In order to avoid numerical instabilities due to the multiscale flow behavior in the case of low Mach number limit, numerical simulations are typically realized for the perturbations, which satisfy the following equations

$$\begin{aligned}
&\partial_t \rho' + \nabla\cdot(\rho\mathbf{u}) = 0,\\
&\partial_t(\rho\mathbf{u}) + \nabla\cdot\left(\rho\mathbf{u}\otimes\mathbf{u} + p'\,\mathrm{Id} - \rho\mu_m\left(\nabla\mathbf{u} + (\nabla\mathbf{u})^T\right)\right) = -\rho' g\mathbf{e}_3 \equiv -\rho' g\begin{pmatrix}0\\0\\1\end{pmatrix},\\
&\partial_t(\rho\theta)' + \nabla\cdot(\rho\theta\mathbf{u} - \rho\mu_h\nabla\theta) = S_\theta,
\end{aligned} \tag{1}$$

where μ_m, μ_h are viscosity and heat conductivity constants. To include the moist dynamics we use in $(1)_3$ instead of the potential temperature for dry air the moist potential temperature. Denoting T the temperature, the moist potential temperature can be approximated as

$$\theta = \frac{R_m}{R}T\left(\frac{p_0}{p}\right)^{R_m/c_p},$$

where $p_0 = 10^5\,(\mathrm{Pa})$ is the reference pressure, $R = 287.05\,(\mathrm{J/(kg\cdot K)})$ is the gas constant of dry air, $R_m = (1 - q_v - q_c - q_r)R + q_v R_v$ is the modified gas constant of moist air and $c_p = 1005\,(\mathrm{J/(kg\cdot K)})$, $R_v = 461.51\,(\mathrm{J/(kg\cdot K)})$. The mass fractions of water vapor, cloud water and rain are denoted by q_v, q_c and q_r, respectively; their evolution equations will be specified below, cf. (2).

In order to close the system we determine pressure from the state equation including moisture $p = p_0 \left(\frac{R\rho\theta}{p_0}\right)^{\gamma_m}$, $\gamma_m = \frac{c_p}{c_p - R_m}$. The source term S_θ that is related to the cloud microphysics for moist processes, cf. (2), expresses the released or absorbed latent heat. For dry case R_m reduces to R and $S_\theta = 0$.

For representing (liquid) clouds in models, different approaches are described in literature, see, e.g., [4, 6] and references therein. In the present study we use a so-called single moment scheme, i.e., evolution equations for mass concentrations of water vapor q_v, cloud water q_c and rain q_r are coupled to the system of Eq. (1). The formulation of cloud models is not possible from first principles, since some approximations and fits to experimental data must be used in order to formulate the equations for mass concentrations only. We chose a consistent approach for modeling the process rates in the cloud model, combining existing approaches in a meaningful way, see, e.g., [9]. The microphysical cloud processes of warm clouds are modeled by the following advection-diffusion-reaction system

$$\begin{aligned} \partial_t(\rho q_v) + \nabla \cdot (\rho q_v \mathbf{u} - \rho \mu_h \nabla q_v) &= -C + E, \\ \partial_t(\rho q_c) + \nabla \cdot (\rho q_c \mathbf{u} - \rho \mu_h \nabla q_c) &= C - A_1 - A_2, \\ \partial_t(\rho q_r) + \nabla \cdot \left(\rho q_r \mathbf{u} - \rho v_q \mathbf{e_3} q_r - \rho \mu_h \nabla q_r\right) &= A_1 + A_2 - E. \end{aligned} \tag{2}$$

The term $\nabla \cdot \left(-\rho v_q \mathbf{e_3} q_r\right)$, where $v_q \sim q_r^{1/5}$ is the raindrop fall velocity, represents the sedimentation of rain water. C, E, A_1, A_2 denote rates of condensation and evaporation (phase transition vapor/water) and collision rates. We assume that cloud water does not sediment by gravity, whereas rain water falls downwards. Thus, we introduce autoconversion A_1 for colliding cloud drops forming rain and accretion A_2 for rain droplets growing by collecting cloud water. Thermodynamic equilibrium is determined by saturation mixing ratio $q_* = q_*(p, T)$. Thus, the source terms can be formulated as follows:

$$C \sim q_c(q_v - q_*), \quad E \sim q_r^{1/2}(q_* - q_v), \quad A_1 \sim q_c^2, \quad A_2 \sim q_c q_r^{19/20}.$$

Note that in general cloud physics parametrisations show stiff behavior. The stiffness is a result of modeling processes with power laws containing exponents α with $0 < \alpha < 1$. To close the coupled model (1), (2) we express the potential temperature source term as

$$S_\theta = \rho \frac{L}{c_p} \frac{\theta}{T} (C - E)$$

with the specific latent heat of vaporization $L = 2.53 \cdot 10^6$ (J/kg). Note that we formulate condensation and evaporation processes explicitly, in contrast to the usual approach of saturation adjustment, see, e.g., [7], which is commonly used in operational weather forecast models. The explicit formulation introduces additional stiff terms and very small time scales.

2 Numerical Scheme

The numerical approximation of the coupled model (1), (2) is realized by the operator splitting approach. We split the whole system into the macroscopic flow equations and microscopic cloud model. The macroscopic model is approximated by the IMEX finite volume scheme. On the other hand the cloud equations are approximated using several subiterations of the same implicit-explicit scheme using smaller time steps. The macroscopic flow equations use the solution of the microscopic cloud model at the last time step and the cloud model uses the solution of the flow equations from the new time step. This yields to a first order splitting. In order to increase the accuracy the second order Strang splitting can be used.

2.1 IMEX Finite Volume Scheme for the Navier–Stokes Equations

In order to take into account multiscale behavior of the solution and to derive an asymptotically stable and accurate scheme, we propose the following splitting of the Navier–Stokes equations into a linear $\mathscr{L}$ and a nonlinear $\mathscr{N}$ part, see also [2] and the references therein. To this end let us rewrite (1) in the following compact form. Let $\mathbf{w} = (\rho', \rho\mathbf{u}, \rho\theta')^T$, $\mathbf{F}(\mathbf{w}) = (\rho\mathbf{u}, \rho\mathbf{u} \otimes \mathbf{u} + p' \,\mathrm{Id}, \rho\theta\mathbf{u})^T$, $\mathbf{D}(\mathbf{w}) = \left(0, \nabla \cdot \left(\rho\mu_m \left(\nabla\mathbf{u} + (\nabla\mathbf{u})^T\right)\right), \nabla \cdot (\rho\mu_h \nabla\theta)\right)^T$ and $\mathbf{S}(\mathbf{w}) = (0, -\rho' g\mathbf{e}_3, S_\theta)^T$, then (1) can be equivalently written as

$$\frac{\partial \mathbf{w}}{\partial t} = -\nabla \cdot \mathbf{F}(\mathbf{w}) + \mathbf{D}(\mathbf{w}) + \mathbf{S}(\mathbf{w}) \equiv \mathscr{L}(\mathbf{w}) + \mathscr{N}(\mathbf{w}). \tag{3}$$

We would like to point out that the choice of the linear and nonlinear operators, $\mathscr{L}$ and $\mathscr{N}$, respectively, is crucial. Indeed, we choose $\mathscr{L}$ to model linear acoustic and gravitational waves as well as a part of viscous fluxes, whereas the operator $\mathscr{N}$ describes resulting nonlinear advective/convective and the remaining viscous effects. Analogously, to split the diffusive terms into linear and nonlinear terms in $\mathbf{w}$, we set $\mathbf{D} = \mathbf{D}_L + \mathbf{D}_N$ with

$$\mathbf{D}_L = \left(0, \mu_m(\Delta(\rho\mathbf{u}) + \Delta(\rho\mathbf{u}^T)), \mu_h \Delta(\rho\theta)'\right)^T,$$

$$\mathbf{D}_N = \left(0, -\mu_m((\mathbf{u} + \mathbf{u}^T)\Delta\rho + \nabla\rho \cdot (\nabla\mathbf{u} + \nabla\mathbf{u}^T)), \mu_h\left(\Delta\left(\overline{\rho\theta}\right) - \theta\Delta\rho - \nabla\rho \cdot \nabla\theta\right)\right)^T.$$

Analogously as in [2] we then set

$$\mathscr{L}(\mathbf{w}) \equiv -\nabla \cdot \mathscr{F}_L(\mathbf{w}) + \mathbf{S}_L(\mathbf{w}) + \mathbf{D}_L(\mathbf{w}) := -\nabla \cdot \begin{pmatrix} \rho\mathbf{u} \\ p' \,\mathrm{Id} \\ \overline{\theta}\rho\mathbf{u} \end{pmatrix} + \begin{pmatrix} 0 \\ -\rho' g\mathbf{e}_3 \\ 0 \end{pmatrix} + \mathbf{D}_L(\mathbf{w})$$

with the linearized pressure $p' = \gamma_m \frac{\overline{p}(\rho\theta)'}{\overline{\rho}\overline{\theta}} \left(\frac{R\overline{\rho}\overline{\theta}}{p_0}\right)^{\gamma_m - \gamma}$, where $\gamma = 1.4$ is the adiabatic constant, and

$$\mathscr{N}(\mathbf{w}) \equiv -\nabla \cdot \mathscr{F}_N(\mathbf{w}) + \mathbf{S}_N(\mathbf{w}) + \mathbf{D}_N(\mathbf{w}) := -\nabla \cdot \begin{pmatrix} 0 \\ \rho \mathbf{u} \otimes \mathbf{u} \\ \theta' \rho \mathbf{u} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ S_\theta \end{pmatrix} + \mathbf{D}_N(\mathbf{w}).$$

Consequently, we discretize the Navier–Stokes equations by the IMEX scheme in time and approximate the linear stiff system at a new time level t_{n+1} and the nonlinear system at the old time level t_n. This yields the first order IMEX scheme. In order to increase the accuracy, the second order IMEX schemes can be applied, see, e.g., [1–3]. In our recent papers [1, 3] we have studied several second order IMEX schemes with respect to their asymptotic preserving properties. Here we confine ourselves with the second order globally stiffly accurate ARS(2,2,2) scheme

$$\begin{aligned} \mathbf{w}^{n+\frac{1}{2}} &= \mathbf{w}^n + \alpha \Delta t \left(\mathscr{L}\left(\mathbf{w}^{n+\frac{1}{2}}\right) + \mathscr{N}\left(\mathbf{w}^n\right)\right), \\ \mathbf{w}^{n+1} &= \mathbf{w}^n + \Delta t \left(\delta \mathscr{N}\left(\mathbf{w}^n\right) + (1-\delta)\mathscr{N}\left(\mathbf{w}^{n+\frac{1}{2}}\right)\right) \\ &\quad + \Delta t \left(\alpha \mathscr{L}\left(\mathbf{w}^{n+1}\right) + (1-\alpha)\mathscr{L}\left(\mathbf{w}^{n+\frac{1}{2}}\right)\right), \end{aligned} \tag{4}$$

where $\alpha = 1 - \frac{1}{\sqrt{2}}$, $\delta = 1 - \frac{1}{2\alpha}$ and $\Delta t = t_{n+1} - t_n$.

Spatial discretization is realized by the finite volume scheme. In particular having a regular rectangular grid we approximate the corresponding divergence operators by applying the Gauss theorem and the numerical flux functions in order to approximate fluxes along the cell interfaces. Let us denote the finite difference in the x_1-direction at the mesh cell $\Omega_{i,j,m} \equiv [x_i - \Delta x_1/2, x_i + \Delta x_1/2] \times [y_j - \Delta x_2/2, y_j + \Delta x_2/2] \times [z_m - \Delta x_3/2, z_m + \Delta x_3/2]$ by $\delta_{x_1} f_{i,j,m} \equiv f_{i+1/2,j,m} - f_{i-1/2,j,m}$; an analogous notation holds in the x_2 and x_3 direction. The finite volume discretization of the operators $\mathscr{L}$ and $\mathscr{N}$ yields

$$\mathscr{L}(\mathbf{w}^\ell)_{i,j,m} = -\sum_{k=1}^{3} \frac{1}{\Delta x_k} \delta_{x_k} \mathscr{F}_L^*(\mathbf{w}^\ell)_{i,j,m} + \mathbf{S}(\mathbf{w}^\ell)_{i,j,m} + \mathscr{D}_L(\mathbf{w}^\ell)_{i,j,m}, \quad \ell = n+1,$$

$$\mathscr{N}(\mathbf{w}^\ell)_{i,j,m} = -\sum_{k=1}^{3} \frac{1}{\Delta x_k} \delta_{x_k} \mathscr{F}_N^*(\mathbf{w}^\ell)_{i,j,m} + \mathscr{D}_N(\mathbf{w}^\ell)_{i,j,m}, \quad \ell = n, n-1.$$

Here $\mathscr{D}_L(\mathbf{w}^\ell)$ and $\mathscr{D}_N(\mathbf{w}^\ell)$ are the second order central difference approximations of the operators $\mathbf{D}_L(\mathbf{w}^\ell)$ and $\mathbf{D}_N(\mathbf{w}^\ell)$, respectively. For the numerical fluxes on cell interfaces, i.e., $\mathscr{F}_L^*(\mathbf{w}^\ell)$ and $\mathscr{F}_N^*(\mathbf{w}^\ell)$, we apply the central difference flux in the linear subsystem $\mathscr{L}$ and the Rusanov numerical flux in the nonlinear subsystem $\mathscr{N}$.

To keep the finite volume approximation of the explicit advection part stable we control the time step by the Courant–Friedrichs–Lewy stability condition

$$CFL_u \equiv \max_{k=1,2,3} \max_{i,j,m} |(u_k)_{i,j,m}| \frac{\Delta t}{\Delta x_k} < 1.$$

2.2 IMEX Finite Volume Scheme for the Cloud Dynamics Model

Similar to the IMEX finite volume discretization of the compressible Navier–Stokes equations in Sect. "IMEX Finite Volume Scheme for the Navier–Stokes Equations", we approximate the advection-diffusion-reaction system (2) by the finite volume method in space and by the second order IMEX scheme in time.

First, we rewrite the system (2) in the following compact form. Let $\mathbf{w}_q = (\rho q_v, \rho q_c, \rho q_r)^T$, $\mathbf{F}_q(\mathbf{w}_q) = (\rho q_v \mathbf{u}, \rho q_c \mathbf{u}, \rho q_r \mathbf{u} - \rho v_q \mathbf{e_3} q_r)^T$, $\mathbf{D}_q(\mathbf{w}_q) = \nabla \cdot (\rho \mu_h \nabla q_v, \rho \mu_h \nabla q_c, \rho \mu_h \nabla q_r)^T$ and $\mathbf{S}_q(\mathbf{w}_q) = (-C + E, C - A_1 - A_2, A_1 + A_2 - E)^T$, then (2) can be equivalently written as

$$\frac{\partial \mathbf{w}_q}{\partial t} = -\nabla \cdot \mathbf{F}_q(\mathbf{w}_q) + \mathbf{D}_q(\mathbf{w}_q) + \mathbf{S}_q(\mathbf{w}_q). \tag{5}$$

Realizing that $\mu_h \nabla \cdot (\rho \nabla q_x) = \mu_h \Delta (\rho q_x) - \mu_h q_x \Delta \rho - \mu_h \nabla \rho \cdot \nabla q_x$ for any $x \in \{c, v, r\}$ we can split $\mathbf{D}_q$ into

$$\mathbf{D}_q^{\text{impl}} = \mu_h \left(\Delta (\rho q_v), \ \Delta (\rho q_c), \ \Delta (\rho q_r)\right)^T,$$
$$\mathbf{D}_q^{\text{expl}} = -\mu_h \left(q_v \Delta \rho + \nabla \rho \cdot \nabla q_v, \ q_c \Delta \rho + \nabla \rho \cdot \nabla q_c, \ q_r \Delta \rho + \nabla \rho \cdot \nabla q_r\right)^T.$$

Then, Eq. (5) can be rewritten as

$$\frac{\partial \mathbf{w}_q}{\partial t} = \mathbf{I}_q(\mathbf{w}_q) + \mathbf{E}_q(\mathbf{w}_q), \tag{6}$$

where $\mathbf{I}_q(\mathbf{w}_q) = \mathbf{D}_q^{\text{impl}}(\mathbf{w}_q)$ and $\mathbf{E}_q(\mathbf{w}_q) = -\nabla \cdot \mathbf{F}_q(\mathbf{w}_q) + \mathbf{D}_q^{\text{expl}}(\mathbf{w}_q) + \mathbf{S}_q(\mathbf{w}_q)$.

Let us point out that the coupled system (1), (2) has a multiscale character, since it combines fast microscopic cloud dynamics with the slower macroscopic fluid flow. Therefore, we choose the time step for the cloud dynamics $\Delta t_{\text{cloud}} = \frac{\Delta t}{\text{const.}}$ for a sufficiently big constant. Time discretization of the cloud dynamics is realized by the second order IMEX ARS(2,2,2) scheme (4), whereas the diffusive terms $\mathbf{I}_q(\mathbf{w}_q)$ are approximated implicitly at a new micro-time level t_{s+1} and the remaining terms $\mathbf{E}_q(\mathbf{w}_q)$ explicitly at the old time level t_s.

The spatial discretization is done by the finite volume scheme. With the same notation as in Sect. "IMEX Finite Volume Scheme for the Navier–Stokes Equations" the discretizations of the operators $\mathbf{I}_q$ and $\mathbf{E}_q$ yield

$$\mathbf{I}_q(\mathbf{w}_q^\ell)_{i,j,m} = \mathscr{D}_q^{\text{impl}}(\mathbf{w}_q^\ell)_{i,j,m},\ \ell = s+1,$$
$$\mathbf{E}_q(\mathbf{w}_q^\ell)_{i,j,m} = -\sum_{k=1}^{3} \frac{1}{\Delta x_k}\delta_{x_k}\mathbf{F}_q^*(\mathbf{w}_q^\ell)_{i,j,m} + \mathscr{D}_q^{\text{expl}}(\mathbf{w}_q^\ell)_{i,j,m} + \mathbf{S}_q(\mathbf{w}_q^\ell)_{i,j,m},\ \ell = s, s-1,$$

where $\mathscr{D}_q^{\text{expl}}(\mathbf{w}_q^\ell)$ is a second order central difference approximation of $\mathbf{D}_q^{\text{expl}}(\mathbf{w}_q^\ell)$; analogous notation holds for $\mathscr{D}_q^{\text{impl}}(\mathbf{w}_q^\ell)$. For the numerical fluxes on the cell interfaces $\delta_{x_k}\mathbf{F}_q^*(\mathbf{w}_q^\ell)$ we apply the Lax–Friedrichs numerical flux.

3 Numerical Test

In this test we simulate free convection of a smooth warm air bubble as proposed in [5], see also [8]. A warm air bubble that is surrounded by cold air is placed on the bottom of the domain. Since the density of the warm air is lower than the surrounding air, the bubble rises up due to the buoyancy force. The experiment was simulated on a two-dimensional ($x_1 - x_3$ plane) domain $\Omega = [0,20.0]\times[0,10.0]\,[\text{km}^2]$.

The initial conditions for the Navier–Stokes equations are chosen as

$$\rho' = \frac{p_0}{R}\pi_e^{\frac{1}{\gamma-1}}\left(\frac{1}{\theta} - \frac{1}{\bar{\theta}}\right) = -\bar{\rho}\frac{\theta'}{\theta}, \quad \pi_e = 1 - \frac{g x_3}{c_p\bar{\theta}}, \quad \bar{\rho} = \frac{p_0}{R\bar{\theta}}\pi_e^{\frac{1}{\gamma-1}},$$
$$\mathbf{u} = 0,$$
$$\theta' = \begin{cases} 0, & r > r_c, \\ 2\cos^2\left(\frac{\pi r}{2}\right), & r \le r_c, \end{cases}$$

where $\bar{\theta} = 300\ [K], r = \|(x_1, x_3)^T - (10.0, 2.0)^T\|_2\,[\text{km}], r_c = 2.0\,[\text{km}], p_0 = \bar{p} = 10^5\,[\text{Pa}]$ and $(x_1, x_3)^T \in \Omega$. For the cloud model the initial conditions are

$$q_v = \min(q_* f(x_3), 0.014), \quad q_c = 0, \quad q_r = 0,$$

where f is the relative humidity and given by

$$f(x_3) = 1 - \frac{3}{4}\left(\frac{x_3}{x_{tr}}\right)^{5/4}, \quad x_{tr} = 12\,[km].$$

We apply the no-flux boundary conditions $\nabla\mathbf{u}\cdot\mathbf{n} = 0$, $\nabla\rho'\cdot\mathbf{n} = 0$, $\nabla(\rho\theta)'\cdot\mathbf{n} = 0$.

In Fig. 1 the time evolution of a moist air bubble, obtained by the ARS(2,2,2) finite volume approximation, is shown. The results for the potential temperature as well as for the vertical velocity are quite similar to the ones by Bryan and Fritsch proposed in [5] which confirms the reliability of our numerical model.

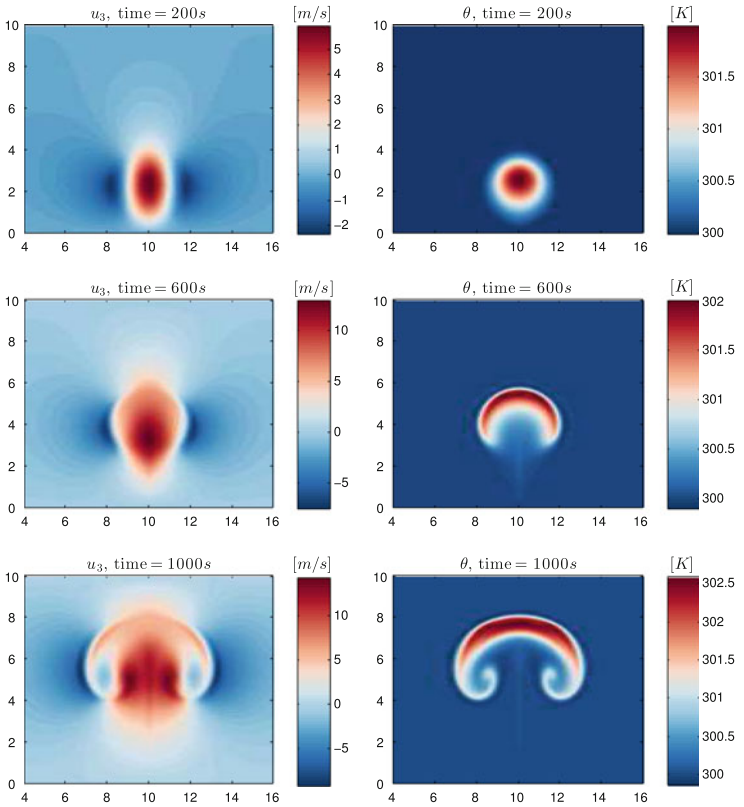


Fig. 1 Rising moist air bubble test computed by the IMEX ARS(2,2,2) finite volume scheme; $\mu_h = 10^{-2}$, $\mu_m = 10^{-3}$, $CFL_{\mathrm{u}} = 0.4$, the mesh resolution $\Delta x_1 = 100$ [m] and $\Delta x_3 = 50$ [m]. The colors correspond to the vertical velocity u_3 (*left column*) and the moist potential temperature θ (*right column*)

Acknowledgements The research leading to these results has been done within the subproject A2 of the Transregional Collaborative Research Center SFB/TRR 165 "Waves to Weather" funded by the German Science Foundation (DFG). The authors acknowledge the support of the Data Center ZDV in Mainz for providing computation time on MOGON cluster.

References

1. Bispen, G., Arun, K.R., Lukáčová-Medvid'ová, M., Noelle, S.: IMEX large time step finite volume methods for low Froude number shallow water flows. Comm. Comput. Phys. **16**, 307–347 (2014)
2. Bispen, G., Lukáčová-Medvid'ová, M., Yelash, L.: IMEX finite volume evolution Galerkin schemes for three-dimensional weakly compresible flows. In: Handlovičová, A (ed.) Algoritmy, pp. 62–73 (2016)
3. Bispen, G., Lukáčová-Medvid'ová, M., Yelash, L.: Asymptotic preserving IMEX finite volume schemes for low Mach number Euler equations with gravitation. J. Comput. Phys. (2017)
4. Brdar, S., Dedner, A., Klöfkorn R. Kränkel, R., Kröner, D.: Simulation of geophysical problems with DUNE-FEM. In: Krause, E (ed.) Computational Sci., & High Performance Computing IV, NNFM 115, pp. 93–106. Springer, Heidelberg (2011)
5. Bryan, G., Fritsch, J.M.: A benchmark simulation for moist nonhydrostatic numerical models. Mon. Weakly Rev. **130**, 2917–2928 (2002)
6. Khain, A.P., Beheng, K.D., Heymsfield, A., Korolev, A., Krichak, S.O., Levin, Z., Pinsky, M., Phillips, V., Prabhakaran, T., Teller, A., van den Heever, S.C., Yano, J.I.: Representation of microphysical processes in cloud-resolving models: spectral (bin) microphysics versus bulk parameterization. Rev. Geophy. **53**, 247–322 (2015)
7. Lamb, D., Verlinde, J.: Physics and Chemistry of Clouds. Cambridge University Press, Cambridge (2011)
8. Schuster, D., Brdar, S., Baldauf, M., Dedner, A., Klöfkorn, R., Kröner, D.: On discontinuous Galerkin approach for atmospheric flow in the mesoscale with and without moisture. Meteorologische Zeitschrift **23**(4), 449–464 (2011)
9. Wacker, U.: Structural stability in cloud physics using parameterized microphysics. Beitr. Phys. Atmosph. **65**, 231–242 (1992)

Hybrid Stochastic Galerkin Finite Volumes for the Diffusively Corrected Lighthill-Whitham-Richards Traffic Model

Raimund Bürger and Ilja Kröker

Abstract The vehicular traffic on an infinite or circular highway can be represented by the diffusively corrected Lighthill-Whitham-Richards (DCLWR) model, but the uncertain knowledge requires to deal with uncertain model parameters. The stochastic Galerkin based methods allow to transform a randomly perturbed PDE to a high-dimensional deterministic system, where the dimension of the system increases rapidly with the increasing number of the uncertain parameters. In this work we consider the application of the hybrid stochastic Galerkin (HSG) finite volume method to the uncertain DCLWR model, which is represented by a random perturbed strongly degenerate parabolic equation. We present the resulting high–dimensional system and corresponding finite volume method. Numerical examples cover the scalar problem with four uncertain parameters.

Keywords Finite volume · Uncertainty quantification · Stochastic galerkin · Traffic modelling

MSC (2010): 65M08 · 68U20 · 35R60

1 Introduction

This work is focused on numerical methods for the quantification of the stochastic variability of solutions $u = u(x, t)$ of the strongly degenerate parabolic equation

$$\partial_t u + \partial_x f(u) = \partial_x^2 A(u), \quad (x, t) \in Q_T := I \times (0, T), \quad I \subset \mathbb{R}, \quad T > 0. \quad (1)$$

R. Bürger
CI^{2}MA and Departamento de Ingeniería Matemática, Universidad de Concepción, Casilla 160-C, Concepcion, Chile
e-mail: rburger@ing-mat.udec.cl

I. Kröker (✉)
IANS, Universtität Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany
e-mail: ilja.kroeker@mathematik.uni-stuttgart.de

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_21

The stochastic variability arises from uncertainty in the parameters that define the function $a = a(u)$, where

$$A(u) = \int_0^u a(s)\,\mathrm{d}s, \quad a \in L^1[0, u_{\max}], \quad a(u) \geq 0 \quad \text{for } 0 \leq u \leq u_{\max}. \tag{2}$$

Equation (1) is posed with suitable initial and boundary conditions. We assume that f is a piecewise smooth, Lipschitz continuous, non-negative function with support on $(0, u_{\max})$, where $u_{\max}$ is a maximum solution value. We assume that $a(u) = 0$ on u-intervals of positive lengths, which motivates why (1) is called *strongly degenerate*. For the corresponding solution values, (1) degenerates into a first-order conservation law, where the location of the type-change interface is unknown beforehand. Stochastic Galerkin-based methods have successfully been applied to several hyperbolic problems in [3, 4, 6, 7, 11, 13]. We herein apply the hybrid stochastic Galerkin (HSG) finite volume method to the diffusively corrected Lighthill-Whitham-Richards (DCLWR) traffic model with four uncertain parameters.

In Sect. 2 we introduce the DCLWR traffic model and extend it by four uncertain parameters. Section 3 contains a short introduction into HSG discretization, and we specify the finite volume method for a degenerate problem. In Sect. 4 we present results of numerical experiments and discuss the accuracy of the method.

2 Diffusively Corrected LWR Traffic Model

The classical Lighthill-Whitham-Richards (LWR) traffic model [9, 12] postulates that vehicular traffic on an infinite or circular highway can be modeled by a first-order conservation law $\partial_t u + \partial_x f(u) = 0$, where unknown u is the local density of cars that is assumed to vary between zero and a maximum value $u_{\max}$, and

$$f(u) = v_{\max} u V(u), \quad V(u) = 1 - u/u_{\max}. \tag{3}$$

Here $v_{\max} > 0$ is a maximum free-way velocity and V is a decreasing function that satisfies $V(0) = 1$ and $V(u_{\max}) = 0$, and which describes drivers' attitude to reduce speed in presence of other cars. We employ the approach by Nelson [10] (see also [1]) to modify the LWR model so that the effects of anticipation distance and reaction time are included. The anticipation length L_{a} may depend on $V(u)$. In fact, the following formula is proposed in [10]:

$$L_{\mathrm{a}}(u) = \max\left\{\frac{(v_{\max} V(u))^2}{2\alpha}, L_{\min}\right\}, \tag{4}$$

where $L_{\min}$ is a minimal anticipation distance and α denotes a deceleration, so that the first argument in (4) denotes the distance required to decelerate from speed $V(u)$ to full stop at deceleration α. In [1, 10] it is proposed to utilize a particular

function V that satisfies $V(u) = \text{const.}$ for $u < u_*$, where $0 < u_* < u_{\max}$, such that the strongly degenerate behaviour of a holds for $u_c := u_*$ and $u = u_{\max}$. Herein we assume, however, that independently of the algebraic definition of V, the critical value u_c acts as a "physiological" threshold value in the sense that reaction times and anticipation lengths are assumed to be effective only whenever the local traffic density u exceeds a critical value u_c. Consequently, our analysis will be based on the following formula:

$$a(u) = \begin{cases} 0 & \text{for } u \le u_c, \\ -uv_{\max}V'(u)\big(L_a(u) + \tau uv_{\max}V'(u)\big) & \text{for } u > u_c. \end{cases} \tag{5}$$

For our formulation of the random perturbed problem we replace the "physiological" threshold u_c, reaction time τ, deceleration α, and the anticipation length $L_{\min}$ by random parameters $p_i := \bar{p}_i\left(1 + \left(\theta_i - \frac{1}{2}\right)\sigma_i\right)$, for $i = 1, \ldots, 4$, where θ_i is a random variable whose law is the uniform law on $(0, 1)$, which is written $\theta_i \sim \mathscr{U}(0, 1)$. Here $\bar{p}_i$ represents the mean and σ_i the magnitude of the random perturbation. For the final formulation we obtain the following form: For a final time $T > 0$ and $A(u(x, t, \boldsymbol{\omega}), \boldsymbol{\omega}) := \int_0^{u(x,t,\boldsymbol{\omega})} a(s, \boldsymbol{\omega})\,\mathrm{d}s$ find $u : \mathbb{R} \times [0, T] \times \Omega^4 \to \mathbb{R}$ that satisfies

$$\partial_t u(x, t, \boldsymbol{\omega}) + \partial_x f(u(x, t, \boldsymbol{\omega})) = \partial_x^2 A(u(x, t, \boldsymbol{\omega}), \boldsymbol{\omega}) \quad \text{on } \mathbb{R} \times (0, T] \times \Omega^4, \tag{6}$$

$$u(x, 0, \boldsymbol{\omega}) = u_0(x) \quad \text{on } \mathbb{R} \times \Omega^4. \tag{7}$$

3 Hybrid Stochastic Galerkin (HSG) Discretization

We now give a short overview on the HSG discretization. We start with an introduction to the general polynomial chaos (gPC) expansion and its extension to the HSG discretization. Then we proceed with their applications to (1).

Let $\boldsymbol{\theta}(\boldsymbol{\omega}) := \{\theta_1(\omega_1), \ldots, \theta_N(\omega_N)\}$ be an N–dimensional random vector of i.i.d. (independent identically distributed) random variables defined on the probability spaces $(\Omega_i, \mathscr{F}_i, \mathscr{P}_i)$, $i = 1, \ldots, N$. We define a multivariate polynomial $\Phi_{\mathfrak{p}}$ for a multi-index $\mathfrak{p} \in \mathbb{N}^N$ by $\Phi_{\mathfrak{p}}(\boldsymbol{\omega}) := \phi_{\mathfrak{p}_1}(\theta_1) \cdot \ldots \cdot \phi_{\mathfrak{p}_N}(\theta_N)$. The choice of the orthonormal polynomial $\phi_{\mathfrak{p}_i}$, $\mathfrak{p}_i \in \mathbb{N}_0$ depends on the law of the random variable θ_i. In presented work we assume that the random variables are uniformly distributed $\theta_i \sim \mathscr{U}(0, 1)$ and use therefore re-scaled Legendre polynomials. The family of the multivariate polynomials $\left\{\Phi_{\mathfrak{p}}\right\}_{\mathfrak{p}\in\mathbb{N}_0^N}$ is orthonormal w.r.t. the scalar product on $L^2(\Omega^N) := L^2(\Omega_1 \times \cdots \times \Omega_N)$, i.e.,

$$\left\langle \Phi_{\mathfrak{p}},\, \Phi_{\mathfrak{q}} \right\rangle_{L^2(\Omega^N)} := \int_{\Omega_1} \cdots \int_{\Omega_N} \Phi_{\mathfrak{p}}(\boldsymbol{\theta}(\boldsymbol{\omega}))\Phi_{\mathfrak{q}}(\boldsymbol{\theta}(\boldsymbol{\omega}))\,\mathrm{d}\mathscr{P}_1(\omega_1) \cdots \mathscr{P}_N(\omega_N) = \delta_{\mathfrak{p}\mathfrak{q}}.$$

Further, the polynomial chaos expansion of a random variable with finite variance $w = w(x, t, \boldsymbol{\theta}(\boldsymbol{\omega}))$, $(x, t) \in \mathbb{R} \times [0, T]$, $\boldsymbol{\omega} \in \Omega^N$ is given by

$$\begin{aligned} w(x, t, \boldsymbol{\theta}(\boldsymbol{\omega})) &= \sum_{q=0}^{\infty} \sum_{|\mathfrak{p}|=q} w^{\mathfrak{p}}(x, t) \Phi_{\mathfrak{p}}(\boldsymbol{\theta}(\boldsymbol{\omega})), \qquad \text{for } (x, t) \in \mathbb{R} \times [0, T],\ \boldsymbol{\omega} \in \Omega^N. \\ w^{\mathfrak{p}}(x, t) &= \left\langle w(x, t, \cdot),\ \Phi_{\mathfrak{p}} \right\rangle_{L^2(\Omega_1 \times \cdots \times \Omega_N)}, \end{aligned}$$

The truncation at N_o leads to a sum of $\frac{(N_\mathrm{o}+N)!}{N_\mathrm{o}!N!}$ terms.

The main idea of the HSG method is the decomposition of the stochastic domain $[0, 1]^N$ (we assume $\theta_i \sim \mathscr{U}(0, 1)$) into 2^{NN_r} ($N_\mathrm{r} \in \mathbb{N}_0$) sub-domains

$$I_{N,l}^{N_\mathrm{r}} := I_{l_1}^{N_\mathrm{r}} \times \cdots \times I_{l_N}^{N_\mathrm{r}}, \quad l = (l_1, \ldots, l_N) \in \mathscr{I} := \left\{0, \ldots, 2^{N_\mathrm{r}} - 1\right\}^N.$$

Here the interval $I_{l_i}^{N_\mathrm{r}}$ for $l_i = 0, \ldots, 2^{N_\mathrm{r}} - 1$ is given by

$$I_{l_i}^{N_\mathrm{r}} := [2^{-N_\mathrm{r}} l_i, 2^{-N_\mathrm{r}}(l_i + 1)].$$

The space of the multivariate piecewise polynomial functions $S_N^{N_\mathrm{o},\, N_\mathrm{r}}$ is given by

$$S_N^{N_\mathrm{o},\, N_\mathrm{r}} := \left\{ w : [0, 1]^N \to \mathbb{R} \;\middle|\; w|_{I_{N,l}^{N_\mathrm{r}}} \in \mathbb{Q}_{N_\mathrm{o}}^N[\boldsymbol{\theta}],\ \forall l \in \mathscr{I} \right\}.$$

Here $\mathbb{Q}_{N_\mathrm{o}}^N[\boldsymbol{\theta}]$ denotes the space of N-variate polynomials of degree $\leq N_\mathrm{o}$. The basis of the space $S_N^{N_\mathrm{o},\, N_\mathrm{r}}$ can be given by the polynomials

$$\Phi_{\mathfrak{p},l}^{N_\mathrm{r}}(\boldsymbol{\theta}) := \begin{cases} 2^{NN_\mathrm{r}/2} \Pi_{k=1}^N \phi_{\mathfrak{p}_k}(2^{N_\mathrm{r}} \theta_k - l_k) & \text{for } \boldsymbol{\theta} \in I_{N,l}^{N_\mathrm{r}}, \\ 0 & \text{otherwise,} \end{cases} \qquad (8)$$
$$\text{for } \mathfrak{p} \in \mathbb{N}_0^N,\ |\mathfrak{p}| \leq N_\mathrm{o},\ l \in \mathscr{I},$$

where $\{\phi_k\}_{k \geq 0}$ are the re-scaled orthonormal Legendre polynomials. Therefore also the N-variate polynomials (8) satisfy the orthogonality relation

$$\left\langle \Phi_{\mathfrak{p},l}^{N_\mathrm{r}},\ \Phi_{\mathfrak{q},k}^{N_\mathrm{r}} \right\rangle_{L^2(\Omega^N)} = \delta_{\mathfrak{p}\mathfrak{q}} \delta_{lk}, \quad \text{for } \mathfrak{p}, \mathfrak{q} \in \mathbb{N}_0^N,\ k, l \in \mathscr{I}. \qquad (9)$$

Similarly as for the gPC expansion we define the projection $\Pi^{N_\mathrm{o}, N_\mathrm{r}} : L^2(\Omega^N) \to S_N^{N_\mathrm{r},\, N_\mathrm{o}}$ of the random variable $w(x, t, \boldsymbol{\theta}(\boldsymbol{\omega})) \in L^2(\Omega^N)$ for $(x, t) \in \mathbb{R} \times [0, T]$ by

$$\Pi^{N_\mathrm{o}, N_\mathrm{r}}[w](x, t, \boldsymbol{\theta}) := \sum_{l \in \mathscr{I}} \sum_{|\mathfrak{p}| \leq N_\mathrm{o}} w_l^{\mathfrak{p}}(x, t) \Phi_{\mathfrak{p},l}^{N_\mathrm{r}}(\boldsymbol{\theta}), \quad w_{\mathfrak{p},l}^{N_\mathrm{r}} := \left\langle w,\ \Phi_{\mathfrak{p},l}^{N_\mathrm{r}} \right\rangle_{L^2(\Omega^N)},$$

for $l \in \mathscr{I}$, $\mathfrak{p} \in \mathbb{N}_0^N$. The convergence of Π^{N_o,N_r} for $N_r, N_o \to \infty$ was shown in [2] and by the Cameron-Martin theorem [5]. Due to the assumptions on the random variable w the expectation and variance of $\Pi^{N_o,N_r}[w]$ can be computed as follows:

$$\mathbf{E}\left[\Pi^{N_o,N_r}[w]\right](x,t) := \sum_{l\in\mathscr{I}} w_{0,l}^{N_r}(x,t)\left\langle \Phi_{0,l}^{N_r},\ \Phi_{0,0}^{0}\right\rangle_{L^2(\Omega^N)},$$

$$\mathbf{Var}\left[\Pi^{N_o,N_r}[w]\right](x,t) := \sum_{l\in\mathscr{I}}\sum_{|\mathfrak{p}|\le N_o} w_{\mathfrak{p},l}^{N_r}(x,t)^2 - \left(\mathbf{E}\left[\Pi^{N_o,N_r}[w]\right](x,t)\right)^2.$$

In order to apply the HSG approach to the final model we replace the unknown u in the final Eqs. (6), (7) by its projection onto $S_N^{N_r,N_o}$ for $N,\ N_r,\ N_o \in \mathbb{N}_0$, denoted by $\Pi^{N_o,N_r}[u]$. The HSG approach of the modified equation yields the following problem: find coefficients $u_{\mathfrak{p},l}^{N_r} : \mathbb{R}\times[0,T]\to\mathbb{R}$ such that

$$\int_{\Omega^N} \left(\ \partial_t(\Pi^{N_o,N_r}[u]) + \partial_x f\left(\Pi^{N_o,N_r}[u]\right) - \partial_x^2 A\left(\Pi^{N_o,N_r}[u]\right)\ \right)\Phi_{\mathfrak{p},l}^{N_r}\,\mathrm{d}\mathscr{P}(\boldsymbol{\omega}) = 0,$$

$$\text{for all } (\mathfrak{p},l)\in\mathbb{N}_0^N\times\mathscr{I},\ |\mathfrak{p}|\le N_o.$$

By using of the orthogonal relation (9) we obtain for $(x,t,\boldsymbol{\omega})\in\mathbb{R}\times(0,T]\times\Omega^N$, $\alpha=(\mathfrak{p},l)\in\mathbb{N}_0^N\times\mathscr{I}$, $|\mathfrak{p}|\le N_o$ the $M := 2^{NN_r}\frac{(N_o+N)!}{N_o!N!}$ dimensional system

$$\partial_t u^\alpha + \partial_x\left(\left\langle f\left(\Pi^{N_o,N_r}[u]\right),\ \Phi_\alpha\right\rangle_{L^2(\Omega^N)}\right) = \partial_x^2\left\langle A\left(\Pi^{N_o,N_r}[u]\right),\ \Phi_\alpha\right\rangle_{L^2(\Omega^N)}. \quad (10)$$

3.1 Finite Volume Method

For the numerical approach we extend the central-upwind scheme, which was introduced in [8, 14], and successful used together with HSG discretization in [4, 6, 7] with a second order term and obtain for $j\in\mathbb{Z}$ the following numerical scheme:

$$\frac{\mathrm{d}\bar{\mathbf{u}}_{j+1/2}(t)}{\mathrm{d}t} = -\frac{\mathbf{F}_{j+1}(t)-\mathbf{F}_j(t)}{\triangle x} + \frac{\mathbf{A}(\bar{\mathbf{u}}_{j+3/2}(t)) - 2\mathbf{A}(\bar{\mathbf{u}}_{j+1/2}(t)) + \mathbf{A}(\bar{\mathbf{u}}_{j-1/2}(t))}{(\triangle x)^2},$$

$$\mathbf{F}_j(t) := \frac{a_j^+\mathbf{f}(\mathbf{u}_j^+) + a_j^-\mathbf{f}(\mathbf{u}_j^-)}{a_j^+ + a_j^-} + \frac{a_j^+ a_j^-}{a_j^+ + a_j^-}\left(\mathbf{u}_j^+ - \mathbf{u}_j^-\right).$$

Here the cell average on $[x_j, x_{j+1}]$ denoted by $\bar{\mathbf{u}}_{j+1/2}$ and the piecewise polynomial reconstructions denoted by $\mathbf{u}_j^\pm$ are given by

$$\bar{\mathbf{u}}_{j+1/2} = \left(\bar{u}_{j+1/2}^0,\ldots,\bar{u}_{j+1/2}^{M-1}\right)^{\mathrm{T}} \quad\text{and}\quad \mathbf{u}_j^\pm = \left((u_j^\pm)^0,\ldots,(u_j^\pm)^{M-1}\right)^{\mathrm{T}}.$$

The vectors $\mathbf{F}(\mathbf{u}_j)$ and $\mathbf{A}(\mathbf{u}_j)$ are given by

$$\mathbf{F}(\mathbf{u}_j) := \left(F^0(\mathbf{u}_j), \ldots, F^{M-1}(\mathbf{u}_j)\right)^{\mathrm{T}}, \quad \mathbf{A}(\mathbf{u}_j) := \left(A^0(\mathbf{u}_j), \ldots, A^{M-1}(\mathbf{u}_j)\right)^{\mathrm{T}},$$

$$F^\alpha(\mathbf{u}_j) := \left\langle F\left(\sum_{\beta=0}^{M-1} u_j^\beta \Phi_\beta\right), \Phi_\alpha \right\rangle_{L^2(\Omega^4)}, \quad A^\alpha(\mathbf{u}_j) := \left\langle A\left(\sum_{\beta=0}^{M-1} u_j^\beta \Phi_\beta\right), \Phi_\alpha \right\rangle_{L^2(\Omega^4)}.$$

The so-called local speeds $a_j^\pm$ are derived from the Jacobian of $\mathbf{F}$, we refer to [8, 14] for details. The time discretization is given by the second-order Runge-Kutta (Heun) method.

4 Numerical Experiments

For the DCLWR traffic model we use the flux function (3) with $v_{\max} = 100\,\mathrm{km/h} = 2.78 \times 10^{-2}\,\mathrm{m/s}$. The unknown density u is measured in cars per kilometre, and we assume that $u_{\max} = 120\,\mathrm{cars/km}$. The other model parameters are chosen similar to those used in [1]. The random perturbed critical density is set to u_c corresponding to $\bar{p}_1 = 10\,\mathrm{cars/km}$ and $\sigma_1 = 0.25$. For the reaction and anticipation terms, we set $\alpha = \bar{p}_3 = 0.1g = 0.981\,\mathrm{m/s^2}$, $\sigma_3 = 0.5$, where g is the acceleration of gravity, $\tau = \bar{p}_2 = 2\,\mathrm{s}$, $\sigma_2 = 0.25$ and $L_{\min} = \bar{p}_4 = 80\,\mathrm{m}$, $\sigma_4 = 0.125$. The interval $I = (0, 5.6)$ km is discretized by 200 points. The functions a and L_a are defined in (5) and (4) respectively. Initial distribution is given by

$$u_0(x) := \begin{cases} 80\ {}^{\mathrm{cars}}/_{\mathrm{km}} & \text{for } 2.0\ \mathrm{km} \le x \le 2.5\ \mathrm{km}, \\ 0\ {}^{\mathrm{cars}}/_{\mathrm{km}} & \text{otherwise.} \end{cases} \tag{11}$$

The numerical results were computed with HSG-FV with $N_r = 0, \ldots, 2$, $N_o = 0, \ldots, 2$ on 200 points. Figure 1(a) shows expectation and variance at $T = 100$. Figure 1(b) shows reconstructions of the numerical solution for ten parameter sets at $T = 100$. Those reconstructions illustrates the dependence of the solution on the possible random perturbed parameter values and explain the shape of the variance on the previous figure. Figure 1(c), (d) show the evolution of the expectation and variance of the numerical solution for $t \in [0, 100]$. Figure 1(e), (f) show expectation and variance computed with several choices of N_r and N_o compared with Monte Carlo (MC) solutions computed with 100000 samples. Table 1 shows $L^1(\mathbb{R})$- and $L^2(\mathbb{R})$-error for expectation and $L^2(\mathbb{R})$- and $L^4(\mathbb{R})$-error for variance of the HSG-FV approach compared with MC solution computed with 100000 samples.

The presented numerical experiments show the expected behaviour of expectation and variance.

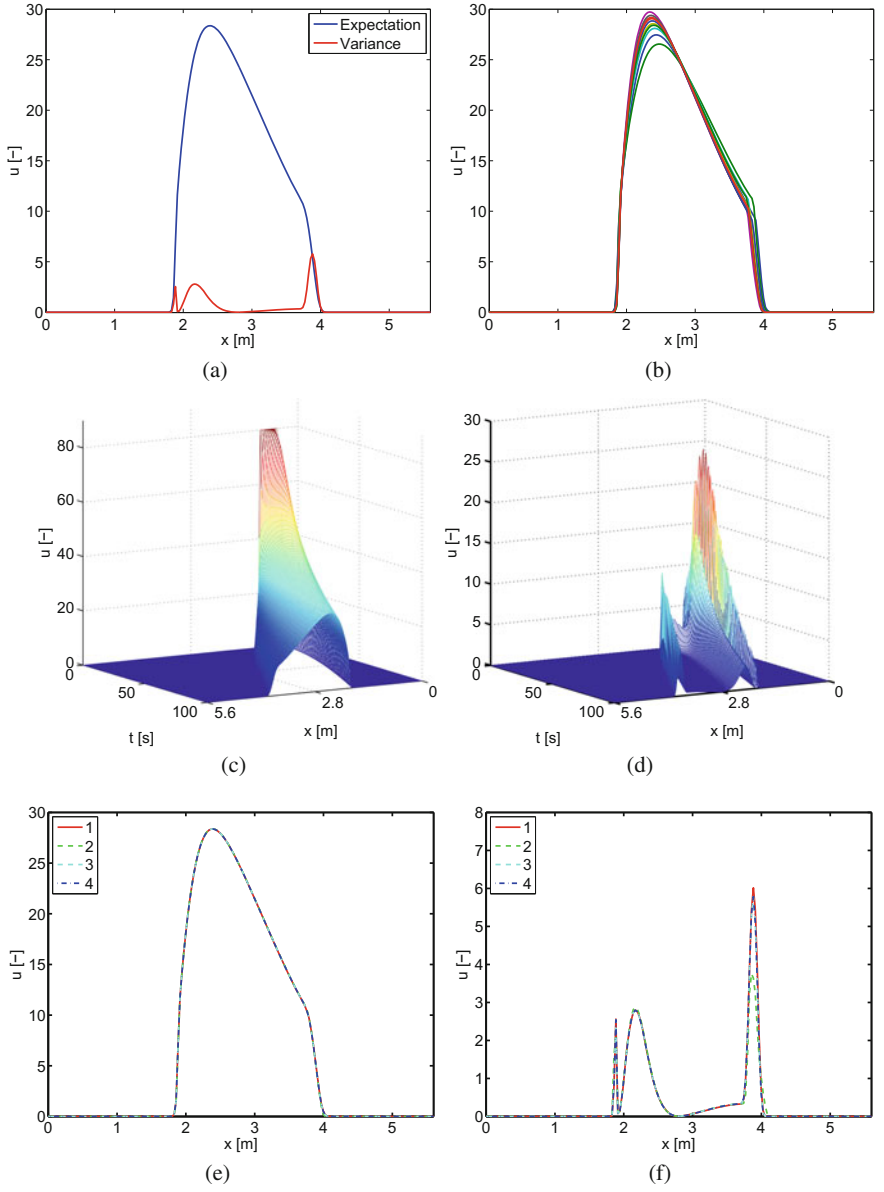


Fig. 1 Numerical results for $T = 100$ s. and 200 points, computed with central-upwind and HSG for $N_r = 2$, $N_o = 2$. (**a**) Expectation and variance; (**b**) reconstructions for 10 several parameter sets; (**c**) time series of the expectation of the solution for $t \in [0, 100]$; (**d**) time series of the variance of the solution for $t \in [0, 100]$; comparison of expectations (**e**) and variances (**f**): (*1*) MC with 100000 samples, (*2*) HSG with $N_r = 0$, $N_o = 2$, (*3*) HSG with $N_r = 1$, $N_o = 2$, (*4*) HSG with $N_r = 2$, $N_o = 2$

Table 1 Expectation: (a) $L^1(\mathbb{R})$-error, (b) $L^2(\mathbb{R})$-error of the HSG-FV approach compared with MC. Variance: (c) $L^2(\mathbb{R})$-error, (d) $L^4(\mathbb{R})$-error of the HSG-FV approach compared with MC

N_o	$N_r = 0$	$N_r = 1$	$N_r = 2$
(a)			
0	2.63e-01	6.64e-02	1.72e-02
1	1.19e-01	5.38e-02	2.26e-02
2	1.11e-01	6.06e-02	2.81e-02
(b)			
0	2.54e-01	5.92e-02	1.50e-02
1	1.35e-01	5.49e-02	2.31e-02
2	1.20e-01	6.11e-02	2.60e-02
(c)			
0	2.48e+00	5.78e-01	1.39e-01
1	7.76e-01	2.58e-01	9.82e-02
2	7.18e-01	2.66e-01	1.04e-01
(d)			
0	3.26e+00	7.51e-01	1.75e-01
1	1.18e+00	4.36e-01	1.67e-01
2	1.19e+00	4.55e-01	1.76e-01

Acknowledgements R. B. is supported by Fondecyt project 1130154; BASAL project CMM, Universidad de Chile and CI^{2}MA, Universidad de Concepción; Fondef ID15I10291; and CRHIAM, Proyecto Conicyt Fondap 15130015. I. K. would like to thank the German Research Foundation (DFG) for financial support of the project within the Cluster of Excellence in Simulation Technology (EXC 310/1) at the University of Stuttgart.

References

1. Acosta, C., Bürger, R., Mejía, C.: Efficient parameter estimation in a macroscopic traffic flow model by discrete mollificatio. Transp. A Transp. Sci. **11**, 702–715 (2015)
2. Alpert, B.K.: A class of bases in L^2 for the sparse representation of integral operators. SIAM J. Math. Anal. **24**(1), 246–262 (1993). doi:10.1137/0524016
3. Barth, A., Bürger, R., Kröker, I., Rohde, C.: Computational uncertainty quantification for a clarifier-thickener model with several random perturbations: a hybrid stochastic Galerkin approach. Comput. Chem. Eng. **89**, 11–26 (2016)
4. Bürger, R., Kröker, I., Rohde, C.: A hybrid stochastic Galerkin method for uncertainty quantification applied to a conservation law modelling a clarifier-thickener unit. ZAMM Z. Angew. Math. Mech. **94**(10), 793–817 (2014). doi:10.1002/zamm.201200174
5. Cameron, R.H., Martin, W.T.: The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals. Ann. Math. **2**(48), 385–392 (1947)

6. Köppel, M., Kröker, I., Rohde, C.: Stochastic modeling for heterogeneous two-phase flow. In: J. Fuhrmann, M. Ohlberger, C. Rohde (eds.) Finite Volumes for Complex Applications VII-Methods and Theoretical Aspects, *Springer Proceedings in Mathematics & Statistics*, vol. 77, pp. 353–361. Springer International Publishing (2014). doi:10.1007/978-3-319-05684-5_34
7. Kröker, I., Nowak, W., Rohde, C.: A stochastically and spatially adaptive parallel scheme for uncertain and nonlinear two-phase flow problems. Comput. Geosci. pp. 1–16 (2015)
8. Kurganov, A., Petrova, G.: Central-upwind schemes on triangular grids for hyperbolic systems of conservation laws. Numer. Methods Partial. Differ. Equ. **21**(3), 536–552 (2005)
9. Lighthill, M., Whitham, G.: On kinematic waves: II. A theory of traffic flow on long crowded roads. Proc. R. Soc. A **229**, 317–345 (1955)
10. Nelson, P.: Traveling-wave solutions of the diffusively corrected kinematic-wave model. Math. Comput. Model. **35**, 561–579 (2002)
11. Poëtte, G., Després, B., Lucor, D.: Uncertainty quantification for systems of conservation laws. J. Comput. Phys. **228**(7), 2443–2467 (2009). doi:10.1016/j.jcp.2008.12.018
12. Richards, P.: Shock waves on the highway. Oper. Res **4**, 42–51 (1956)
13. Tryoen, J., Le Maître, O., Ndjinga, M., Ern, A.: Intrusive Galerkin methods with upwinding for uncertain nonlinear hyperbolic systems. J. Comput. Phys. **229**(18), 6485–6511 (2010)
14. Wang, G., Ge, C.: Semidiscrete central-upwind scheme for conservation laws with a discontinuous flux function in space. Appl. Math. Comput. **217**(17), 7065–7073 (2011)

The RS-IMEX Scheme for the Rotating Shallow Water Equations with the Coriolis Force

Hamed Zakerzadeh

Abstract In this note, we comment on the applicability of the recently-presented RS-IMEX scheme for the rotating shallow water equations. We show the asymptotic consistency of the scheme for the quasi-geostrophic distinguished limit. We also test the quality of the computed solution by a numerical example.

Keywords Asymptotic preserving scheme · Rotating shallow water system

MSC (2010) : 65M08 · 76U05 · 76M45 · 86A05

1 Introduction

Accurately enough, one can claim that the modern era of numerical schemes for geophysical flows has been started with [7]; it simplified the meteorological equations by *filtering out* noises (fast gravity waves) which do not contribute to the bulk motion of the fluid and derived the so-called *barotropic quasi-geostrophic equations*; see [12]. Nowadays, one is able to handle the computational cost of more sophisticated models like the shallow water equations; but, due to the presence of fast waves, the system is stiff and requires using very fine grids or devising schemes covering several scales in time and/or space at once. An interesting example is the large-scale rotating shallow water equations (RSWE) when the Rossby and Froude numbers approach zero in the so-called *quasi-geostrophic limit*, as the scaling parameter ε vanishes; see (3). For this limit, it is proved in [11] that the RSWE converge to the quasi-geostrophic equations, which are the equations derived formally in [7]. This ensures that, at least in the continuous level, there is a convergence for $\varepsilon \to 0$. The so-called *Asymptotic Preserving (AP) schemes* are defined to preserve such a convergence in the discrete level; they are supposed to be consistent, stable and efficient uniformly

H. Zakerzadeh (✉)
Institut Für Geometrie Und Praktische Mathematik RWTH Aachen University, Templergraben 55, 52056 Aachen, Germany
e-mail: h.zakerzadeh@igpm.rwth-aachen.de

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_22

in ε, cf. [8]. We aim here to design and analyze an AP scheme for this singularly-perturbed example as such a study has not been performed to its merit in the literature; see [2, 3, 5, 10] for some contributions.

We consider the Reference Solution Implicit-Explicit scheme, which has been shown to be well-behaved for the shallow water equations with topography in [13, 14], and we see if the scheme works well with the additional Coriolis force. In Sect. 2 we introduce the RS-IMEX scheme for the RSWE. Then, in Sect. 3 we discuss the asymptotic consistency of the scheme, followed by a numerical example in Sect. 4.

2 RS-IMEX Scheme for the Rotating Shallow Water Equations

The 2d RSWE in the domain $\Omega \subset \mathbb{R}^2$ lying in the $\boldsymbol{x} = (x_1, x_2)$ plane write [12]

$$\begin{aligned} &\partial_t h + \nabla_{\boldsymbol{x}} \cdot (h\boldsymbol{u}) = 0, \\ &\partial_t (h\boldsymbol{u}) + \nabla_{\boldsymbol{x}} \cdot \left(h\boldsymbol{u} \otimes \boldsymbol{u} + \frac{gh^2}{2} \mathbb{I}_2 \right) = -gh\nabla_{\boldsymbol{x}} \eta^b - fh\boldsymbol{u}^{\perp}, \end{aligned} \tag{1}$$

where h is the water height, η^b is the bottom function, $\boldsymbol{u} = (u_1, u_2)$ is the 2d velocity vector, $\boldsymbol{u}^{\perp} = (-u_2, u_1)$ is the *orthogonal velocity*, g is the gravity acceleration constant, $\mathbb{I}_2$ is the 2×2 identity matrix and the constant f is the Coriolis parameter. We limit our focus on periodic domains $\Omega = \mathbb{T}^2$ for simplicity.

The *Buckingham π-theorem* [6] implies that there are three different dimensionless groups for this system: the Strouhal number St, the Froude number Fr and the Rossby number Ro. But since we consider two height scales, as in [11, Chap. 4], we should also introduce another dimensionless group Θ. The height scales are $H_\circ$ for the mean water level chosen equal to the actual mean water level H_{mean}, and $Z_\circ$ for the surface perturbation from H_{mean} (denoted by z as in [4, 13]), i.e., $h = H_{\text{mean}} + z - \eta^b$. Defining dimensionless variables as $\hat{\boldsymbol{x}} := \boldsymbol{x}/L_\circ$, $\hat{t} := t/t_\circ$, $\hat{\boldsymbol{u}} := \boldsymbol{u}/U_\circ$, $\hat{z} := z/Z_\circ$, $\hat{\eta}^b := \eta^b/Z_\circ$ and $\hat{h} := h/H_\circ$, where characteristic states are denoted by subscript $\circ$, one obtains $\hat{h} = 1 + \Theta(\hat{z} - \hat{\eta}^b)$ and can re-write (1) as (cf. [11])

$$\begin{aligned} &St\, \partial_{\hat{t}} \left(\Theta \hat{z} \right) + \nabla_{\hat{\boldsymbol{x}}} \cdot \left(\hat{h}\hat{\boldsymbol{u}} \right) = 0, \\ &St\, \partial_{\hat{t}} \left(\hat{h}\hat{\boldsymbol{u}} \right) + \nabla_{\hat{\boldsymbol{x}}} \cdot \left(\hat{h}\hat{\boldsymbol{u}} \otimes \hat{\boldsymbol{u}} + \frac{\hat{h}^2}{2Fr^2} \mathbb{I}_2 \right) = -\frac{\Theta}{Fr^2} \hat{h} \nabla_{\hat{\boldsymbol{x}}} \hat{\eta}^b - \frac{\hat{h}}{Ro} \hat{\boldsymbol{u}}^{\perp}, \end{aligned} \tag{2}$$

with the following definitions for St, Fr, Ro and Θ:

$$St := \frac{L_\circ}{U_\circ t_\circ}, \quad Fr := \frac{U_\circ}{\sqrt{gH_\circ}}, \quad Ro := \frac{U_\circ}{fL_\circ}, \quad \Theta := \frac{Z_\circ}{H_\circ}.$$

The relation between these groups characterizes the *distinguished limit*. We choose $St = 1$, which is suitable for long time dynamics of the system. Also, defining $F^{1/2} := fL_\circ/\sqrt{gH_\circ} = \mathcal{O}(1)$, we choose (cf. [11])

$$Ro = \varepsilon \ll 1, \qquad Fr = F^{1/2}\varepsilon, \qquad \Theta = F\varepsilon. \tag{3}$$

This is the so-called *quasi-geostrophic distinguished limit*, i.e., the Rossby and Froude numbers are small, there is an exact balance between pressure gradient and the Coriolis force, and the variation of the bottom topography and surface perturbation are very mild compared to the height of the water column, owing to $\Theta \sim \varepsilon$, i.e., $\|z\|, \|\nabla_{\boldsymbol{x}}\eta^b\| = \mathcal{O}(\varepsilon)$ (see [11, 12]). Using that, and with similar notations as [4, 13, 14], we can re-write (2) as (after suppressing hats):

$$\begin{aligned}
&\partial_t z + \frac{1}{\Theta}\nabla_{\boldsymbol{x}} \cdot \boldsymbol{m} = 0,\\
&\partial_t \boldsymbol{m} + \nabla_{\boldsymbol{x}} \cdot \left(\frac{\boldsymbol{m} \otimes \boldsymbol{m}}{\Theta z - b} + \frac{\Theta z^2 - 2bz}{2\varepsilon}\mathbb{I}_2 \right) = -\frac{1}{\varepsilon} z \nabla_{\boldsymbol{x}} b - \frac{1}{\varepsilon}\boldsymbol{m}^\perp,
\end{aligned} \tag{4}$$

where $\boldsymbol{m} := (\Theta z - b)\boldsymbol{u}$ is the momentum vector and b is the dimensionless water depth measured from H_{mean} (scaled by $H_\circ$) with a negative sign, i.e., $1 - \Theta\eta^b = -b$.

It is proved in [11] that for $\varepsilon \to 0$ the system (4) converges to the quasi-geostrophic equations (QGE):

$$\boldsymbol{u}_{(0)} = \nabla_{\boldsymbol{x}}^\perp z_{(0)} \quad \Longleftrightarrow \quad \Delta_{\boldsymbol{x}} z_{(0)} = \zeta_{(0)} \tag{5a}$$

$$\left(\partial_t + \boldsymbol{u}_{(0)} \cdot \nabla_{\boldsymbol{x}}\right)\left(\zeta_{(0)} - F z_{(0)} + F\eta^b_{(0)}\right) = 0 \tag{5b}$$

where subscript (0) stands for the leading order term in the Poincaré expansion, ζ is the magnitude of the *vorticity* $\nabla_{\boldsymbol{x}} \times \boldsymbol{u}$ and $\nabla_{\boldsymbol{x}}^\perp := (-\partial_{x_2}, \partial_{x_1})$. Equation (5a) means that the solution is at *geostrophic equilibrium* locally in time. It also implies that the surface perturbation $z_{(0)}$ can be read as the stream function ψ, i.e., $\nabla_{\boldsymbol{x}}^\perp \psi = \boldsymbol{u}_{(0)}$; so, the velocity field is solenoidal. Equation (5b) is the conservation of the (leading order of the) *potential vorticity* $\xi := \zeta_{(0)} - F z_{(0)} + F b_{(0)}$ while $\zeta_{(0)}$ is given by (5a).

2.1 RS-IMEX Scheme for the Rotating Shallow Water System

Consider the RSWE (4) in the form $\partial_t \boldsymbol{U} + \nabla_{\boldsymbol{x}} \cdot \boldsymbol{F} = \boldsymbol{S}^B + \boldsymbol{S}^C$. The main idea of the RS-IMEX scheme is to decompose the solution $\boldsymbol{U}$ as $\overline{\boldsymbol{U}} + \boldsymbol{V}$, where $\overline{\boldsymbol{U}}$ is a chosen reference solution and $\boldsymbol{V}$ is its perturbation. We pick $\overline{\boldsymbol{U}}$ as a solution which is asymptotically close to $\boldsymbol{U}$, e.g., the solution of the incompressible Euler equations for the Euler equations, or the QGE for the RSWE. Then, we use a Taylor expansion around $\overline{\boldsymbol{U}}$ to split the flux and source terms into reference ($\overline{\boldsymbol{F}}$, $\overline{\boldsymbol{S}}$), linear stiff ($\widetilde{\boldsymbol{F}}$, $\widetilde{\boldsymbol{S}}$) and non-linear non-stiff parts ($\widehat{\boldsymbol{F}}$, $\widehat{\boldsymbol{S}}$):

$$\begin{aligned}
\boldsymbol{F}(\boldsymbol{U}) &= \boldsymbol{F}(\overline{\boldsymbol{U}}) + \boldsymbol{F}'(\overline{\boldsymbol{U}})\boldsymbol{V} + \big(\boldsymbol{F}(\boldsymbol{U}) - \boldsymbol{F}(\overline{\boldsymbol{U}}) - \boldsymbol{F}'(\overline{\boldsymbol{U}})\boldsymbol{V}\big) =: \overline{\boldsymbol{F}} + \widetilde{\boldsymbol{F}} + \widehat{\boldsymbol{F}},\\
\boldsymbol{S}(\boldsymbol{U}) &= \boldsymbol{S}(\overline{\boldsymbol{U}}) + \boldsymbol{S}'(\overline{\boldsymbol{U}})\boldsymbol{V} + \big(\boldsymbol{S}(\boldsymbol{U}) - \boldsymbol{S}(\overline{\boldsymbol{U}}) - \boldsymbol{S}'(\overline{\boldsymbol{U}})\boldsymbol{V}\big) =: \overline{\boldsymbol{S}} + \widetilde{\boldsymbol{S}} + \widehat{\boldsymbol{S}}.
\end{aligned}$$

Then, one is left with the following system for the perturbation $\boldsymbol{V} = (v_1, v_2, v_3)^T$:

$$\partial_t \boldsymbol{V} + \nabla_{\boldsymbol{x}} \cdot \big(\widetilde{\boldsymbol{F}}(\overline{\boldsymbol{U}}, \boldsymbol{V}) + \widehat{\boldsymbol{F}}(\overline{\boldsymbol{U}}, \boldsymbol{V})\big) = \widetilde{\boldsymbol{S}}(\overline{\boldsymbol{U}}, \boldsymbol{V}) + \widehat{\boldsymbol{S}}(\overline{\boldsymbol{U}}, \boldsymbol{V}) - \overline{\boldsymbol{T}}(\overline{\boldsymbol{U}}), \tag{6}$$

where $\overline{\boldsymbol{T}}(\overline{\boldsymbol{U}})$ is the residual of the reference solution and reads

$$\overline{\boldsymbol{T}}(\overline{\boldsymbol{U}}) := \partial_t \overline{\boldsymbol{U}} + \nabla_{\boldsymbol{x}} \cdot \overline{\boldsymbol{F}}(\overline{\boldsymbol{U}}) - \overline{\boldsymbol{S}}(\overline{\boldsymbol{U}}). \tag{7}$$

For the RSWE (4) and with $\boldsymbol{U} = (z, m_1, m_2)^T$, one can identify $\boldsymbol{F}$, $\boldsymbol{S}^B$ and $\boldsymbol{S}^C$ as

$$\boldsymbol{F} = \begin{bmatrix} m_1/\Theta & m_2/\Theta \\ \dfrac{m_1^2}{\Theta z - b} + \dfrac{\Theta z^2 - 2zb}{2\varepsilon} & \dfrac{m_1 m_2}{\Theta z - b} \\ \dfrac{m_1 m_2}{\Theta z - b} & \dfrac{m_2^2}{\Theta z - b} + \dfrac{\Theta z^2 - 2zb}{2\varepsilon} \end{bmatrix}, \boldsymbol{S}^B = \begin{bmatrix} 0 \\ -z b_x/\varepsilon \\ -z b_y/\varepsilon \end{bmatrix}, \boldsymbol{S}^C = \begin{bmatrix} 0 \\ m_2/\varepsilon \\ -m_1/\varepsilon \end{bmatrix}.$$

Assuming the reference solution $\overline{\boldsymbol{U}} = (\overline{z}, \overline{m_1}, \overline{m_2})^T$ to be solution of the QGE, the RS-IMEX procedure gives the stiff part as:

$$\widetilde{\boldsymbol{F}} = \begin{bmatrix} v_2/\Theta & v_3/\Theta \\ -\dfrac{\overline{m_1}^2 v_1 \Theta}{(\Theta\overline{z} - b)^2} + \dfrac{2\overline{m_1} v_2}{\Theta\overline{z} - b} + \dfrac{(\Theta\overline{z} - b)}{\varepsilon} v_1 & -\dfrac{\overline{m_1 m_2} v_1 \Theta}{(\Theta\overline{z} - b)^2} + \dfrac{\overline{m_1} v_3}{\Theta\overline{z} - b} + \dfrac{\overline{m_2} v_2}{\Theta\overline{z} - b} \\ -\dfrac{\overline{m_1 m_2} v_1 \Theta}{(\Theta\overline{z} - b)^2} + \dfrac{\overline{m_1} v_3}{\Theta\overline{z} - b} + \dfrac{\overline{m_2} v_2}{\Theta\overline{z} - b} & -\dfrac{\overline{m_2}^2 v_1 \Theta}{(\Theta\overline{z} - b)^2} + \dfrac{2\overline{m_2} v_3}{\Theta\overline{z} - b} + \dfrac{(\Theta\overline{z} - b)}{\varepsilon} v_1 \end{bmatrix},$$

$$\widetilde{\boldsymbol{S}}^B = \begin{bmatrix} 0 \\ -v_1 b_x/\varepsilon \\ -v_1 b_y/\varepsilon \end{bmatrix}, \qquad \widetilde{\boldsymbol{S}}^C = \begin{bmatrix} 0 \\ v_3/\varepsilon \\ -v_2/\varepsilon \end{bmatrix},$$

while $\widehat{\boldsymbol{S}}^B = \widehat{\boldsymbol{S}}^C = \boldsymbol{0}$ and $\widehat{\boldsymbol{F}}(\overline{\boldsymbol{U}}, \boldsymbol{V}) = \boldsymbol{F}(\overline{\boldsymbol{U}} + \boldsymbol{V}) - \overline{\boldsymbol{F}}(\overline{\boldsymbol{U}}) - \widetilde{\boldsymbol{F}}(\overline{\boldsymbol{U}}, \boldsymbol{V})$. One can verify that the Jacobian matrices $\widehat{\boldsymbol{F}}'$ and $\widetilde{\boldsymbol{F}}'$ have complete sets of eigenvectors and that the eigenvalues of $\widehat{\boldsymbol{F}}'$ are non-stiff. This can be readily seen from the expression of the non-stiff flux $\widehat{\boldsymbol{F}}_1$ (and similarly $\widehat{\boldsymbol{F}}_2$)

$$\widehat{\boldsymbol{F}}_1 = \begin{bmatrix} 0 \\ \frac{m_1^2}{\Theta z - b} + \frac{\Theta z^2 - 2zb}{2\varepsilon} - \frac{\overline{m_1}^2}{\Theta\bar{z} - b} - \frac{\Theta\bar{z}^2 - 2\bar{z}b}{2\varepsilon} + \frac{\overline{m_1}^2 v_1 \Theta}{(\Theta\bar{z} - b)^2} - \frac{2\overline{m_1} v_2}{\Theta\bar{z} - b} - \frac{(\Theta\bar{z} - b)}{\varepsilon} v_1 \\ \frac{m_1 m_2}{\Theta z - b} - \frac{\overline{m_1 m_2}}{\Theta\bar{z} - b} + \frac{\overline{m_1 m_2} v_1 \Theta}{(\Theta\bar{z} - b)^2} - \frac{\overline{m_1} v_3}{\Theta\bar{z} - b} - \frac{\overline{m_2} v_2}{\Theta\bar{z} - b} \end{bmatrix}, \quad (8)$$

as, after simplification, it does not contain any $\mathcal{O}(1/\varepsilon)$ term.

The RS-IMEX approach discretizes (6) for each cell $(i, j) \in \{1, 2, \ldots, N\}^2$ in a square computational domain Ω_N with spatial steps $\Delta x_1 = \Delta x_2$ and time step Δt:

$$\boldsymbol{V}_{ij}^{n+1} = \boldsymbol{V}_{ij}^n - \Delta t \overline{\boldsymbol{T}}_{ij}^{n+1} + \Delta t \left(-\nabla_{h,\boldsymbol{x}} \cdot \widetilde{\boldsymbol{F}}_{ij} + \widetilde{\boldsymbol{S}}_{ij}^B + \widetilde{\boldsymbol{S}}_{ij}^C \right)^{n+1} - \Delta t \left(\nabla_{h,\boldsymbol{x}} \cdot \widehat{\boldsymbol{F}}_{ij} \right)^n,$$

with central difference gradient operator $\nabla_{h,\boldsymbol{x}}$, Rusanov-type fluxes and central discretization of source terms. Defining $\Delta_{h,\boldsymbol{x}}$ as the central discretization of the Laplacian, one can write $\overline{\boldsymbol{T}}_{ij}^{n+1}$ as (with $\overline{\alpha}$ denoting the numerical diffusion coefficient)

$$\overline{\boldsymbol{T}}_{ij}^{n+1} = \frac{\overline{\boldsymbol{U}}_{ij}^{n+1} - \overline{\boldsymbol{U}}_{ij}^n}{\Delta t} + \nabla_{h,x_1} \overline{\boldsymbol{F}}_{1,ij}^{n+1} + \nabla_{h,x_2} \overline{\boldsymbol{F}}_{2,ij}^{n+1} - \left(\overline{\boldsymbol{S}}^B + \overline{\boldsymbol{S}}^C \right)_{ij}^{n+1} - \frac{\overline{\alpha} \Delta x_1}{2} \Delta t \Delta_{h,\boldsymbol{x}} \boldsymbol{V}_{ij}^{n+1}.$$

Having this, the RS-IMEX scheme can be written as

$$\begin{aligned} \boldsymbol{V}_{ij}^{n+1/2} &= \boldsymbol{V}_{ij}^n - \Delta t \nabla_{h,x_1} \widehat{\boldsymbol{F}}_{1,ij}^n - \Delta t \nabla_{h,x_2} \widehat{\boldsymbol{F}}_{2,ij}^n + \frac{\widehat{\alpha} \Delta x_1}{2} \Delta t \Delta_{h,\boldsymbol{x}} \boldsymbol{V}_{ij}^n, \\ \boldsymbol{V}_{ij}^{n+1} &= \boldsymbol{V}_{ij}^{n+1/2} - \Delta t \nabla_{h,x_1} \widetilde{\boldsymbol{F}}_{1,ij}^{n+1} - \Delta t \nabla_{h,x_2} \widetilde{\boldsymbol{F}}_{2,ij}^{n+1} + \frac{\widetilde{\alpha} \Delta x_1}{2} \Delta t \Delta_{h,\boldsymbol{x}} \boldsymbol{V}_{ij}^{n+1} \\ &\quad + \Delta t \left(\widetilde{\boldsymbol{S}}^B + \widetilde{\boldsymbol{S}}^C \right)_{ij}^{n+1} - \Delta t \overline{\boldsymbol{T}}_{ij}^{n+1}, \end{aligned} \quad (9)$$

where $\widehat{\alpha}$ is chosen as the largest eigenvalue of $\widehat{\boldsymbol{F}}'$ over the domain, $\overline{\alpha} = \widehat{\alpha}$, and $\widetilde{\alpha} = \mathcal{O}(1)$ is computed for $\varepsilon = 1$ not to add excessive diffusion.

The remaining point to be clarified is how to solve the QGE to find the reference solution. We employ the Arakawa method [1] for this purpose: having ψ^n, we obtain ξ^{n+1} using the *Arakawa Jacobian*. Then, we solve for ψ^{n+1} and redo the procedure to correct the predicted solution. We refer the reader to consult [9].

3 Main Result

Theorem 1 (Formal asymptotic consistency of the RS-IMEX scheme) *Consider the rotating shallow water equations with topography (4), on a periodic domain and with well-prepared initial data* $(z_{0,\varepsilon}, \boldsymbol{u}_{0,\varepsilon})$ *such that*

$$z(0,\cdot) = z_{0,\varepsilon} = z^0_{(0)} + \varepsilon z^0_{(1),\varepsilon}, \qquad \boldsymbol{u}(0,\cdot) = \boldsymbol{u}_{0,\varepsilon} = \boldsymbol{u}^0_{(0)} + \varepsilon \boldsymbol{u}^0_{(1),\varepsilon},$$

where $(z^0_{(0)}, \boldsymbol{u}^0_{(0)})$ *is the solution of the QGE (5a)–(5b). Then, the RS-IMEX scheme (9) has a unique solution for all* $\varepsilon > 0$*, which is formally consistent with the QGE up to* $\mathscr{O}(\varepsilon)$*. So, it is asymptotically consistent and provides a consistent discretization of QGE.*

Sketch of the proof Let us assume $\widetilde{\alpha} = 0$ for simplicity. Regarding solvability, the linear system of the implicit step with the companion matrix J_ε can be written as $J_\varepsilon := \mathbb{I}_{3N^2} + \Delta t\, \Xi_\varepsilon$ with a Δt-independent matrix Ξ_ε. It is plausible to conclude that for a suitable choice of Δt, none of the eigenvalues of $\Delta t\, \Xi_\varepsilon$ is equal to -1; so, J_ε is non-singular, and the implicit step is solvable; see [13, 14] for details.

Regarding asymptotic consistency, we take a formal approach and put the Poincaré expansion into the scheme and balance the equal powers of ε to show that the solution is asymptotically consistent with the limit, formally.

Firstly, we show that the explicit step is ε-stable, i.e., $\|\boldsymbol{V}_\Delta^{n+1/2}\| = \mathscr{O}(1)$ given $\|\boldsymbol{V}_\Delta^{n}\| = \mathscr{O}(1)$ (which is compatible with the well-prepared initial data). Since $\widehat{\boldsymbol{F}}_{1,1} = \widehat{\boldsymbol{F}}_{2,1} = 0$, one can immediately conclude that $\|\boldsymbol{V}_{1,\Delta}^{n+1/2}\| = \mathscr{O}(1)$. Owing to (8), one can simply confirm that for $\boldsymbol{V}_{2,\Delta}$ (and similarly $\boldsymbol{V}_{3,\Delta}$)

$$\lim_{\varepsilon\to 0} \left(\nabla_{h,x_1} \widehat{\boldsymbol{F}}^n_{1,2,ij} + \nabla_{h,x_2} \widehat{\boldsymbol{F}}^n_{2,2,ij}\right) = \mathscr{O}(1),$$

recalling that $\Theta = F\varepsilon$. So, the explicit step does not change the leading order of $\boldsymbol{V}_{2,\Delta}$ (and $\boldsymbol{V}_{3,\Delta}$). This concludes the ε-stability proof of the explicit step.

Completing the asymptotic consistency analysis, we show that the implicit step is consistent with the limit. We *assume* that $\|\boldsymbol{V}_\Delta^{n+1}\| = \mathscr{O}(1)$ to justify the use of Poincaré expansion and will discuss this assumption somewhere else. From the v_1-update, $\overline{\boldsymbol{T}}$ and $\widetilde{\boldsymbol{F}}$, the momentum field (up to $\mathscr{O}(\varepsilon)$) is solenoidal, i.e.,

$$\nabla_{h,x_1} (\overline{m_1} + v_2)^{n+1}_{ij} + \nabla_{h,x_2} (\overline{m_2} + v_3)^{n+1}_{ij} = \mathscr{O}(\varepsilon).$$

Using v_2-update (similarly for v_3), one can balance $\mathscr{O}(1/\varepsilon)$ terms, which give

$$-\nabla_{h,x_1}\left(bv_{1(0)}+b\overline{z}\right)_{ij}^{n+1}=-\left(\overline{z}+v_{1(0)}\right)_{ij}^{n+1}\nabla_{h,x_1}b_{ij}+(\overline{m_2}+v_{3(0)})_{ij}^{n+1}. \tag{10}$$

This is a consistent discretization of (5a), since the bottom is almost flat, $\|\nabla_{h,\boldsymbol{x}}b_{ij}\|=\mathscr{O}(\varepsilon)$. In other words, (10) implies that

$$\nabla_{h,\boldsymbol{x}}\left(\overline{z}+v_{1(0)}\right)_{ij}^{n+1}=\boldsymbol{u}_{(0),ij}^{\perp,n+1}.$$

Thus, up to $\mathscr{O}(\varepsilon)$ terms, the solution is consistent with the limit. Using this, the asymptotic consistency of the scheme can be concluded. □

4 Numerical Example

We discuss the 2d stationary (and non quasi-geostrophic) vortex in the periodic domain $[0,1)^2$ as in [3]. The initial data in the polar co-ordinates write

$$\boldsymbol{u}_0(r,\theta)=\vartheta_\theta(r)\hat{\boldsymbol{\theta}},\quad z_0'(r)=\vartheta_\theta(r)+\frac{\varepsilon}{r}\vartheta_\theta^2(r),\quad \vartheta_\theta(r):=5r\mathbf{1}_{[r<\frac{1}{5}]}+(2-5r)\mathbf{1}_{[\frac{1}{5}\le r<\frac{2}{5}]},$$

where r is the distance to $(0.5,0.5)^T$.

Figure 1 indicates the time evolution of relative perturbation from the equilibrium, on the 30×30 grid and for different ε. The time step is chosen as the smaller value the Arakawa method and the (non-stiff) explicit step require, with CFL $=0.45$; so, Δt is uniform in ε. It appears that, like for the scheme devised in [3], the error decreases with ε. Also in Fig. 2, the perturbation from the equilibrium has been plotted for different ε and longer times, which implies that the scheme provides accurate results particularly for $\varepsilon\ll 1$. Note that unlike [3], no well-balancing mechanism has been incorporated into the scheme; so, for the $\varepsilon=\mathscr{O}(1)$ regime, one should expect less

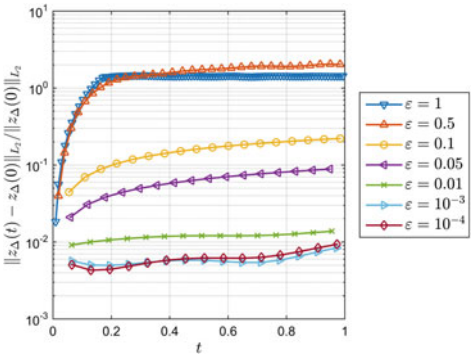
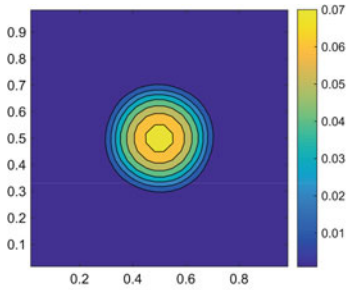
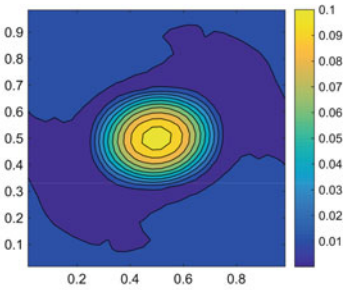


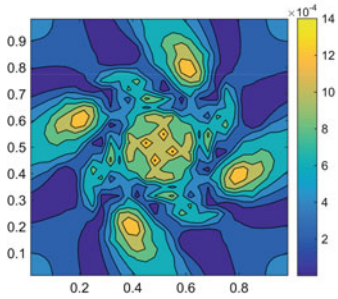
Fig. 1 Evolution of the relative error for surface perturbation in time, for the RS-IMEX scheme on the 30×30 grid, and for different ε. z_Δ stands for the numerically computed surface perturbation



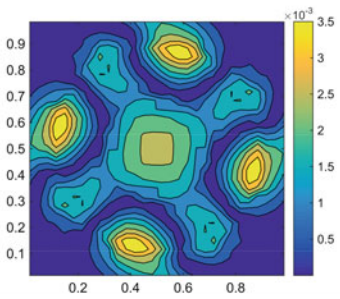
(a) $t = 1$ and $\varepsilon = 0.1$.



(b) $t = 20$ and $\varepsilon = 0.1$.



(c) $t = 1$ and $\varepsilon = 10^{-4}$.



(d) $t = 20$ and $\varepsilon = 10^{-4}$.

Fig. 2 Perturbation from the equilibrium, $|z_\Delta(t) - z_\Delta(0)|$, for the RS-IMEX scheme, computed on the 30×30 grid and for $\varepsilon = 0.1, 10^{-4}$ and $t = 1, 20$

accurate results, compared to [3]. We emphasize that this example does fit into the assumptions of Theorem 1, as the initial condition and the solution are close to the limit manifold for $\varepsilon \ll 1$.

Acknowledgements The research was supported by RWTH Aachen University through *Graduiertenförderung nach Richtlinien zur Förderung des wissenschaftlichen Nachwuchses (RFwN).*

References

1. Arakawa, A.: Computational design for long-term numerical integration of the equations of fluid motion: two-dimensional incompressible flow. Part I. J. Comput. Phys. **1**(1), 119–143 (1966)
2. Audusse, E., Dellacherie, S., Do Minh Hieu, P.O., Penel, Y.: Godunov type scheme for the linear wave equation with Coriolis source term. HAL: hal-01254888 (2015)
3. Audusse, E., Klein, R., Nguyen, D., Vater, S.: Preservation of the discrete geostrophic equilibrium in shallow water flows. In: Finite Volumes for Complex Applications VI Problems & Perspectives, pp. 59–67. Springer (2011)

4. Bispen, G., Arun, K.R., Lukáčová-Medvid'ová, M., Noelle, S.: IMEX large time step finite volume methods for low Froude number shallow water flows. Commun. Comput. Phys. **16**, 307–347 (2014)
5. Bouchut, F., Le Sommer, J., Zeitlin, V.: Frontal geostrophic adjustment and nonlinear wave phenomena in one-dimensional rotating shallow water. Part 2. High-resolution numerical simulations. J. Fluid Mech. **514**, 35–63 (2004)
6. Buckingham, E.: Model experiments and the forms of empirical equations. Trans. Am. Soc. Mech. Eng. **37**, 263–296 (1915)
7. Charney, J.G.: On the scale of atmospheric motions. Geofysiske Publikasjoner **17**(2) (1948)
8. Hu, J., Jin, S., Li, Q.: Asymptotic-preserving schemes for multiscale hyperbolic and kinetic equations. Handb. Numer. Anal. (2016)
9. Kacimi, A., Aliziane, T., Khouider, B.: The Arakawa Jacobian method and a fourth-order essentially nonoscillatory scheme for the beta-plane barotropic equations. Int. J. Numer. Anal. Model. (2011)
10. Lukáčová-Medvid'ová, M., Noelle, S., Kraft, M.: Well-balanced finite volume evolution Galerkin methods for the shallow water equations. J. Comput. Phys. **221**(1), 122–147 (2007)
11. Majda, A.: Introduction to PDEs and waves for the atmosphere and Ocean. Am. Math. Soc. 9 (2003)
12. Pedlosky, J.: Geophysical Fluid Dynamics. Springer Science & Business Media (2013)
13. Zakerzadeh, H.: Asymptotic analysis of the RS-IMEX scheme for the shallow water equations in one space dimension. IGPM report 455, RWTH Aachen University (2016). Submitted for publication
14. Zakerzadeh, H.: Asymptotic consistency of the RS-IMEX scheme for the low-Froude shallow water equations: analysis and numerics. In: Proceedings of XVI International Conference on Hyperbolic Problems, Aachen (2016)

Analysis of Apparent Topography Scheme for the Linear Wave Equation with Coriolis Force

Emmanuel Audusse, Minh Hieu Do, Pascal Omnes and Yohan Penel

Abstract The shallow water equations can be used to model many phenomena in geophysical fluid mechanics. For large scales, the Coriolis force plays an important role and the geostrophic equilibrium which corresponds to the balance between the pressure gradient and the Coriolis force is an important feature. In this communication, we investigate the stability condition and the behavior of the so-called Apparent Topography scheme which is capable of capturing a discrete version of the geostrophic equilibrium.

Keywords Shallow water flows · Finite volume method · Coriolis force · Well-balanced schemes

MSC (2010): 65M08 · 65M12 · 76U05

E. Audusse · M.H. Do (✉) · P. Omnes
LAGA, CNRS UMR 7539, Université Paris 13, Sorbonne Paris Cité, 99 av. J.-B. Clément, 94430 Villetaneuse, France
e-mail: do@math.univ-paris13.fr

E. Audusse
e-mail: audusse@math.univ-paris13.fr

P. Omnes
e-mail: omnes@math.univ-paris13.fr; pascal.omnes@cea.fr

P. Omnes
CEA Saclay, DEN, DM2S, STMF, LMSF, 91191 Gif-sur-Yvette Cedex, France

Y. Penel
Team ANGE, Inria–Paris, CEREMA, UPMC and CNRS, 2 Rue Simone Iff, CS 42112, 75589 Paris Cedex 12, France
e-mail: yohan.penel@inria.fr

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_23

1 Introduction

In order to study the Shallow Water equations with Coriolis source term, we consider the dimensionless formulation on the rotating frame which is given by

$$\begin{cases} \mathsf{St}\partial_t h + \nabla \cdot (h\bar{\mathbf{u}}) = 0, & \text{(1a)} \\ \mathsf{St}\partial_t (h\bar{\mathbf{u}}) + \nabla \cdot (h\bar{\mathbf{u}} \otimes \bar{\mathbf{u}}) + \dfrac{1}{\mathsf{Fr}^2}\nabla\left(\dfrac{h^2}{2}\right) = -\dfrac{1}{\mathsf{Fr}^2} h\nabla b - \dfrac{1}{\mathsf{Ro}} h\bar{\mathbf{u}}^{\perp}, & \text{(1b)} \end{cases}$$

where unknowns h and $\bar{\mathbf{u}}$ respectively denote the water depth and the average velocity over the water column and function $b(\mathbf{x})$ denotes the topography of the considered oceanic basin and is a given function. Dimensionless numbers St, Fr and Ro respectively stand for the Strouhal, the Froude and the Rossby numbers defined below. In the sequel, we shall focus on cases where

$$\mathsf{St} := \frac{L}{UT} = \mathscr{O}\left(\frac{1}{\mathsf{M}}\right), \quad \mathsf{Fr} := \frac{U}{\sqrt{gH}} = \mathscr{O}(\mathsf{M}), \quad \mathsf{Ro} := \frac{U}{\Omega L} = \mathscr{O}(\mathsf{M}),$$

with M a small parameter. The parameters g and Ω denote the gravity coefficient and the angular velocity of the Earth. Constants U, H, L and T are some characteristic velocity, vertical and horizontal lengths and time. These orders of magnitude correspond to the study of short-time dynamics and standard conditions for large scale oceanic flows.

For data independent of y and with a flat topography, the solution of System (1) then satisfies at the leading order the quasi-1d linear wave equation with Coriolis source term (see [2] for the derivation)

$$\begin{cases} \partial_t r + a_\star \partial_x u = 0, \\ \partial_t u + a_\star \partial_x r = \omega v, \\ \partial_t v = -\omega u, \end{cases} \tag{2}$$

where $a_\star$ and ω are constants of order $\mathscr{O}(1)$—respectively related to the wave velocity and to the rotating velocity—r is the first order perturbation of the water depth h and (u, v) is the leading order for the velocity field. The stationary state corresponding to System (2) is the 1d version of the so-called *geostrophic equilibrium* and is given by

$$u = 0, \quad a_\star \partial_x r = \omega v. \tag{3}$$

A first study of the accuracy of numerical schemes applied to system (2) for initial data that are close to the kernel (3) was performed in [2]. It was shown that the standard Godunov scheme applied to the linear wave equation with Coriolis source term is inaccurate at low Froude number and the numerical viscosity on the pressure equation is the main reason for this inaccuracy. A modified *low Froude* Godunov scheme was proposed to cure the problem. The scheme was shown to be L^2 stable

under a suitable CFL condition. The proofs extend the ideas introduced in [5] for the study of the homogeneous wave equations in *low Mach* number regimes.

In this paper, our objective is to study in the same context the numerical scheme introduced in [3] as a well-balanced (WB) scheme for the Shallow Water equations with Coriolis source term (1). In particular we prove the L^2 stability of the scheme under suitable CFL conditions. Moreover we compare this scheme, called *apparent topography* scheme in the following, and the *low Froude* one in terms of dispersion relations and accuracy for some test cases. Note that a high order extension of the *apparent topography* scheme for the non-linear SW equations with Coriolis source term has been studied in [4], where the authors also paid attention to the linear dispersion relation (hence related to (2)).

2 The Numerical Schemes

Both *low Froude* and *apparent topography* schemes are collocated finite volume schemes and can be interpreted as a way to modify the numerical diffusion of the classical Godunov scheme on the pressure equation. In the *low Froude* scheme proposed in [2], the numerical diffusion on the pressure equation is simply deleted. In the *apparent topography* scheme introduced in [3], the diffusion term of the classical Godunov scheme remains and an additional consistent term is introduced in the pressure equation such that the numerical diffusion vanishes when applied to an element of the linear kernel (3). The name *apparent topography* comes from the fact that the scheme was first developed in the context of WB methods for the shallow water equation with topography, see [1]. The two aforementioned semi-discrete schemes applied to (2) read

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} r_j + a_\star \frac{u_{j+1}-u_{j-1}}{2\Delta x} - \frac{\kappa_r a_\star \Delta x}{2} \frac{r_{j+1}-2r_j+r_{j-1}}{\Delta x^2} + \frac{\kappa_r \omega}{2} \frac{v_{j+1}-v_{j-1}}{2} = 0, \\ \frac{\mathrm{d}}{\mathrm{d}t} u_j + a_\star \frac{r_{j+1}-r_{j-1}}{2\Delta x} - \frac{\kappa_u a_\star \Delta x}{2} \frac{u_{j+1}-2u_j+u_{j-1}}{\Delta x^2} = \omega f(v_{j-1}, v_j, v_{j+1}), \\ \frac{\mathrm{d}}{\mathrm{d}t} v_j = -\omega f(u_{j-1}, u_j, u_{j+1}). \end{cases} \tag{4}$$

where the *low Froude* scheme corresponds to the choice $\kappa_r = 0$, $f(x, y, z) = y$ and the *apparent topography* scheme to the choice $\kappa_r = \kappa_u$, $f(x, y, z) = \frac{x+2y+z}{4}$.

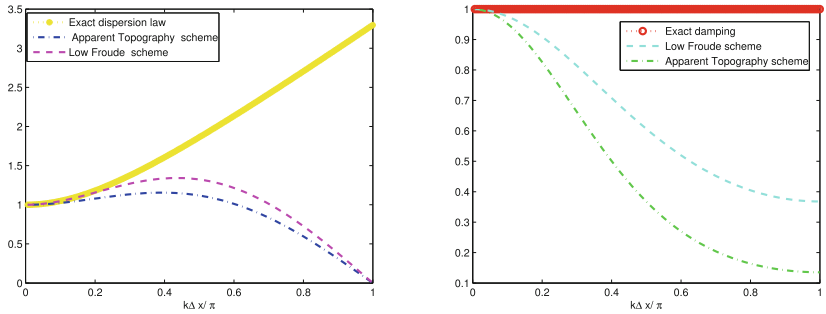
2.1 Study of the Semi–discrete Scheme–Dispersion Relations

We now study the stability of the semi-discrete Godunov type schemes by means of Fourier modes:

$$r_j(t) = \varphi_r(t) e^{ikx_j}, \quad u_j(t) = \varphi_u(t) e^{ikx_j} \quad \text{and} \quad v_j(t) = \varphi_v(t) e^{ikx_j}.$$

Table 1 The eigenvalues corresponding to the inertia-gravity modes for small $k\Delta x$

Wave equation	$\pm i\sqrt{a_\star^2 k^2 + \omega^2}$
Low Froude	$a_\star \frac{\kappa_u}{2} \frac{\sin^2(\frac{k\Delta x}{2})}{\frac{\Delta x}{2}} \pm i\sqrt{a_\star^2 \left(\frac{\sin(k\Delta x/2)}{\Delta x/2}\right)^2 \left[\cos^2(\frac{k\Delta x}{2}) - \left(\frac{\kappa_u}{2}\right)^2 \sin^2(\frac{k\Delta x}{2})\right] + \omega^2}$
Apparent Topography	$a_\star \kappa_u \frac{\sin^2(\frac{k\Delta x}{2})}{\frac{\Delta x}{2}} \pm i\sqrt{a_\star^2 \left(\frac{\sin(k\Delta x)}{\Delta x}\right)^2 + \omega^2 \left(\frac{1+\cos(k\Delta x)}{2}\right)^2}$



(a) Dispersion relation $\Im(\lambda)/\omega$ (b) Damping error $e^{-\Re(\lambda)}$

Fig. 1 Numerical properties of the semi-discrete schemes with the Rossby deformation $R_d := \frac{a_\star}{\omega} = \Delta x$ and $(\kappa_r, \kappa_u) = (0, 1)$ for LF, $(\kappa_r, \kappa_u) = (1, 1)$ for AT

Substituting these expressions into (4), we obtain the following linear system of differential equations

$$\begin{pmatrix} \varphi_r'(t) \\ \varphi_u'(t) \\ \varphi_v'(t) \end{pmatrix} + \begin{pmatrix} \kappa_r a_\star \frac{\sin^2(\frac{k\Delta x}{2})}{\frac{\Delta x}{2}} & i a_\star \frac{\sin(k\Delta x)}{\Delta x} & i \frac{\kappa_r \omega \Delta x}{2} \frac{\sin(k\Delta x)}{\Delta x} \\ i a_\star \frac{\sin(k\Delta x)}{\Delta x} & \kappa_u a_\star \frac{\sin^2(k\Delta x/2)}{\Delta x/2} & -\omega\zeta \\ 0 & \omega\zeta & 0 \end{pmatrix} \begin{pmatrix} \varphi_r(t) \\ \varphi_u(t) \\ \varphi_v(t) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad (5)$$

where $\zeta = 1$ for the *low Froude* scheme and $\zeta = \cos^2(\frac{k\Delta x}{2})$ for the *apparent topography* scheme. The first eigenvalue of the amplification matrix is $\lambda = 0$, corresponding to the discrete stationary state (3). The other two, corresponding to the inertia-gravity modes, are given in Table 1. Their real part $\Re(\lambda)$ characterizes the decay of Fourier modes k. Since $\Re(\lambda) > 0$, both *low Froude* and *apparent topography* schemes are damping. The damping rate of the *apparent topography* scheme is twice larger than that of the *low Froude* scheme. The imaginary part $\Im(\lambda)$ characterizes the propagation properties of the Fourier modes. Note that for the *low Froude* scheme, the eigenvalues may be real for $k\Delta x$ close to π which means the corresponding modes do not propagate and are only damped. For small $a_* k/\omega$, the dispersion relation $\Im(\lambda)/\omega$ of the *low Froude* scheme is closer to the exact one (for the rotating wave Eq. (2)) whereas the converse holds for large $a_* k/\omega$ (Fig. 1).

2.2 Study of the Fully Discrete Scheme: Kernel and L^2-Stability

The fully discrete *apparent topography* scheme applied to (2) can be written as

$$\begin{cases} \frac{r_j^{n+1}-r_j^n}{\Delta t} + a_\star \frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x} - \frac{\kappa_r a_* \Delta x}{2} \frac{r_{j+1}^n - 2r_j^n + r_{j-1}^n}{\Delta x^2} + \frac{\kappa_r \omega}{2} \frac{(v_{j+1}^n - v_{j-1}^n)}{2} = 0, \\ \frac{u_j^{n+1}-u_j^n}{\Delta t} + a_\star \frac{r_{j+1}^n - r_{j-1}^n}{2\Delta x} - \frac{\kappa_u a_* \Delta x}{2} \frac{u_{j+1}^n - 2u_j^n + u_{j-1}^n}{\Delta x^2} \\ \qquad = \omega \left[\theta_1 \frac{v_{j+1}^n + 2v_j^n + v_{j-1}^n}{4} + (1-\theta_1) \frac{v_{j+1}^{n+1} + 2v_j^{n+1} + v_{j-1}^{n+1}}{4} \right], \\ \frac{v_j^{n+1}-v_j^n}{\Delta t} = -\omega \left[\theta_2 \frac{u_{j+1}^n + 2u_j^n + u_{j-1}^n}{4} + (1-\theta_2) \frac{u_{j+1}^{n+1} + 2u_j^{n+1} + u_{j-1}^{n+1}}{4} \right] \end{cases} \tag{6}$$

for $j \in \{1, \dots, N\}$ and $0 \le \theta_1, \theta_2 \le 1$. Setting $q = (r, u, v)$, periodic boundary conditions read $q_0^{n+1} = q_N^{n+1}$, $q_{N+1}^{n+1} = q_1^{n+1}$. For practical reasons, we assume that the cell number N is odd.

Lemma 1 *The kernel of the Apparent Topography scheme* (6) *is given by*

$$\mathscr{E}_{\omega \neq 0}^h = \mathbf{ker} L_{\kappa,h} = \left\{ q = (r, u, v) \;\middle|\; u_j = 0,\ a_\star \frac{r_{j+1} - r_j}{\Delta x} = \omega \frac{v_{j+1} + v_j}{2} \right\}. \tag{7}$$

Proof A stationary state has to satisfy relations (6) with $q^{n+1} = q^n$. In particular, the third equation, with an odd number of points and given periodic boundary conditions, leads to

$$u_j^n = 0, \quad \forall j \in \{0, \dots, N+1\}.$$

We then deduce from the first two relations that

$$a_\star \frac{r_{j+1}^n - 2r_j^n + r_{j-1}^n}{\Delta x} = \omega \frac{v_{j+1}^n - v_{j-1}^n}{2} \quad \text{and} \quad a_\star \frac{r_{j+1}^n - r_{j-1}^n}{\Delta x} = \omega \frac{v_{j+1}^n + 2v_j^n + v_{j-1}^n}{2}.$$

Summing the two equations yields the discrete kernel (7). Conversely, any element satisfying (7) is a stationary state of relations (6). This discrete kernel is a consistent discretisation, defined at the cell interfaces, of the continuous kernel (3).

Remark 1 Let us recall that the discrete kernel of the *low Froude* scheme is

$$u_j = 0, \qquad a_\star \frac{r_{j+1} - r_{j-1}}{2\Delta x} = \omega v_j$$

(see [2]) which is another consistent discretisation, defined at the cell centers, of the continuous kernel (3).

Remark 2 When the number of points is even, checkerboard modes for velocity u may exist in the discrete kernel of the *apparent topography* scheme. Note that the *low Froude* scheme suffers the same drawback, but for the pressure r.

We will now investigate the L^2 stability of the *apparent topography* scheme. Let us first mention that when $0 < \theta_1, \theta_2 < 1$, the *apparent topography* scheme requires to solve a linear system at each time step, which leads to an additional computational cost. On the other hand, the case $\theta_1 = \theta_2 = 1$, that corresponds to a fully explicit scheme, is known to be unstable—see [4]. Therefore, we restrict our study to the two cases $\theta_1 = 0, \theta_2 = 1$ and $\theta_1 = 1, \theta_2 = 0$. Note that in [2], the L^2 stability of the *low Froude* scheme was studied for all values of $(\theta_1, \theta_2) \in [0, 1]^2$.

Lemma 2 *Under the hypothesis*

$$\kappa_r \kappa_u \leq 1 + \frac{\omega^2 \Delta x^2}{4 a_\star^2}, \tag{8}$$

the apparent topography scheme is L^2 stable under the CFL condition

$$\Delta t \leq \min\{\Delta t_a, \Delta t_b, \Delta t_c\}$$

where

$$\Delta t_a := \begin{cases} \dfrac{-\frac{|a_\star|}{\Delta x} + \sqrt{\frac{a_\star^2}{\Delta x^2} + (\kappa_r + \kappa_u)\kappa_r \omega^2}}{\kappa_r \omega^2} & \text{if } \kappa_r \neq 0, \\ \dfrac{\kappa_u}{2} \dfrac{\Delta x}{|a_\star|} & \text{otherwise.} \end{cases}$$

and

$$\Delta t_b := min \left\{ \frac{1}{\kappa_r}, \frac{1}{\kappa_u} \right\} \frac{\Delta x}{|a_\star|} \quad , \qquad \Delta t_c := \frac{2}{\omega}$$

Remark 3 Note that the choice $\kappa_r = 0$ is similar to the *low Froude* scheme, but with a discretisation of the Coriolis term at the interfaces. We then retrieve the same CFL condition as that of the cell-centered *low Froude* scheme, see [2].

Remark 4 Hypothesis (8) is not restrictive since the *low Froude* scheme always satisfies this condition and the classical choice for the *apparent topography* scheme is to take $\kappa_r = \kappa_u = 1$.

Remark 5 The bound Δt_c is the classical CFL condition for the inertial oscillations phenomenon.

Remark 6 The bound Δt_b is one of the classical CFL conditions for the problem without rotation. For $\Delta x \ll 1$, the asymptotic expansion of the bound Δt_a leads to the other classical CFL condition for the problem without rotation

$$\Delta t_a = \frac{\kappa_r + \kappa_u}{2} \frac{\Delta x}{|a_\star|}.$$

Proof We perform a Von Neumann analysis to investigate the stability condition. Let us denote

$$\sigma = \frac{\Delta t}{\Delta x}, \quad \gamma = \omega \Delta t, \quad s = \sin\left(\frac{k\Delta x}{2}\right), \quad \mu = \cos^2\left(\frac{k\Delta x}{2}\right) = 1 - s^2.$$

By substituting the discrete Fourier modes $r_j^n = \varphi_r^n e^{ikx_j}$, $u_j^n = \varphi_u^n e^{ikx_j}$ and $v_j^n = \varphi_v^n e^{ikx_j}$ into the fully discrete scheme (6), we obtain $\mathscr{A}\varphi^{n+1} = \mathscr{B}\varphi^n$ where the matrices $\mathscr{A}$ and $\mathscr{B}$ are given by

$$\mathscr{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -(1-\theta_1)\gamma\mu \\ 0 & (1-\theta_2)\gamma\mu & 1 \end{pmatrix}$$

and

$$\mathscr{B} = \begin{pmatrix} 1 - 2\kappa_r|a_\star|\sigma s^2 & -a_\star\sigma i \sin(k\Delta x) & -\frac{\kappa_r \omega \Delta t}{2} i \sin(k\Delta x) \\ -a_\star\sigma i \sin(k\Delta x) & 1 - 2\kappa_u|a_\star|\sigma s^2 & \theta_1\gamma\mu \\ 0 & -\theta_2\gamma\mu & 1 \end{pmatrix}.$$

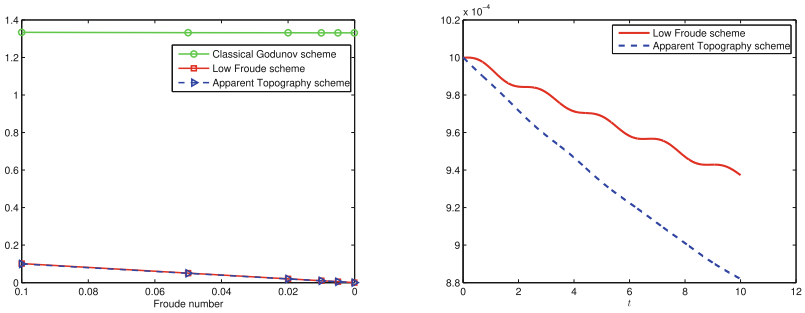
We then search for the eigenvalues of the amplification matrix $\mathscr{C} = \mathscr{A}^{-1}\mathscr{B}$, that are the roots of the third order polynomial $\mathscr{P}(\lambda) = \det(\mathscr{B} - \lambda\mathscr{A})$. Easy computations lead to

$$\mathscr{P}(\lambda) = (1-\lambda)(\Lambda\lambda^2 + \xi\lambda + \zeta) \tag{9}$$

with

$$\begin{aligned} \Lambda &= 1 + \gamma^2\mu^2(1-\theta_1)(1-\theta_2) > 0 \\ \xi &= -2 + \gamma^2\mu^2(\theta_1 + \theta_2 - 2\theta_1\theta_2) \\ &\quad + 2(\kappa_r + \kappa_u)|a_\star|\sigma s^2 + 2\kappa_r|a_\star|\sigma s^2\gamma^2\mu^2(1-\theta_1)(1-\theta_2) \\ \zeta &= 1 + \gamma^2\mu^2\theta_1\theta_2 - 2(\kappa_r + \kappa_u)|a_\star|\sigma s^2 \\ &\quad + 4a_\star^2\sigma^2 s^2(1-s^2) + 4\kappa_r\kappa_u a_\star^2\sigma^2 s^4 + 2\kappa_r|a_\star|\sigma s^2\gamma^2\mu^2\theta_2(1-\theta_1). \end{aligned}$$

The eigenvalue $\lambda_0 = 1$ corresponds to the discrete kernel (7). In order to ensure that the other two roots of (9) are in the unit circle ($|\lambda_\pm| \leq 1$), the coefficients Λ, ξ and ζ have to satisfy $|\zeta| \leq \Lambda$ and $|\xi| \leq \Lambda + \zeta$. Computations are then similar to the ones in [2] and lead to the results. More precisely, condition $\zeta \leq \Lambda$ leads to the condition involving Δt_a and condition $|\xi| \leq \Lambda + \zeta$ leads to conditions involving Δt_b and Δt_c.



(a) Maximum in time of the deviation $\|q_h - \hat{q}_h^0\|$ depending on the Froude number (b) Evolution of the deviation $\|q_h - \hat{q}_h^0\|$ for $M = 10^{-3}$

Fig. 2 Comparisons of classical and WB schemes

3 Numerical Results

Let us fix the parameters $a_\star = 1$, $\omega = 1$, $\theta_1 = 1, \theta_2 = 0$ and consider the initial condition on the domain $(0, 2\pi)$

$$q_i^0 = \hat{q}_i^0 + M \frac{\tilde{q}_i^0}{\|\tilde{q}_i^0\|} \quad \text{where} \begin{cases} \hat{q}_h^0(x) = \big(\sin(\omega x), 0, a_\star \cos(\omega x)\big) \in \mathscr{E}^h_{\omega \neq 0}, \\ \tilde{q}_h^0(x) = \big(a_\star \cos(\omega x), 1, \sin(\omega x)\big) \in \mathscr{E}^{h,\perp}_{\omega \neq 0}, \end{cases}$$

which is close to the kernel $\mathscr{E}^h_{\omega \neq 0}$ up to a perturbation of order M. We solve the 1D linear wave equation (2) by means of the *Apparent Topography* scheme (6), the *low Froude* scheme and the classical Godunov scheme. We observe on Fig. 2 (left) that the classical Godunov scheme is inaccurate since the deviation from the kernel does not remain of order M, while the two schemes designed for the geostrophic regime have the correct behaviour as the Froude number goes to 0. We now investigate the accuracy with time at a fixed Froude number. As exhibited for the semi-discrete scheme, we see on Fig. 2 (right) that the *Apparent Topography* scheme is more diffusive than the *low Froude* scheme for the part of the signal which is in the orthogonal of the kernel.

In future works, the authors will apply the two schemes to linear 2D cases before considering nonlinear applications in order to discriminate them.

References

1. Audusse, E., Bouchut, F., Bristeau, M.O., Klein, R., Perthame, B.: A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water flows. SIAM J. Sci. Comput. **25**(6), 2050–2065 (2004)
2. Audusse, E., Dellacherie, S., Do, M.H., Omnes, P., Penel, Y.: Godunov type scheme for the linear wave equation with Coriolis source term. Accepted in ESAIM:ProcS
3. Bouchut, F., Le Sommer, J., Zeitlin, V.: Frontal geostrophic adjustment and nonlinear wave phenomena in one-dimensional rotating shallow water. II. High-resolution numerical simulations. J. Fluid Mech. **514**, 35–63 (2004)
4. Castro, M.J., López, J.A., Parés, C.: Finite volume simulation of the geostrophic adjustment in a rotating shallow-water system. SIAM J. Sci. Comput. **31**(1), 444–477 (2008)
5. Dellacherie, S.: Analysis of Godunov type schemes applied to the compressible Euler system at low Mach number. J. Comput. Phys. **229**(4), 978–1016 (2010)

Application of a Combined Finite Element—Finite Volume Method to a 2D Non-hydrostatic Shallow Water Problem

Nora Aïssiouene, Marie-Odile Bristeau, Edwige Godlewski, Anne Mangeney, Carlos Parés and Jacques Sainte-Marie

Abstract We propose a numerical method for a two-dimensional non-hydrostatic shallow water system with topography (Bristeau et al. in Discret Contin Dyn Syst Ser B 20(4):961–988, 2015, [6]). We use a prediction-correction scheme initially introduced by Chorin-Temam (Rannacher in The Navier-Stokes equations II—theory and numerical methods. Springer, Berlin, pp 167–183, 1992, [13], and which has been applied previously to the one dimensional problem in Aïssiouene (Numerical analysis and discrete approximation of a dispersive shallow water model, 2016, [1]). The prediction part leads to solving a shallow water system for which we use finite volume methods (Audusse and Bristeau in J Comput Phys 206(1):311–333, 2005, [3]), while the correction part leads to solving a mixed problem in velocity/pressure using a finite element method. We present an application of the method with a comparison between a hydrostatic and a non-hydrostatic model.

N. Aïssiouene (✉)
Inria Paris -LJLL- UPMC, Paris, France
e-mail: nora.aissiouene@inria.fr

M.-O. Bristeau
Inria Paris, Paris, France
e-mail: Marie-Odile.Bristeau@inria.fr

E. Godlewski
LJLL-UPMC, Paris, France
e-mail: Edwige.Godlewski@upmc.fr

A. Mangeney
IPGP, Universit Paris Diderot, SPC, Paris, France
e-mail: anne.mangeney@ipgp.fr

C. Parés
Uma - Universidad de Málaga, Málaga, Spain
e-mail: pares@uma.es

J. Sainte-Marie
Inria Paris -LJLL-UPMC-CEREMA, Paris, France
e-mail: Jacques.Sainte-Marie@inria.fr

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_24

Keywords Finite element · Finite volume · Dispersive wave · Shallow water model · Non-hydrostatic pressure

MSC (2010): 65M06 · 65M60 · 76M10

1 Introduction

Mathematical models for free surface flows are widely studied, however one still needs to improve the existing models as well as develop robust numerical methods. The most common way to represent the physical behavior of the free surface is to compute the solutions of the Shallow Water equations. These equations are based on a shallowness assumption and lead to assuming the pressure is hydrostatic. Therefore, they are used for many geophysical flows on rivers, lakes, oceans where the characteristic horizontal length is much greater than the depth.

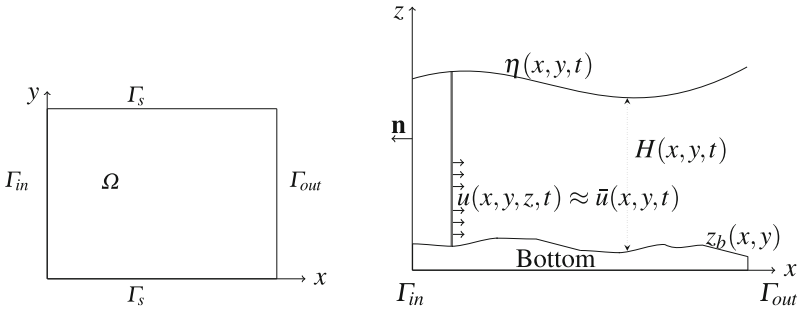
However, when the hydrostatic assumption is no longer valid, what we call dispersive effects appear, then more complex models have to be used to represent these effects. Many free surface models are available to take into consideration this dispersive effect, see [11] for the classical Green-Naghdi (GN) model and [5–7, 9] for other kinds of non hydrostatic models with bathymetry.

In this approach, we propose a new method dealing with a formulation without high order terms, we treat the depth-averaged Euler system developed in [6] where the non-hydrostatic pressure is an unknown of the system. The aim is to provide a robust numerical method for the two-dimensional model on an unstructured grid. The objective is to have a stable method to simulate real cases where the topography can be complex and needs an irregular mesh. Moreover, it gives the possibility to perform adaptive meshes if one wants to refine the mesh in the areas where the dispersive effects are expected. For instance, the dispersive contribution can have a significant impact in the water depth for the propagation of tsunamis [4, 10].

The paper is organized as follows. In the next section, we recall the depth-averaged Euler system. The Sect. 2 is devoted to the Chorin-Temam approach (prediction-correction scheme) applied for the model problem, while in Sect. 4, we give a geophysical application where we compare the results using a hydrostatic model vs a non-hydrostatic model.

2 The Averaged Euler System

We consider a two-dimensional domain $\Omega \subset \mathbb{R}^2$ delimited by the boundary $\Gamma = \Gamma_{in} \cup \Gamma_{out} \cup \Gamma_s$ as described in Fig. 1a. We denote by $H(x, y, t)$ the water depth, $z_b(x, y)$ the topography, $\mathbf{u}(x, y, t)$ the averaged velocity of the fluid $\mathbf{u} = (u, v, w)^t$ and p the non hydrostatic pressure (see Fig. 1b).



(a) View from above. (b) Vertical cross section.

Fig. 1 Model domain and notations

The two-dimensional depth-averaged Euler system described in [6] reads:

$$\frac{\partial H}{\partial t} + \nabla_0 \cdot (H\mathbf{u}) = 0, \tag{1}$$

$$\frac{\partial H\mathbf{u}}{\partial t} + \nabla_0 \cdot (H\mathbf{u} \otimes \mathbf{u}) + \nabla_0(\frac{g}{2}H^2) + \nabla_{sw}(p) = -gH\nabla_0(z_b), \tag{2}$$

$$\mathrm{div}_{sw}(\mathbf{u}) = 0, \tag{3}$$

where we define the operators ∇_0 and div_0 by

$$\nabla_0 f = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ 0 \end{pmatrix}, \quad \mathrm{div}_0 \mathbf{v} = \frac{\partial v_1}{\partial x} + \frac{\partial v_2}{\partial y}. \tag{4}$$

Also, we give an interpretation of the non-hydrostatic contribution by defining a shallow water version of the pressure gradient ∇_{sw} and the divergence operator div_{sw}. Assuming that f and $\mathbf{v} = (v_1, v_2, v_3)^T$ are smooth enough:

$$\nabla_{sw} f = \begin{pmatrix} H\frac{\partial f}{\partial x} + f\frac{\partial (H+2z_b)}{\partial x} \\ H\frac{\partial f}{\partial y} + f\frac{\partial (H+2z_b)}{\partial y} \\ -2f \end{pmatrix}, \tag{5}$$

$$\mathrm{div}_{sw}(\mathbf{v}) = \frac{\partial Hv_1}{\partial x} + \frac{\partial Hv_2}{\partial y} - v_1\frac{\partial (H+2z_b)}{\partial x} - v_2\frac{\partial (H+2z_b)}{\partial y} + 2v_3. \tag{6}$$

Under the assumptions done for the derivation of the non-hydrostatic model, the operator ∇_{sw} (resp. div_{sw}) is the average of the classical operator ∇ (resp. div) in the sens that it corresponds to the gradient averaged in the vertical direction between z_b and η. An important property is that the operators div_{sw} and ∇_{sw} satisfy the duality relation

$$\int_{\Omega} \nabla_{sw}(f) \cdot \mathbf{v} = -\int_{\Omega} \mathrm{div}_{sw}(\mathbf{v}) f + \int_{\Gamma} H f \mathbf{v} \cdot \mathbf{n}, \tag{7}$$

where $\mathbf{n}$ is the outward unit normal vector to the boundary Γ. This property is crucial for the algorithm presented in the following since we will consider a mixed problem in velocity/pressure, which will lead, at the numerical level, to having an operator for the pressure and its transpose for the velocity.

The depth-averaged model (1)–(3) is derived in [6] and is based on the minimization of the energy (see [12], this property provides a consistency with the Euler system [6] in terms of energy.

$$p_{tot} = g\frac{H}{2} + p, \tag{8}$$

where we take into account the hydrostatic pressure $g\frac{H}{2}$.

3 Prediction—Correction Scheme

The problem (1)–(3) is solved using a Chorin-Temam splitting scheme (see [13]).

The prediction-correction method is widely used to approximate the Navier-Stokes equations and is based on a time-splitting scheme. For each time step, the problem is solved in two steps, in the first one, we use a finite-volume method to solve the hyperbolic part which is a Shallow Water system with topography (where the non hydrostatic pressure p is not evaluated). This allows us to get a first predicted state which is not divergence free. In the second step, we update the predicted state with the shallow water version of the gradient pressure evaluated in such a way that the velocity satisfies the divergence free condition (3).

Let us denote by X the vectors of unknowns and $F(X)$ the matrix:

$$X = \begin{pmatrix} H \\ Hu \\ Hv \\ Hw \end{pmatrix}, \quad F(X) = \begin{pmatrix} Hu & Hv \\ Hu^2 + \frac{g}{2}H^2 & Huv \\ Huv & Hv^2 + \frac{g}{2}H^2 \\ Huw & Hvw \end{pmatrix}, \tag{9}$$

and set

$$S(X) = \begin{pmatrix} 0 \\ -gH\frac{\partial z_b}{\partial x} \\ -gH\frac{\partial z_b}{\partial y} \\ 0 \end{pmatrix} \quad \text{and} \quad R_{nh} = \begin{pmatrix} 0 \\ \nabla_{sw}(p) \end{pmatrix}. \tag{10}$$

Then, the system (1)–(3) can be written

$$\frac{\partial X}{\partial t} + \mathrm{div}_0 F(X) + R_{nh} = S(X), \tag{11}$$

$$\mathrm{div}_{sw}(\mathbf{u}) = 0. \tag{12}$$

We set t^0 the initial time and $t^{n+1} = t^n + \Delta t^n$ where Δt^n satisfies a stability condition (CFL) and the state X^n will denote an approximation of $X(t^n)$. For each time step, we consider an intermediate state which will be denoted with the superscript $^{n+1/2}$. The semi discretization in time can be summarized in the following steps:

$$X^{n+1/2} = X^n - \Delta t^n \mathrm{div}_0 F(X^n) + \Delta t S(X^n), \tag{13}$$

$$X^{n+1} + \Delta t^n R_{nh}^{n+1} = X^{n+1/2}, \tag{14}$$

$$\mathrm{div}_{sw}\, \mathbf{u}^{n+1} = 0. \tag{15}$$

So the first step (13) leads to solving the hyperbolic system with source terms in order to get the state $X^{n+1/2} = (H^{n+1/2}, (Hu)^{n+1/2}, (Hv)^{n+1/2}, (Hw)^{n+1/2})^T$. Equation (14) allows us to correct the predicted value $X^{n+1/2}$ in order to obtain a state which satisfies the divergence free condition (15).

The prediction part (13) is solved using a cell centered finite-volume method [3]. For this system, our scheme is second order accurate in time and, if we use a reconstruction algorithm [3] in the hyperbolic step, it is formally second order accurate in space [2, 3]. In the application, we use a kinetic solver for its good mathematical properties. The correction part (14) is solved using a finite element method. To do so, we consider the equations (14)–(15) as a mixed problem [2] and, starting with an appropriate variational formulation of the problem, we apply the finite element method to obtain the pressure p^{n+1} which is solution of an elliptic equation and the velocity $\mathbf{u}^{n+1}$. The elliptic equation of the pressure can be written under the form:

$$\mathrm{div}_{sw}\left(\frac{\nabla_{sw}\, p^{n+1}}{H^{n+1}}\right) = \frac{1}{\Delta t^n}\mathrm{div}_{sw}\left(\frac{(H\mathbf{u})^{n+1/2}}{H^{n+1/2}}\right). \tag{16}$$

We consider a primal mesh which is the triangular mesh and a dual mesh corresponding to the centered finite volume cells. The approximation of the variables is based on the triangular mesh for the finite element scheme and the dual mesh for the finite volume scheme. The finite volume cells are centered on the vertices and built by joining the centers of mass of the triangles surrounding each vertex. The variables H, $H\mathbf{u}$ are estimated first as constant mean values on the cells by the finite volume scheme, which gives the intermediate state $X^{n+1/2}$. For the finite element scheme, the state X^{n+1} is approximated at the vertices of the triangles. The algorithm uses an iterative method of Uzawa type to solve the elliptic equation in pressure involved in the problem. The details of the combined method and the treatment of the boundary conditions will be detailed in a forthcoming paper.

4 Numerical Results

In this section we test the depth-averaged model (1)–(3) on a numerical application. We generate small amplitude waves at the inlet of a domain of dimensions $[0, 10] \times [0, 6]$ and we observe the propagation of the waves over an obstacle. The channel is also ended by a slope of 40%. This simulation allows us to confront our method to a test case where we have a variable bottom with strong variations of the elevation and wet/dry interfaces. The dimensions of the case are described in Fig. 2 and the obstacle is defined by the topography function:

$$z_b(x, y) = \min\left(z_m, Ae^{-\left((a(x-x_0)^2)+b(y-y_0)^2\right)}\right), \tag{17}$$

where we set $z_m = 0.5\,\text{m}$, $A = 2\,\text{m}$, $a = 3.3\,\text{m}$, $b = 1.51\,\text{m}$ and $x_0 = 3\,\text{m}$, $y_0 = 3\,\text{m}$. We set an initial free surface $\eta_0 = 0.6\,\text{m}$ and a sinusoidal wave given at the inlet with an amplitude of 0.02 m. The test is performed over an unstructured mesh of 45506 nodes for the fine mesh. The numerical solution is computed with a P1-iso-P2/P1 approximation (see [1] for more details on the choices of approximation spaces). We compare the solutions obtained using the Shallow Water model and using the depth-averaged Euler model (1)–(3) in order to observe the effects of the dispersion on the propagation and the wave interactions. Figure 3 shows the simulations at instant $t_1 = 4.54531\,s$ (Fig. 3a, b) for the Shallow Water model (left) and the dispersive model (right). The figures represent the free surface η. We clearly observe the impact of the dispersive effects around the obstacle and on the forms of the waves. In Fig. 4 we show the free surface over the time at different points around the obstacle and compare the solution obtained for the Shallow Water model and the depth averaged model. We can recover the same kind behavior in one dimensions

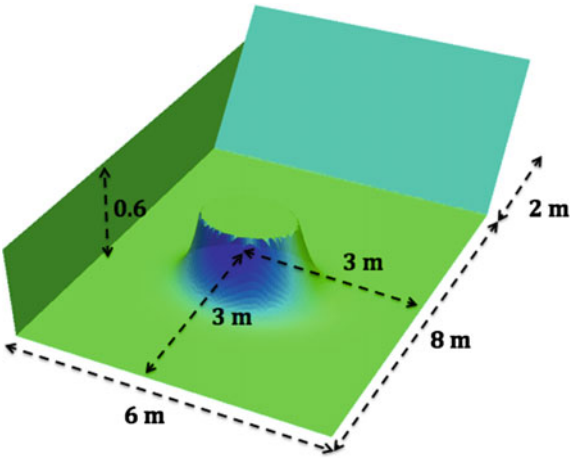
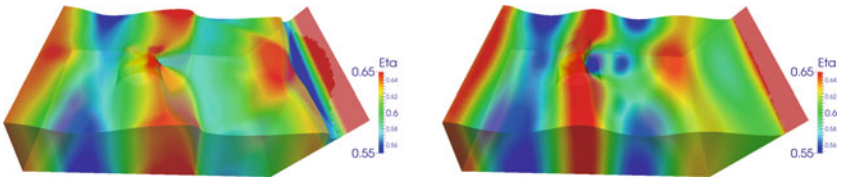
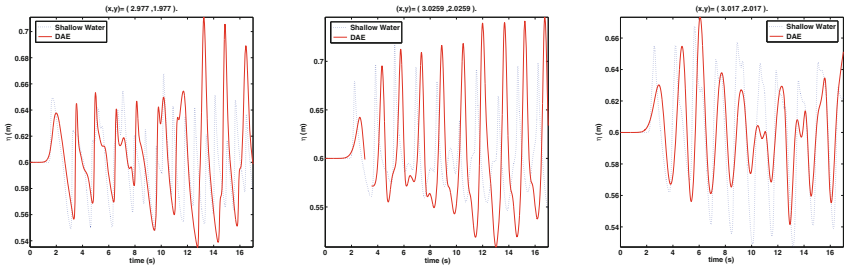


Fig. 2 Dimension of the test case



(a) Hydrostatic simulation at time $t_1 = 4.54531\,s$ (b) Non-hydrostatic simulation at time t_1

Fig. 3 Free surface obtained with a hydrostatic simulation and a non-hydrostatic simulation



(a) Point 1 (b) Point 2 (c) Point 3

Fig. 4 Comparison of the free surface over the time for the selected points between solutions computed with a hydrostatic model (· · ·) and the depth-averaged model (—)

for a very classical test case which is known as the Dingemans experiment [8], these effects occur when we have a strong variation of the topography with a strong gradient of the elevation.

5 Conclusion

In this paper, we have presented an application a the combined finite-volume/finite element method for a two dimensional dispersive shallow water model on an unstructured mesh. We solve a mixed problem using a finite element method to obtain the velocity and the non-hydrostatic pressure.

Acknowledgements The authors acknowledge the Inria Project Lab Algae in Silicio for its financial support. The first author received a partial grant from the Fondation Ledoux. This research is also supported by the CEREMA, the ERC SLIDEQUAKES and the ANR MIMOSA project.

References

1. Aïssiouene, N.: Numerical analysis and discrete approximation of a dispersive shallow water model. Theses, Pierre et Marie Curie, Paris VI (2016). https://hal.archives-ouvertes.fr/tel-01418676
2. Aïssiouene, N., Bristeau, M.O., Godlewski, E., Sainte-Marie, J.: A combined finite volume-finite element scheme for a dispersive shallow water system. Netw. Heterog. Media **11**(1), 1–27 (2016). doi:10.3934/nhm.2016.11.1. https://hal.inria.fr/hal-01160718
3. Audusse, E., Bristeau, M.O.: A well-balanced positivity preserving second-order scheme for Shallow Water flows on unstructured meshes. J. Comput. Phys. **206**(1), 311–333 (2005)
4. Behrens, J., Dias, F.: New computational methods in tsunami science. Philos. Trans. A Math. Phys. Eng. Sci. **373**(2053) (2015). (Oct 28)
5. Bonneton, P., Barthelemy, E., Chazel, F., Cienfuegos, R., Lannes, D., Marche, F., Tissier, M.: Recent advances in Serre-Green Naghdi modelling for wave transformation, breaking and runup processes. Eur. J. Mech. B Fluids **30**(6), 589–597 (2011). doi:10.1016/j.euromechflu.2011.02.005. http://www.sciencedirect.com/science/article/pii/S0997754611000185. (Special Issue: Nearshore Hydrodynamics)
6. Bristeau, M.O., Mangeney, A., Sainte-Marie, J., Seguin, N.: An energy-consistent depth-averaged Euler system: derivation and properties. Discret. Contin. Dyn. Syst. Ser. B **20**(4), 961–988 (2015). doi:10.3934/dcdsb.2015.20.961. http://aimsciences.org/journals/displayArticlesnew.jsp?paperID=10801
7. Chazel, F., Lannes, D., Marche, F.: Numerical simulation of strongly nonlinear and dispersive waves using a Green-Naghdi model. J. Sci. Comput. **48**(1–3), 105–116 (2011). doi:10.1007/s10915-010-9395-9. http://dx.doi.org/10.1007/s10915-010-9395-9
8. Dingemans, M.W.: Wave propagation over uneven bottoms. Advanced Series on Ocean Engineering-World Scientific (1997)
9. Duran, A., Marche, F.: Discontinuous-Galerkin discretization of a new class of Green-Naghdi equations. Commun. Comput. Phys. 130 (2014). https://hal.archives-ouvertes.fr/hal-00980826
10. Glimsdal, S., Pedersen, G.K., Harbitz, C.B., Løvholt, F.: Dispersion of tsunamis: does it really matter? Nat. Hazards Earth Syst. Sci. **13**(6), 1507–1526 (2013). doi:10.5194/nhess-13-1507-2013. http://www.nat-hazards-earth-syst-sci.net/13/1507/2013/
11. Green, A., Naghdi, P.: A derivation of equations for wave propagation in water of variable depth. J. Fluid Mech. **78**, 237–246 (1976)
12. Levermore, C., Sammartino, M.: A shallow water model with eddy viscosity for basins with varying bottom topography. Nonlinearity **14**(6), 1493–1515 (2001)
13. Rannacher, R.: On Chorin's projection method for the incompressible Navier-Stokes equations. In: Heywood John, G., Masuda, K., Rautmann, R., Solonnikov Vsevolod, A.(eds.) The Navier-Stokes Equations II—Theory and Numerical Methods, *Lecture Notes in Mathematics*, vol. 1530, pp. 167–183. Springer, Berlin, Heidelberg (1992). doi:10.1007/BFb0090341. http://dx.doi.org/10.1007/BFb0090341

A Relaxation Scheme for the Simulation of Low Mach Number Flows

Emanuela Abbate, Angelo Iollo and Gabriella Puppo

Abstract A scheme for the simulation of inviscid flows with low Mach number is derived. The scheme is built on a relaxation system and it is based on a linear implicit time discretization. The advective part is discretized by a convex combination of upwind and centered schemes, in order to recover the correct limit when the Mach number goes to zero. The implicit treatment allows to stabilize the central approximation in the low Mach limit and also to avoid demanding constraints on the time step in low Mach flows. The scheme applies to steady or unsteady flows and to general equations of state. We discuss examples pertaining to both gas and liquid flows.

Keywords Low mach flows · Relaxation method · Compressible flows

MSC (2010): 65M08 · 65N08 · 35Q31

1 Introduction

In the numerical simulation of fluid flows, the low Mach regime induces severe stiffness and, consequently, stability problems for standard computational techniques. This is due to the fact that the upwind schemes adopted when solving compressible

E. Abbate (✉) · G. Puppo
Dip. di Scienza e Alta Tecnologia, Univ. dell'Insubria, via Valleggio, Como, Italy
e-mail: eabbate.studenti@uninsubria.it; emanuela.abbate@u-bordeaux.fr;
emanuela.abbate@inria.fr

G. Puppo
e-mail: gabriella.puppo@uninsubria.it

E. Abbate · A. Iollo
Memphis Team, Inria Bordeaux Sud-Ouest, Talence, France
e-mail: angelo.iollo@u-bordeaux.fr

E. Abbate · A. Iollo
University Bordeaux, IMB, UMR 5251, 33400 Talence, France

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_25

flows are not "asymptotic preserving", namely they do not preserve the low Mach number limit behaviour, which tends to incompressible flow [6, 7]. It has been proved in [4] that upwind schemes lead to pressure fluctuations of the order of the Mach number M, while in the continuous case the pressure fluctuations are of order M^2: this is one of the main reasons why compressible flow solvers perform so poorly in the low Mach number regime.

In this work, the derivation of a scheme for solving low Mach inviscid flows is addressed with a novel implicit relaxation approach. The relaxation method introduced in [5] is adopted, approximating the original system with a larger one. This new system is linear except for a lower order source term and thanks to linearity, implicit time discretizations can be easily implemented. Therefore, both upwind and centered spatial discretizations can be used without having stability problems. The centered scheme allows to recover the correct limit on the pressure gradients in the low Mach regime. Nevertheless, the upwind discretization is needed in presence of Mach numbers of order one, in order to introduce enough numerical viscosity. Since the adopted spatial discretization is a convex combination of these two, it is able to approximate flows in different regimes and thus the scheme can be seen as an "all-speed" scheme, as the ones proposed in [2, 3]. Moreover, thanks to the absolute stability of implicit schemes, the CFL is not limited by the acoustic constraint, which becomes extremely demanding in low Mach regimes.

2 The Relaxation Method

We briefly revise the relaxation method developed in [5]. A general hyperbolic system of balance laws has the following structure:

$$\partial_t \boldsymbol{\psi} + \partial_x \mathbf{F}(\boldsymbol{\psi}) = \mathbf{Q}(\boldsymbol{\psi}), \tag{1}$$

where $\boldsymbol{\psi} \in \mathbb{R}^n$ is the vector of the conservative variables, $\mathbf{F}(\boldsymbol{\psi})$ the vector of the fluxes and $\mathbf{Q}(\boldsymbol{\psi})$ a source term. By introducing the relaxation variables vector $\mathbf{v} \in \mathbb{R}^n$, the relaxation system approximating the original system (1) reads:

$$\begin{cases} \partial_t \boldsymbol{\psi} + \partial_x \mathbf{v} = \mathbf{Q}(\boldsymbol{\psi}) \\ \partial_t \mathbf{v} + \mathbf{A}\partial_x \boldsymbol{\psi} = \dfrac{1}{\varepsilon}(\mathbf{F}(\boldsymbol{\psi}) - \mathbf{v}), \quad \varepsilon > 0, \end{cases} \tag{2}$$

where $\mathbf{A} = diag\{a_i\},\ i = 1, .., n$ is a positive diagonal matrix. The small positive parameter ε is called relaxation rate. The right hand side of the second equation is a stiff lower order term. Our motivation in choosing this specific relaxation method relies in the resulting linearity of the advective part. This allows for an easy implementation of implicit schemes: the introduction of modified Riemann solvers is not required, as it is when adopting the Suliciu relaxation approach [1].

In order to ensure the dissipative nature of the relaxation, it is necessary to respect the Liu *subcharacteristic condition* (for details see [5])

$$\mathbf{A} - \mathbf{F}'(\boldsymbol{\psi})^2 \geq 0 \quad \forall \boldsymbol{\psi} \tag{3}$$

when building the relaxation system. By enforcing this condition, the coefficients of the relaxation matrix $\mathbf{A}$ are derived in such a way that the characteristic speeds $\lambda_i,\ i = 1, ..n$ of the original system (1) interlace with those of the relaxation system $\mu_j = \pm\sqrt{a_i},\ j = 1, ..2n,\ i = 1, ..n$. In our computations, $\mathbf{A}$ is built by taking $a_i = max_x \lambda_i^2$. This is the maximum over the domain of the a-priori estimated speed.

3 Numerical Relaxation Scheme

System (2) is discretized with finite volumes on a uniform mesh. For 1D problems, the grid spacing is $\Delta x = x_{i+1/2} - x_{i-1/2}$ and the time step $\Delta t = t_{n+1} - t_n$. Let $\mathbf{w}_i^n$ be the approximate cell average of a quantity $\mathbf{w}$ in the cell $\left[x_{i-1/2}, x_{i+1/2}\right]$ at time t_n and $\mathbf{w}_{i+1/2}^n$ the approximate point value of $\mathbf{w}$ at an interface $x = x_{i+1/2}$ and at $t = t_n$.

In general, the spatial discretization for system (2) can be written as

$$\begin{cases} \partial_t \boldsymbol{\psi}_i + \dfrac{\mathbf{v}_{i+1/2} - \mathbf{v}_{i-1/2}}{\Delta x} = \mathbf{Q}\left(\boldsymbol{\psi}_i\right) \\ \partial_t \mathbf{v}_i + \mathbf{A}\dfrac{\boldsymbol{\psi}_{i+1/2} - \boldsymbol{\psi}_{i-1/2}}{\Delta x} = \dfrac{1}{\varepsilon}\left(\mathbf{F}\left(\boldsymbol{\psi}_i\right) - \mathbf{v}_i\right). \end{cases} \tag{4}$$

This is valid with an accuracy of $\mathcal{O}\left(\Delta x^2\right)$ and for sufficiently accurate reconstructions of the quantities at the interfaces, noting that

$$\mathbf{F}\left(\boldsymbol{\psi}_i\right) = \mathbf{F}\left(\frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} \boldsymbol{\psi}\, dx\right) = \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} \mathbf{F}\left(\boldsymbol{\psi}\right) dx + \mathcal{O}\left(\Delta x^2\right) = \mathbf{F}_i + \mathcal{O}\left(\Delta x^2\right).$$

3.1 Spatial Discretization

In (4), the interface values have to be computed with a specific scheme. Let $\mathbf{u}$ be a generic variable. We define an hybrid discretization in the following way:

$$\left(\frac{\mathbf{u}_{i+1/2} - \mathbf{u}_{i-1/2}}{\Delta x}\right)_{hyb} = f\left(M_{loc}\right)\left(\frac{\mathbf{u}_{i+1/2} - \mathbf{u}_{i-1/2}}{\Delta x}\right)_{upw} + \left(1 - f\left(M_{loc}\right)\right)\left(\frac{\mathbf{u}_{i+1/2} - \mathbf{u}_{i-1/2}}{\Delta x}\right)_{cent}, \tag{5}$$

where the first term defines the spatial discretization through an upwind method and the second through a centered one. M_{loc} is the local Mach number and we choose $f\left(M_{loc}\right) = \min\{1, M_{loc}\}$ in order to have a convex combination. This way the correct numerical viscosity for each regime is recovered.

The upwind scheme is derived as in [5], computing the interface values by using the characteristic variables $\mathbf{v} \pm \mathbf{A}^{1/2}\boldsymbol{\psi}$ of system (2). At first order we have:

$$\begin{cases} \left(\dfrac{\mathbf{v}_{i+1/2} - \mathbf{v}_{i-1/2}}{\Delta x}\right)_{upw} = \dfrac{1}{2\Delta x}\left(\mathbf{v}_{i+1} - \mathbf{v}_{i-1}\right) - \dfrac{\mathbf{A}^{1/2}}{2\Delta x}\left(\boldsymbol{\psi}_{i+1} - 2\boldsymbol{\psi}_i + \boldsymbol{\psi}_{i-1}\right) \\ \left(\dfrac{\boldsymbol{\psi}_{i+1/2} - \boldsymbol{\psi}_{i-1/2}}{\Delta x}\right)_{upw} = \dfrac{1}{2\Delta x}\left(\boldsymbol{\psi}_{i+1} - \boldsymbol{\psi}_{i-1}\right) - \dfrac{\mathbf{A}^{-1/2}}{2\Delta x}\left(\mathbf{v}_{i+1} - 2\mathbf{v}_j + \mathbf{v}_{i-1}\right). \end{cases} \tag{6}$$

For second order approximations, a Van Leer MUSCL scheme is employed. Since in this work we test second order schemes only for the computation of smooth solutions, non-linear flux limiters are not introduced. The centered spatial discretization is second order accurate and reads:

$$\begin{cases} \left(\dfrac{\mathbf{v}_{i+1/2} - \mathbf{v}_{i-1/2}}{\Delta x}\right)_{cent} = \dfrac{1}{2\Delta x}\left(\mathbf{v}_{i+1} - \mathbf{v}_{i-1}\right) \\ \left(\dfrac{\boldsymbol{\psi}_{i+1/2} - \boldsymbol{\psi}_{i-1/2}}{\Delta x}\right)_{cent} = \dfrac{1}{2\Delta x}\left(\boldsymbol{\psi}_{i+1} - \boldsymbol{\psi}_{i-1}\right). \end{cases} \tag{7}$$

3.2 Implicit Time Discretization

To ensure that the centered discretization (7) is stable and to avoid the acoustic CFL, we propose a fully implicit relaxation scheme, instead of adopting classic explicit relaxation schemes [5, 8]. The implicit scheme at first order reads:

$$\begin{cases} \dfrac{\boldsymbol{\psi}_i^{n+1} - \boldsymbol{\psi}_i^n}{\Delta t} + \dfrac{\mathbf{v}_{i+1/2}^{n+1} - \mathbf{v}_{i-1/2}^{n+1}}{\Delta x} = \mathbf{Q}\left(\boldsymbol{\psi}_i^{n+1}\right) \\ \dfrac{\mathbf{v}_i^{n+1} - \mathbf{v}_i^n}{\Delta t} + \mathbf{A}\dfrac{\boldsymbol{\psi}_{i+1/2}^{n+1} - \boldsymbol{\psi}_{i-1/2}^{n+1}}{\Delta x} = \dfrac{1}{\varepsilon}\left(\mathbf{F}\left(\boldsymbol{\psi}_i^{n+1}\right) - \mathbf{v}_i^{n+1}\right), \end{cases} \tag{8}$$

which is a simple backward Euler. For a second order implicit approximation, a BDF (Backward Differentiation Formula) second order accurate is employed.

The implicit non linear part in the source is dealt with applying one iteration of the Newton method, namely the fluxes are approximated with the Taylor expansion

$$\mathbf{F}\left(\boldsymbol{\psi}^{n+1}\right) = \mathbf{F}\left(\boldsymbol{\psi}^n\right) + \mathbf{F}'\left(\boldsymbol{\psi}^n\right)\left(\boldsymbol{\psi}^{n+1} - \boldsymbol{\psi}^n\right), \tag{9}$$

where $\mathbf{F}'\left(\boldsymbol{\psi}^n\right)$ is the Jacobian of the flux and can be computed analytically. This constitutes a significant simplification with respect to the use of complex solvers that

deal with non-linear fluxes inside implicit schemes. When a non-linear source term $\mathbf{Q}(\boldsymbol{\psi})$ is present, it can be treated in the same way. With linearization (9), the derived implicit scheme possesses the following zero relaxation limit $\varepsilon \to 0^+$

$$\frac{\boldsymbol{\psi}_i^{n+1} - \boldsymbol{\psi}_i^n}{\Delta t} + \frac{\mathbf{F}\left(\boldsymbol{\psi}_{i+1/2}^{n+1}\right) - \mathbf{F}\left(\boldsymbol{\psi}_{i-1/2}^{n+1}\right)}{\Delta x} = \mathbf{Q}\left(\boldsymbol{\psi}_i^{n+1}\right),$$

which is a consistent and stable discretization of (1).

In this framework, the full linear system to be solved reads

$$\begin{cases} \mathbf{M}\boldsymbol{\Psi}^{n+1} + \mathbf{N}\mathbf{V}^{n+1} = \mathbf{r} \\ \mathbf{P}\boldsymbol{\Psi}^{n+1} + \mathbf{R}\mathbf{V}^{n+1} = \mathbf{s}, \end{cases}$$

where $\boldsymbol{\Psi}^{n+1}$ and $\mathbf{V}^{n+1}$ are the vectors containing the grid point values of the conservative and of the relaxation variables respectively. In 1D problems, **M**, **N**, **P** and **R** have tridiagonal sub-blocks for first order discretizations. Linearization (9) only adds terms on the diagonals of the sub-blocks of **P**, not increasing the computational effort in the inversion algorithms.

The scheme is unconditionally stable. The acoustic Courant number is defined as $\nu_{ac} = \mu_{max}\Delta t/\Delta x$ and with the proposed implicit scheme it can be taken significantly larger than one. This is very useful in terms of computational time when simulating low Mach flows, since the acoustic waves are extremely fast (for Euler equations $\mu_{max} \geq |u + c|$ from condition (3), u being the flow velocity and c the sound speed). Therefore, the time step Δt is not excessively limited as it would be with an explicit scheme requiring $\nu_{ac} < 1$ for stability. Moreover, in approximating material waves, a "material CFL condition" can be employed: the material Courant number is defined as $\nu_{mat} = \mu_{mat}\Delta t/\Delta x$ with $\mu_{mat} \geq |u|$. To ensure accuracy on the material waves, the time step is chosen by imposing $\nu_{mat} \leq 1$. Thus, Δt is not constrained by the speed of the acoustic waves but only by the speed of the material wave.

4 Numerical Results

The proposed scheme is applied to the simulation of gas and liquid flows, governed by the one dimensional Euler equations with a generalized equation of state. In all simulations we employ $\varepsilon = 10^{-8}$. Initial and boundary conditions on the relaxation variables are set to be consistent with the equilibrium state $\mathbf{v} = \mathbf{F}(\boldsymbol{\psi})$ [5].

4.1 Steady State Example: Nozzle Flow

We validate the scheme by computing the steady state of a Laval nozzle flow in the quasi-1D approximation. In the general formulation (1), the conservative variables, the fluxes and the source term are:

$$\boldsymbol{\psi} = \begin{bmatrix} S\rho \\ S\rho u \\ SE \end{bmatrix}, \quad \mathbf{F}(\boldsymbol{\psi}) = \begin{bmatrix} S\rho u \\ S\left(\rho u^2 - p\right) \\ S(E-p)u \end{bmatrix}, \quad \mathbf{Q}(\boldsymbol{\psi}) = \begin{bmatrix} 0 \\ p\partial_x S \\ 0 \end{bmatrix},$$

where ρ is the density, u is the flow velocity, $E = \rho u^2/2 + \rho e$ is the total energy (sum of the kinetic part and the internal energy e) and p is the pressure. $S = S(x)$ is the cross sectional area of the nozzle, which is a smooth function of the axial coordinate x. In this approximation, all flow variables depend on only the coordinate x.

The scheme is able to deal with different state laws. In particular, we simulate:

- *perfect gas* flows, with the state law $e = p/(\rho(\gamma - 1))$ ($\gamma = c_P/c_V$ is the heat capacity ratio and it is equal to 1.4 for biatomic gases);
- *stiffened gas* flows, with the state law $e = p/(\rho(\gamma - 1)) + p_\infty/\rho$, where the p_∞ term models the intermolecular forces. This law describes flows of liquids (the parameters for water are $\gamma = 4.4$ and $p_\infty = 6.8 \cdot 10^8$Pa).

Three different configurations of the nozzle are simulated:

1. *test 1* is a perfect gas flow with Mach number in the range $M \in [0.45; 0.7]$. At the inlet the total pressure and temperature $P_{tot} = 1$Pa and $T_{tot} = 1$K are imposed, at the outlet $p_{out} = 0.9$Pa;
2. *test 2* is a perfect gas flow with Mach number in the range $M \in [4; 9] \cdot 10^{-3}$. At the inlet $P_{tot} = 1$Pa and $T_{tot} = 1$K are imposed, at the outlet $p_{out} = 0.99999$Pa;
3. *test 3* is a water flow with Mach number in the range $M \in [7.26; 8.67] \cdot 10^{-5}$. At the inlet $P_{tot} = 10$Pa and $T_{tot} = 280$K are imposed, at the outlet $p_{out} = 1$Pa.

Tests 2 and 3 are low Mach flows, with the last one being almost incompressible. We remark that pressure and temperature values are here chosen with the only purpose of demonstrating the properties of the scheme.

The pressure profiles obtained with the first and second order implicit scheme are shown in Fig. 1a, b, c for the three tests and they are compared to the results of the classic explicit-upwind relaxation scheme of [5]. The comparison shows that for test 1 the two schemes are equivalent. Instead, for low Mach flows (tests 2 and 3) our implicit scheme is more accurate than the explicit one, which gives a solution shifted from the exact one and presents oscillations, due to the wrong numerical viscosity. The implicit scheme computes pressure profiles that are superimposed to the exact solution in all the three cases thanks to discretization (5). In Fig. 1d, the convergence analysis for the implicit scheme is carried out and the correct convergence rates are recovered.

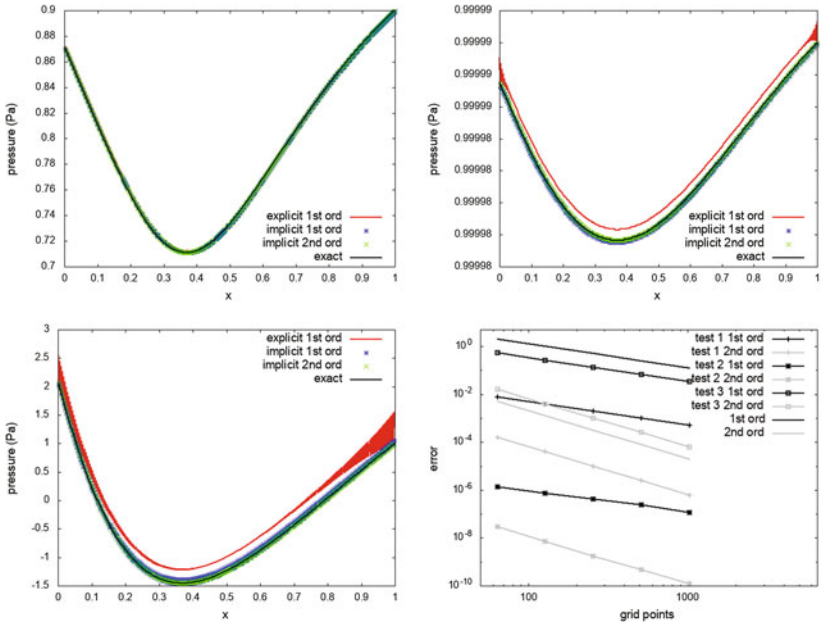


Fig. 1 Nozzle flow: pressure profiles obtained with first and second order implicit schemes and with first order explicit-upwind relaxation scheme of [5] (500 grid points). Panel **d** convergence analysis of the implicit scheme for the three tests

The time step for the implicit scheme is chosen by enforcing $\nu_{ac} = 100$ at first order and $\nu_{ac} = 40$ at second order. Thanks to this constraint, our implicit scheme is computationally less demanding. For example in test 2, a precision of order $2 \cdot 10^{-7}$ is obtained by the first order implicit scheme in 64 s of CPU time and by the explicit in 3464 s.

4.2 Material Wave Simulation: Gas Tube

We give a numerical illustration of the performances of the scheme for the simulation of material waves, by solving a low Mach perfect gas flow in a tube. The Euler system in the general formulation (1) reads

$$\boldsymbol{\psi} = \begin{bmatrix} \rho \\ \rho u \\ E \end{bmatrix}, \quad \mathbf{F}(\boldsymbol{\psi}) = \begin{bmatrix} \rho u \\ \rho u^2 - p \\ (E - p)\, u \end{bmatrix}, \quad \mathbf{Q}(\boldsymbol{\psi}) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Let x_d be the discontinuity position, which is in the middle of the tube when $t = 0$. The Riemann problem initial data are $\rho_L = \rho_R = 1\text{Kg/m}^3$, $u_L = 0\text{m/s}$, $u_R = 0.008\text{m/s}$,

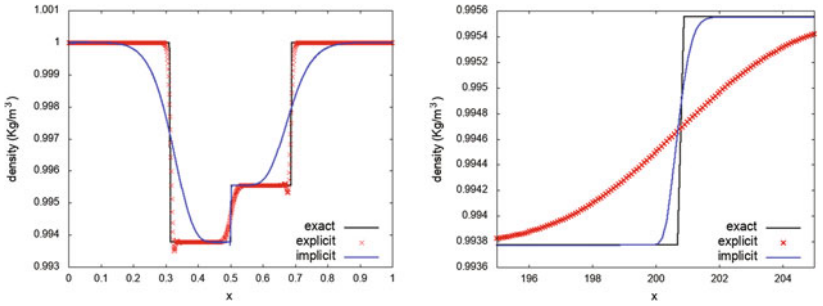


Fig. 2 Low Mach gas tube flow: density profiles obtained with the implicit scheme and with the explicit-upwind relaxation scheme of [5], both at first order

$p_L = 0.4$Pa and $p_R = 0.399$Pa. The solution consists in two fast expansion waves and a very slow material wave. On this latter wave we have $M \simeq 6 \cdot 10^{-3}$. The results at time $t = 0.25$s are shown in Fig. 2a: the wave has moved from $x_d = 0.5$ to $x_d = 0.501167$m, namely, for a grid spacing $\Delta x = 10^{-3}$, it has crossed 1 cell. The length of this tube is 1m. In order to see the material wave move, we run a simulation for long times: at time $t = 150$s (Fig. 2b) the wave has moved from $x_d = 200$ to $x_d = 200.7$m, namely, for a grid spacing $\Delta x = 10^{-1}$, it has crossed 7 cells. The length of this second tube is 400m.

The results obtained using the implicit scheme with $\nu_{mat} = 0.2$ (giving $\Delta t \simeq 2.2 \cdot 10^{-2}$ on the chosen grid) are compared to the ones of the explicit-upwind relaxation scheme of [5], where instead an acoustic condition has to be enforced ($\nu_{ac} = 0.4$ giving $\Delta t \simeq 5 \cdot 10^{-4}$). Due to the material constraint on Δt, the implicit scheme is of course very diffusive on the acoustic waves. If accuracy on the rarefactions is needed, an acoustic CFL has to be enforced, thus recovering the same precision of the explicit scheme on these waves. Nevertheless, the implicit scheme is able to accurately reproduce and keep the contact wave sharp as it moves to the right, thanks to spatial discretization (5). The explicit scheme, instead, is very diffusive on this wave, which is completely smeared after some time (see Fig. 2b).

5 Conclusions

We presented a novel implicit relaxation scheme, able to simulate low Mach flows, with first and second order accuracy. The low Mach limit is well reproduced in steady state problems and in the approximation of material waves. The scheme is simple and can be easily adapted to different state laws. In the future, simulation of waves in compressible elastic materials will be exploited.

References

1. Chalons, C., Coquel, F., Marmignon, C.: Well-balanced time implicit formulation of relaxation schemes for the Euler equations. SIAM J. Sci. Comput. **30**(1), 394–415 (2008)
2. Cordier, F., Degond, P., Kumbaro, A.: An asymptotic-preserving all-speed scheme for the Euler and Navier-Stokes equations. J. Comput. Phys. **231**(17), 5685–5704 (2012)
3. Degond, P., Tang, M.: All speed scheme for the low Mach number limit of the isentropic Euler equation (2009). arXiv:0908.1929
4. Guillard, H., Viozat, C.: On the behaviour of upwind schemes in the low Mach number limit. Comput. Fluids **28**(1), 63–86 (1999)
5. Jin, S., Xin, Z.: The relaxation schemes for systems of conservation laws in arbitrary space dimensions. Commun. Pure Appl. Math. **48**(3), 235–276 (1995)
6. Klainerman, S., Majda, A.: Compressible and incompressible fluids. Commun. Pure Appl. Math. **35**(5), 629–651 (1982)
7. Klein, R.: Semi-implicit extension of a Godunov-type scheme based on low Mach number asymptotics i: one-dimensional flow. J. Comput. Phys. **121**(2), 213–237 (1995)
8. Pareschi, L., Russo, G.: Implicit-explicit Runge-Kutta schemes and applications to hyperbolic systems with relaxation. J. Sci. Comput. **25**(1–2), 129–155 (2005)

Comparison of Wetting and Drying Between a RKDG2 Method and Classical FV Based Second-Order Hydrostatic Reconstruction

Stefan Vater, Nicole Beisiegel and Jörn Behrens

Abstract We compare the treatment of wetting and drying for shallow water flows at the coast using a discontinuous Galerkin (DG) scheme with classical finite volumes in one space dimension. The presented DG scheme employs piecewise linear ansatz functions and is formally second-order accurate. The core of the method is a velocity based "limiting" of the momentum, which provides stable and accurate solutions in the computation of inundation events. Artificial gradients of the water surface elevation which are introduced by the DG discretization at the wet/dry interface are specially handled to prevent spurious velocities. The finite volume method is based on second-order hydrostatic reconstruction. In general, both methods show comparable results in terms of stability and accuracy. For certain situations the DG method is slightly superior.

Keywords Shallow water equations · Discontinuous galerkin · Finite volumes · Wetting and drying · Well-balancing

MSC (2010): 65M08 · 35Q86 · 86-08

S. Vater (✉) · J. Behrens
Department of Mathematics, Universität Hamburg, Bundesstraße 55, 20146 Hamburg, Germany
e-mail: stefan.vater@uni-hamburg.de

J. Behrens
e-mail: joern.behrens@uni-hamburg.de

N. Beisiegel
School of Mathematics & Statistics, University College Dublin, Science Centre – North, G.03, Belfield, Dublin 4, Ireland
e-mail: nicole.beisiegel@ucd.ie

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_26

1 Introduction

Coastal inundation processes play a crucial role in geoscientific applications such as the simulation of tsunamis and storm surges. This became repeatedly obvious in recent disasters, for instance the 2011 Japan Tsunami, or the 2013 Super-Taifun Haiyan hitting the Philippines. In order to model such events, the numerical algorithms applied must inherit fundamental properties from the continuous problem, i.e., conservation of mass or the ability to keep the still water state at rest. Even more important is the stability of the schemes, which manifests itself in the positivity of the water depth or the suitability for complex bathymetry configurations.

In this article we compare a newly developed scheme, which is based on the discontinuous Galerkin (DG) discretization [17], to a classical finite volume (FV) method with hydrostatic reconstruction [1] applied to the shallow water equations. The latter has been proved to be very robust in several applications. On the other hand, the DG method has become quite popular in recent years, because of its attractive features such as the possibility to extend the scheme to higher order. But there is still a shortage of robust algorithms when it comes to complex applications such as inundation events at the coast. Besides the few methods developed so far [3, 10, 11, 18] there have been also some comparison studies between FV and DG discretizations in this field [12]. The comparison between different numerical methods is always biased by the choice which discretization details are to be fixed among the schemes. Often, solutions with the same number of degrees of freedom are compared to each other. In the presented results we choose to use the same underlying grid and the same time discretization for both schemes.

The shallow water equations are a system of balance laws for mass and momentum. They are given by

$$\mathbf{U}_t + \mathbf{F}(\mathbf{U})_x = \mathbf{S}(\mathbf{U}) \,, \tag{1}$$

where the vector of unknowns is given by $\mathbf{U} = (h, hu)^T$. The quantity $h = h(x,t)$ denotes the fluid depth of a uniform density water layer and $u = u(x,t)$ is the depth-averaged particle velocity. The flux function is defined by $\mathbf{F}(\mathbf{U}) = (hu, hu^2 + \frac{g}{2}h^2)^T$, where g is the gravitational constant. Furthermore, the bathymetry or bottom topography $b = b(x)$ is taken into account by the source term $\mathbf{S}(\mathbf{U}) = (0, -ghb_x)^T$.

2 Numerical Methods

While the FV method can be directly derived from the principles of conservation of mass and momentum, the DG formulation is based on the weak formulation of the associated system of partial differential equations within each element. This means that (formally) the latter requires more regularity, but it also provides a pointwise solution almost everywhere. The FV computes cell mean values. Both methods

are derived as a semi-discretization in space, where the resulting ordinary differential equation is solved by Heun's method, which is the second-order representative of a standard Runge-Kutta total-variation diminishing (TVD) scheme [8, 14]. In this work, the governing equations (1) are solved on the one-dimensional domain $[x_{\min}, x_{\max}]$, which is divided into intervals (cells) $I_i = (x_{i-1/2}, x_{i+1/2})$.

2.1 Finite Volume Second Order Hydrostatic Reconstruction

Application of conservation of mass and momentum to each cell I_i leads to $\frac{\mathrm{d}\mathbf{U}_i}{\mathrm{d}t} = -\frac{1}{\Delta x_i}\left(\mathbf{F}_{i+1/2} - \mathbf{F}_{i-1/2}\right) + \mathbf{S}_i$, where $\mathbf{U}_i$ and Δx_i are the cell average of $\mathbf{U}$ and the width of cell I_i, respectively. $\mathbf{F}_{i\pm1/2}$ are the fluxes at the cell interfaces, and $\mathbf{S}_i$ is the mean effect of the source term. Note that this formula is still exact. The fluxes are usually computed by the solution of a Riemann problem, applied to reconstructed left and right states at the cell interfaces, i.e., $\mathbf{F}_{i+1/2} = \tilde{\mathbf{F}}_{i+1/2}(\mathbf{U}_{i+1/2,-}, \mathbf{U}_{i+1/2,+})$. To obtain second order accuracy in space, one can reconstruct the inner cell states by central differences, where the gradient must be scaled by a slope limiter to avoid spurious oscillations.

Special care must be taken in the discretization of the source term to maintain balanced states such as the "lake at rest" also in the discrete scheme up to machine accuracy. Here we follow the approach of [1], which essentially involves a hydrostatic reconstruction in the surface elevation $h + b$ instead of the fluid depth h. Furthermore, the source term must be discretized in a way that it exactly cancels the flux divergence when the data is initially balanced.

In this particular study, the Riemann problem is approximately solved by the HLLE solver [6]. The reconstruction is limited by the monotonized central-difference (MC) limiter [13] applied to surface elevation $h + b$ and velocity u.

2.2 RKDG2 with Limiter Based Wetting and Drying

For the derivation of the DG Method the equations from system (1) are multiplied by a test function φ and integrated over each interval. Integration by parts of the flux term leads to the weak DG formulation

$$\int_{I_i} \mathbf{U}_t \varphi \,\mathrm{d}x - \int_{I_i} \varphi_x \mathbf{F}(\mathbf{U}) \,\mathrm{d}x + \left[\mathbf{F}^*(\mathbf{U})\varphi\right]_{x_{i-1/2}}^{x_{i+1/2}} = \int_{I_i} \mathbf{S}(\mathbf{U})\varphi \,\mathrm{d}x \,. \tag{2}$$

Note, that the interface flux $\mathbf{F}^*$ is not defined in general, since the solution can have different values in the adjacent cells. This problem is circumvented by using the (approximate) solution of the corresponding Riemann problem applied to the left and right states. Here, we used the Rusanov solver [13]. System (2) is further discretized in space with a piecewise linear ansatz for the discrete solution components and

test functions. Exact 2-point Gauß-Legendre quadrature is applied for the integrals, which is necessary to achieve well-balancing of the inner cell pressure flux term ghh_x and the source ghb_x in the momentum equation. For a more complete presentations of Runge-Kutta DG methods the reader is referred to e.g., [7, 9].

As in the FV method each Runge-Kutta stage involves a limiting procedure in order to avoid spurious oscillations near discontinuities. This limiting step is further used to obtain a stable discretization of wetting and drying in the DG discretization. The basis is the limiter of [2], which limits the solution within one element such that it does not exceed cell mean values of the surrounding elements. We chose this limiter, because it does not alter a well-balanced state in most circumstances. If an element is fully wet, the depth is adjusted by limiting in the surface elevation $h + b$. In contrast, at the wet/dry interface, where the fluid depth gets under a given wet tolerance at one node, we blend the limiting in $h + b$ with limiting in the fluid depth h. In these cells, the artificial gradient in the surface elevation is further neglected to not introduce spurious waves.

Another problem which arises at the wet/dry interface is the ill-conditioning of the velocity $u = hu/h$, which is due to the fact that fluid depth and momentum go to zero at the same time. In FV methods this is handled by limiting in u instead of hu. For the DG case this is not possible, since the solution (h, hu) is directly modified. Therefore, we compute minimum and maximum values of the velocity in each element on the basis of the surrounding cell mean values and modify the linear momentum distribution according to these thresholds. Full details of this new algorithm are given in [17].

3 Results

In all simulations the gravitational constant is set to $g = 9.81$. Here and below we omit the dimensions of the physical quantities, which should be thought in the standard SI system with m (meter), s (seconds) etc. as basic units. The discrete initial conditions and the bottom topography are derived from the analytical ones by interpolation at the nodal (cell interface) points for the DG discretization and by averaging over each cell for the FV scheme. For both schemes, we set a wet tolerance of the fluid depth to 10^{-8}, under which a state is considered to be dry, and therefore, the velocity is set to zero. In the visualization, we display cell mean values for the FV method and the piecewise polynomial representations for the DG method. We do not show well-balancing results for the "lake at rest" testcase, since this is commonly known for the finite volume hydrostatic reconstruction. For the DG method, this testcase was presented in [17].

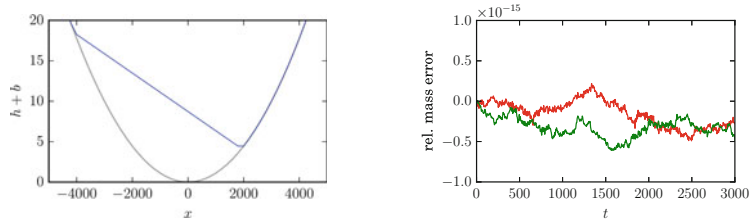


Fig. 1 *Left* Initial free surface elevation of the oscillatory flow in a parabolic bowl. *Right* Mass error of the numerical DG (*red*) and FV (*green*) solutions

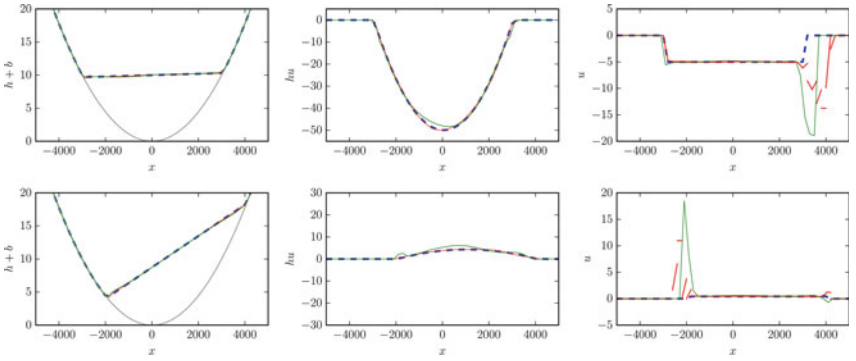


Fig. 2 Free surface elevation, momentum and velocity of an oscillatory flow in a parabolic bowl. DG (*red*), FV (*green*) and exact solution (*blue dashed*) at times $t = 1000$ (*top*), $t = 2000$ (*bottom*)

3.1 Oscillatory Flow in a Parabolic Bowl

A numerically challenging test, for which the analytical solution is known, goes back to [15] and has become a standard test problem for inundation schemes (e.g., [11, 18]). On the domain $[-5000, 5000]$ consider an oscillatory flow in a domain with parabolic bottom topography, which is defined by $b(x) = h_0(x/a)^2$. Here, $a = 3000$ and $h_0 = 10$ define the shape of the parabolic basin. Note that the boundary conditions for the domain are irrelevant since the boundary is in the dry part of the solution. An analytic solution of the water surface is then given by

$$h(x,t) + b(x) = h_0 - \frac{B^2}{4g}\left(1 + \cos(2\omega t)\right) - \frac{Bx}{2a}\sqrt{\frac{8h_0}{g}}\cos\omega t,$$

where we set $\omega = \sqrt{2gh_0}/a$ and $B = 5$. The initial momentum at $t = 0$ is set to zero over the whole domain. The resulting solution involves a periodical movement of the wet/dry interface at both sides of the basin. The problem is discretized on a relatively coarse grid using 50 cells, where differences become visible. The timestep is set to 4.0, which corresponds to a maximum CFL number of 0.3 during the simulation.

The initial data is shown in Fig. 1 (left). Snapshots of the simulation are displayed in Fig. 2 at times 1000 and 2000. One can see that the surface elevation is well approximated by both discretizations and only small differences appear compared to the exact solution. This is different for the momentum variable, where the FV method shows clear deviations from the exact solution. The biggest deviations become visible in the velocity variable which is derived by the quotient of momentum and fluid depth. Note that we have set the velocity to zero when h gets below the wet tolerance. Large values appear, when both, fluid depth and momentum are very small. This occurs especially in the drying process, when the water recedes, and the velocity computation is ill-conditioned. The deviations are nearly twice as large for the FV method compared to the DG discretization, but they do not get as far into the dry area as for the DG method. However, these spurious velocities do neither grow in time, nor seem to influence the overall stability of the scheme. Furthermore, we have displayed the relative global mass error produced by both schemes in Fig. 1 (right). It shows that both methods are conservative up to machine accuracy.

Since efficiency is also of interest in comparing numerical methods, we measured the computation times in this test case. In general it is hard to compare two different codes, since a lot depends on the particular implementation. However, both presented methods were implemented in python and they even share some functionalities. The result is that the DG method takes approximately 16% longer for one time step.

3.2 Tsunami Runup onto a Sloping Beach

In order to test the methods for their applicability to tsunami simulations, the propagation of a tsunami wave onto a uniformly sloping beach is simulated [16]. Besides the slope of the beach the initial surface elevation and momentum with $(hu)(x, 0) \equiv 0$ is given at intervals of 50 (Fig. 3). The solution is sought on the domain $[-500, 50\,000]$ and the bottom topography is set to $b(x) = 5000 - 0.1x$. The crucial task is to correctly simulate the inundation process. The analytical solution can be derived by the initial-value-problem technique introduced in [4]. According to the given data the domain is discretized into cells with size 50. The timestep is 0.05, which approximately corresponds to a CFL number of 0.22 at the deepest point (right side) of the domain.

Figure 4 displays the simulation results at times $t = 160$, 175 and 220. For the surface elevation the exact solution is given for comparison. At time 160 the water

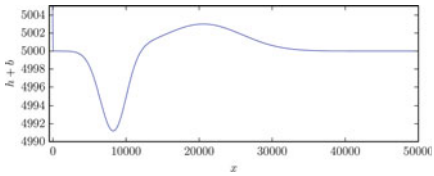


Fig. 3 Tsunami runup onto a sloping beach. Initial surface elevation at $t = 0$

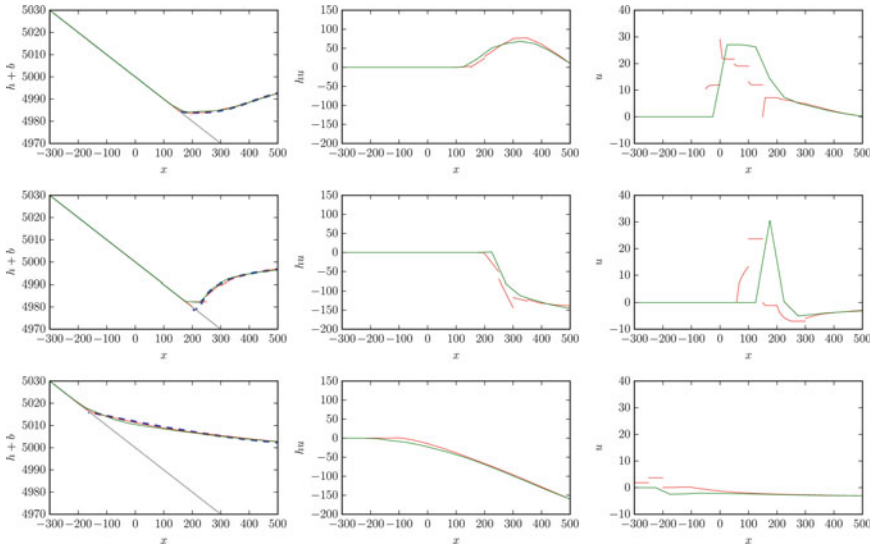


Fig. 4 Tsunami runup onto a sloping beach. Computed surface elevation, momentum and velocity from 1D DG method (*red*) compared to the FV method (*green*). The analytical solution for the surface elevation (*dashed*) at times $t = 160$ (*top*), $t = 175$ (*middle*), $t = 220$ (*bottom*)

still recedes, whereas $t = 175$ is the reversal point between drainage and inundation. At $t = 220$ the coast is still flooded. Also in this test case the surface elevation is well approximated by both discretizations. Small differences can be observed for the momentum between the two methods. As in the previous test case, the spurious velocity deviations can be seen especially in the drying process ($t = 160, 175$). These are approximately of the same size for both discretizations.

3.3 Low-Resolution Bathymetry

As a third testcase we apply both schemes to a supercritical steady flow on an inclined plane with constant slope α. In [5] it was shown that this test case combined with a low-resolution bathymetry is unsuitable for the finite volume hydrostatic reconstruction. The steady state for h and hu is given by

$$(hu)(x,t) = q_0 = \text{const.} \quad \text{and} \quad h^3 + h^2\left(\alpha x - \frac{q_0^2}{2gh_0^2} - h_0\right) + \frac{q_0^2}{2g} = 0,$$

where h_0 and q_0 are the fluid depth and momentum at $x = 0$. The domain is chosen to $[0, 10]$ and the parameters are given by $h_0 = 0.02$ and $q_0 = 0.01$. The simulations are performed with a uniform grid with cell size $\Delta x = 0.1$ and a fixed timestep

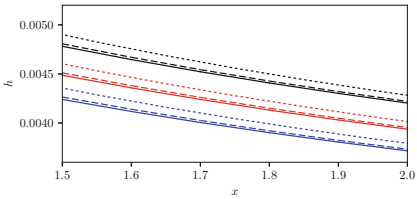


Fig. 5 Fluid depth for supercritical steady flow with different slopes $\alpha = 13\%$ (*black*), 15% (*red*) and 17% (*blue*). exact solution (*solid line*), DG method (*dashed*) and FV method (*dotted*)

which satisfies the CFL condition. The results using three different negative slopes $\alpha = 13\%$, 15% and 17% are presented in Fig. 5. As in [5] the FV method fails to compute the correct fluid depth. The DG method performs much better in this case.

4 Conclusions

In this study we have compared a new treatment to deal with inundation in piecewise linear DG discretizations to a classical FV method with hydrostatic reconstruction. The results show that both schemes are well balanced and are able to compute stable solutions when small perturbations around the still water state at rest are introduced. In case of rapid wetting and drying both schemes result in similar results for the fluid depth variable, even on relatively coarse discretizations. For the momentum variable the DG discretization is superior compared to the FV method. Larger deviations from the exact solution can be seen in the velocity, especially in the drying process, when the water recedes. Although, this is a secondary variable in both schemes, it is crucial for stability because it must be computed in the flux computation and directly enters the CFL stability criterion. However, the velocity remains bounded and does not remarkably influence the stability of the schemes. In general, both schemes display comparable results in terms of accuracy and stability. The DG scheme is more accurate in certain situations. However, it is left to future research how the DG scheme behaves in more complex situations and to test its applicability to two-dimensional problems.

Acknowledgements The authors gratefully acknowledge support through the ASCETE (Advanced Simulation of Coupled Earthquake and Tsunami Events) project sponsored by the Volkswagen foundation and through ASTARTE—Assessment, STrategy And Risk Reduction for Tsunamis in Europe. Grant 603839, 7th FP (ENV.2013.6.4-3)

References

1. Audusse, E., Bouchut, F., Bristeau, M.O., Klein, R., Perthame, B.: A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water flows. SIAM J. Sci. Comput. **25**(6), 2050–2065 (2004). doi:10.1137/S1064827503431090

2. Barth, T.J., Jespersen, D.C.: The design and application of upwind schemes on unstructured meshes. AIAA Paper 89-0366 (1989)
3. Bunya, S., Kubatko, E.J., Westerink, J.J., Dawson, C.: A wetting and drying treatment for the Runge-Kutta discontinuous Galerkin solution to the shallow water equations. Comput. Methods Appl. Mech. Eng. **198**, 1548–1562 (2009). doi:10.1016/j.cma.2009.01.008
4. Carrier, G.F., Wu, T.T., Yeh, H.: Tsunami run-up and draw-down on a plane beach. J. Fluid Mech. **475**, 79–99 (2003). doi:10.1017/S0022112002002653
5. Delestre, O., Cordier, S., Darboux, F., James, F.: A limitation of the hydrostatic reconstruction technique for shallow water equations. C. R. Acad. Sci. Paris Ser. I **350**, 677681 (2012)
6. Einfeldt, B.: On Godunov-type methods for gas dynamics. SIAM J. Numer. Anal. **25**(2), 294–318 (1988). doi:10.1137/0725021
7. Giraldo, F.X., Warburton, T.: A high-order triangular discontinuous Galerkin oceanic shallow water model. Int. J. Numer. Methods Fluids **56**(7), 899–925 (2008). doi:10.1002/fld.1562
8. Gottlieb, S., Shu, C.W., Tadmor, E.: Strong stability-preserving high-order time discretization methods. SIAM Rev. **43**(1), 89–112 (2001). doi:10.1137/S003614450036757X
9. Hesthaven, J.S., Warburton, T.: Nodal Discontinuous Galerkin Methods: Algorithms, Analysis, and Applications. Springer (2008)
10. Kärnä, T., de Brye, B., Gourgue, O., Lambrechts, J., Comblen, R., Legat, V., Deleersnijder, E.: A fully implicit wetting-drying method for DG-FEM shallow water models, with an application to the scheldt estuary. Comput. Methods Appl. Mech. Eng. **200**(5–8), 509–524 (2011). doi:10.1016/j.cma.2010.07.001
11. Kesserwani, G., Liang, Q.: Well-balanced RKDG2 solutions to the shallow water equations over irregular domains with wetting and drying. Comput. Fluids **39**, 2040–2050 (2010). doi:10.1016/j.compfluid.2010.07.008
12. Kesserwani, G., Wang, Y.: Discontinuous Galerkin flood model formulation: Luxury or necessity? Water Resour. Res. **50**(8), 6522–6541 (2014). doi:10.1002/2013WR014906
13. LeVeque, R.J.: Finite Volume Methods for Hyperbolic Problems, Cambridge Texts in Applied Mathematics, vol. 31. Cambridge University Press (2002)
14. Shu, C.W., Osher, S.: Efficient implementation of essentially non-oscillatory shock-capturing schemes. J. Comput. Phys. **77**(2), 439–471 (1988). doi:10.1016/0021-9991(88)90177-5
15. Thacker, W.C.: Some exact solutions to the nonlinear shallow-water wave equations. J. Fluid Mech. **107**, 499–508 (1981). doi:10.1017/S0022112081001882
16. The Third International Workshop on Long-Wave Runup Models: Benchmark problem #1: Tsunami runup onto a plane beach (2004). http://isec.nacse.org/workshop/2004_cornell/bmark1.html
17. Vater, S., Beisiegel, N., Behrens, J.: A limiter-based well-balanced discontinuous Galerkin method for shallow-water flows with wetting and drying: One-dimensional case. Adv. Water Resour. **85**, 1–13 (2015). doi:10.1016/j.advwatres.2015.08.008
18. Xing, Y., Zhang, X., Shu, C.W.: Positivity-preserving high order well-balanced discontinuous Galerkin methods for the shallow water equations. Adv. Water Resour. **33**(12), 1476–1493 (2010). doi:10.1016/j.advwatres.2010.08.005

A Discontinuous Galerkin Method for Non-hydrostatic Shallow Water Flows

Anja Jeschke, Stefan Vater and Jörn Behrens

Abstract In this work a non-hydrostatic depth-averaged shallow water model is discretized using the discontinuous Galerkin (DG) Method. The model contains a non-hydrostatic pressure component, similar to Boussinesq-type equations, which allows for dispersive gravity waves. The scheme is a projection method and consists of a predictor step solving the hydrostatic shallow water equations by the Runge-Kutta DG method. In the correction the non-hydrostatic pressure component is computed by satisfying a divergence constraint for the velocity. This step is discretized by application of the DG discretization to the first order elliptic system. The numerical tests confirm the correct dispersion behavior of the method, and show its validity for simple test cases.

Keywords Shallow water equations · Non-hydrostatic · Discontinuous galerkin method

MSC (2010): 65M08 · 35Q86 · 86-08

A. Jeschke · S. Vater (✉) · J. Behrens
Department of Mathematics, Universität Hamburg, Bundesstraße 55,
20146 Hamburg, Germany
e-mail: stefan.vater@uni-hamburg.de

A. Jeschke
e-mail: anja.jeschke@uni-hamburg.de

J. Behrens
e-mail: joern.behrens@uni-hamburg.de

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_27

1 Introduction

Exact numerical modeling of long surface gravity waves is important to understand and predict physical phenomena such as tsunamis and storm surges. For its description, the shallow water equations provide a good first approximation, but to understand the behavior of the next smaller scale, a more detailed view is in need. In general, surface gravity waves are dispersive or "non-hydrostatic", meaning that waves of different wave lengths travel at different wave speeds. Our goal is to present a numerical discretization of the so called depth-averaged non-hydrostatic extension for shallow water equations of [9] in the context of a discontinuous Galerkin discretization.

Discontinuous Galerkin methods are promising, because they combine the local conservation property and the locality of finite volume methods with the ability to use higher order basis functions as in (continuous) finite element methods. These approximations are discontinuous along element boundaries and only connected by numerical fluxes. Here we apply the discontinuous Galerkin method presented in [16], where it has been used to derive a robust inundation treatment at the wet/dry interface.

The non-hydrostatic extension for shallow water equations is a system of equations to describe the fluid flow of long dispersive surface gravity waves. A projection method to solve the equations stepwise forms the basis of the numerical method. A splitting of the pressure into a hydrostatic and a non-hydrostatic component [2, 12] is advantageous, because the predictor step can resort on the shallow water equations. The corrector step implements the solution of an elliptic equation for the non-hydrostatic pressure in each timestep. There are multi-layer models [6, 13] as well as depth-averaged models implementing a finite difference [1], finite volume [4] or finite element [7, 17] scheme. Also discontinuous Galerkin discretizations for 3D non-hydrostatic [15] and Boussinesq-type [5] models have been introduced. Here, we present a novel DG discretization of the conservative formulation of the depth-averaged non-hydrostatic extension for shallow water equations presented in [9].

In the following the continuous conservative equations are introduced followed by the description of the projection method and discontinuous Galerkin discretization. Numerical results confirm the ability of the scheme to deal with idealized test situations before we conclude our results in the final section.

2 Conservative Depth-Averaged Non-hydrostatic Model

The one-dimensional shallow water equations in conservative formulation are given as

$$\begin{pmatrix} h \\ hu \end{pmatrix}_t + \begin{pmatrix} hu \\ hu^2 + \frac{g}{2}h^2 \end{pmatrix}_x = \begin{pmatrix} 0 \\ -ghb_x \end{pmatrix} \tag{1}$$

with the vector of unknowns $(h, hu)^T$ and the gravitational constant g. The fluid depth $h = h(x, t)$ is measured from a bathymetry $b = b(x)$ to the water surface. The quantity $u = u(x, t)$ denotes the depth-averaged horizontal fluid velocity. We assume that the water density ρ is constant.

The derivation of the conservative formulation of the depth-averaged non-hydrostatic extension for shallow water equations follows the derivation given in [9] for primitive variables. To extend the conservative formulation to the non-hydrostatic regime, the non-hydrostatic pressure P^{nh} and the vertical velocity W have to be taken into account in the derivation starting from Euler equations of motion and splitting the pressure $P = P^{nh} + P^{hy}$ into a non-hydrostatic component P^{nh} and a hydrostatic component P^{hy}. We assume $P = P^{nh} = 0$ at the water surface. We derive the system of equations in depth-averaged unknowns only and express the non-hydrostatic pressure at the bottom P_b^{nh} in terms of the depth-averaged non-hydrostatic pressure p^{nh}. This results in $P_{-d}^{nh} = f_{\mathrm{nh}} p^{nh} + f_{\mathrm{d}}$, where the scalar f_{nh} determines the vertical profile of the non-hydrostatic pressure P^{nh}. A linear vertical profile ($f_{\mathrm{nh}} = 2$) is traditionally assumed in literature concerning non-hydrostatic models. The adaptation of the profile to be quadratic ($f_{\mathrm{nh}} = 1.5$) as in Boussinesq-type models yields equivalence [9] of non-hydrostatic models to some well-known Boussinesq-type equations. The quadratic profile is also the correct profile in the long wave limit and the linear pressure profile results in a wrong wave front being too high and short. The scalar f_{d} has to be chosen properly to receive equivalence in case of non-constant bathymetry.

Due to the non-hydrostatic pressure, the vertical velocity W is not negligible anymore, which is different than in hydrostatic regime. Here, the vertical velocity W has to fulfill the kinematic boundary conditions

$$W_{h+b} = h_t + u(h + b)_x, \tag{2}$$

$$W_b = ub_x. \tag{3}$$

at the water surface and the fluid bottom.

In order to close the system of equations and to apply the projection method later on, we have to introduce a divergence constraint as an additional equation. As in [9], we assume the vertical profile of W to be linear and denote the depth-averaged vertical velocity with w and the depth-averaged non-hydrostatic pressure with p^{nh}. A linear vertical profile of W is the easiest one to fulfill the given kinematic boundary conditions, and it is often assumed in the literature by other authors. Furthermore it is consistent with the equations, if the horizontal velocities are approximated by

their mean values which are independent from z. Instead of deriving the additional equation in terms of primitive depth-averaged variables h, u, w, we express it in terms of the conservative variables h, hu, hw. Using the kinematic boundary conditions to compute the depth-averaged vertical velocity together with the mass equation leads to the desired divergence constraint (5). All together, the conservative formulation of the depth-averaged non-hydrostatic extension for shallow water equations results in

$$\begin{pmatrix} h \\ hu \\ hw \end{pmatrix}_t + \begin{pmatrix} hu \\ hu^2 + \frac{g}{2}h^2 + \frac{1}{\rho}hp^{nh} \\ hwu \end{pmatrix}_x = \begin{pmatrix} 0 \\ -ghb_x - \frac{1}{\rho}\left(f_{\mathrm{nh}}p^{nh} + f_{\mathrm{d}}\right)b_x \\ \frac{1}{\rho}\left(f_{\mathrm{nh}}p^{nh} + f_{\mathrm{d}}\right) \end{pmatrix} \tag{4}$$

$$2hw - hu(h+2b)_x = -h(hu)_x, \tag{5}$$

where the scalars f_{nh} and f_{d} determine the vertical profile of the non-hydrostatic pressure.

3 Discretization

In our novel approach, system (4)–(5) is solved as a projection method using a discontinuous Galerkin discretization in space [3, 8]. First, an auxiliary system is solved by neglecting the non-hydrostatic pressure component. In a second step the non-hydrostatic pressure is computed, which provides a correction of the velocity field in order to be in compliance with the divergence constraint. For the discretization, we divide the domain $\Omega = [x_l, x_r]$ into uniform cells $I_i = [x_i, x_{i+1}]$.

Neglecting the non-hydrostatic pressure p^{nh} in (4) results in the (hydrostatic) shallow water equations (1) augmented with a passive tracer equation for the vertical momentum hw. This is the auxiliary system to solve in the predictor step. Here, we use a second-order Runge-Kutta DG (RKDG2) method, which was previously presented in [16] in the context of inundation problems. For the DG discretization, each equation is multiplied by a test function φ and integrated over one cell. Integration by parts of the flux term leads to the weak DG formulation

$$\int_{I_i} \mathbf{U}_t \varphi \,\mathrm{d}\mathbf{x} - \int_{I_i} \mathbf{F}(\mathbf{U}) \cdot \nabla\varphi \,\mathrm{d}\mathbf{x} + \int_{\partial I_i} \mathbf{F}^*(\mathbf{U}) \cdot \mathbf{n}\, \varphi \,\mathrm{d}S = \int_{I_i} \mathbf{S}(\mathbf{U})\varphi \,\mathrm{d}\mathbf{x}\,, \tag{6}$$

where $\mathbf{U} = (h, hu, hw)$ is the vector of unknowns, and $\mathbf{n}$ is the outward pointing normal at the edge of the cell I_i. The interface flux $\mathbf{F}^*$ is not defined in general, as the solution can have different values at the interface in the adjacent elements. This problem is circumvented in the discretization by using the (approximate) solution of the corresponding Riemann problem. For the simulations in this study we used the Rusanov solver [14]. The resulting semi-discrete system is solved by Heun's method, which is the second-order representative of a standard Runge-Kutta total-variation diminishing (TVD) scheme [11].

For the correction, we use an implicit Euler discretization. This means that the final momentum is computed according to

$$(hu)^{n+1} = (\widetilde{hu})^{n+1} - \frac{\Delta t}{\rho}\left[(hp^{nh})_x + \left(\frac{f_{\mathrm{nh}}}{h}(hp^{nh}) + f_{\mathrm{d}}\right) b_x\right], \tag{7}$$

$$(hw)^{n+1} = (\widetilde{hw})^{n+1} - \frac{\Delta t}{\rho}\left[\frac{f_{\mathrm{nh}}}{h}(hp^{nh}) + f_{\mathrm{d}}\right], \tag{8}$$

where the values with tilde are the one obtained from the predictor step. Substituting (8) into the divergence constraint (5), we obtain, together with (7), an elliptic system of first order differential equations for the unknowns $(hu)^{n+1}$ and (hp^{nh}). An update of the water height follows the solution of the elliptic system. This system is again discretized by the DG approach. Both equations are multiplied by a test function and integrated over each cell to obtain a weak formulation. Integration by parts is applied to the terms involving first derivatives of the unknowns. The arising numerical traces at the interfaces of each cell are chosen according to [3].

4 Numerical Results

We present two test cases to validate our numerical model. These are the standing wave and the propagating solitary wave. The gravitational constant is set to $g = 9.81$. For simplicity we omit the physical dimensions throughout this manuscript. We use piecewise linear polynomials for the DG discretization in space.

4.1 Standing Wave

The standing wave is the analytical solution

$$h(x,t) = d - a\sin(\kappa x)\cos(\kappa c t)\,, \tag{9}$$

$$hu(x,t) = ac\cos(\kappa x)\sin(\kappa c t)\,, \tag{10}$$

to the linearized system of both the shallow water and the non-hydrostatic extension for shallow water equations around the state of a constant background height d and a zero velocity field. The phase velocity is given by $c = c_{\mathrm{sw}} = \sqrt{gd}$ for the hydrostatic equation set and $c = c_{\mathrm{sw}}\left(1 + \frac{(\kappa d)^2}{2f_{\mathrm{nh}}}\right)^{-0.5}$ for the non-hydrostatic equation

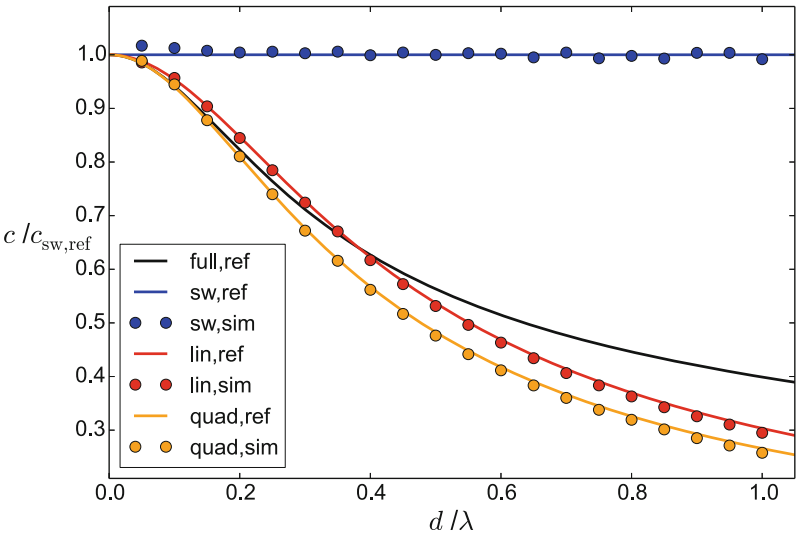


Fig. 1 Comparison of simulated and analytical phase velocities using the standing wave initial condition

set, respectively. We run the non-linear model with parameters $a = 0.01$ and $\kappa = \frac{\pi}{10}$. In different model runs, we vary d to compute the phase velocity (see Fig. 1) for different ratios of water depth d and wave length $\lambda = \frac{2\pi}{\kappa}$. We choose periodic boundary conditions and the CFL stability constant as CFL≈ 0.07. The simulated phase velocities fit to the analytical ones.

4.2 Propagating Solitary Wave

The non-hydrostatic extension for shallow water equations using the quadratic vertical pressure profile has an analytic solitary wave solution on constant bathymetry. This fact comes from their equivalence to the Serre equations [9]. The analytic solitary wave solution is (see [10])

$$h(x,t) = d + a\cosh^{-2}(K(x - ct - x_0)), \tag{11}$$

$$hu(x,t) = c\,(h(x,t) - d) \tag{12}$$

with the amplitude $a = 2.0$, the propagation velocity $c = \sqrt{g(d+a)}$ on a constant depth $d = 10.0$, scale factor $K = \sqrt{\left(\frac{3a}{4d^2(d+a)}\right)}$ and displacement $x_0 = 200.0$ on a domain of length $l = 800.0$. We choose periodic boundary conditions and the CFL

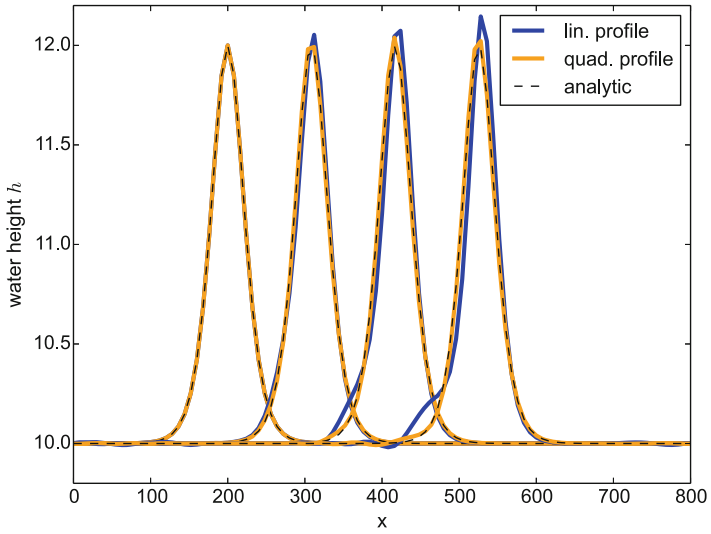


Fig. 2 Comparison of the analytical (*black dashed*) water height of the solitary wave with the simulation results of the quadratic (*yellow*) and linear (*blue*) initial vertical profile and those obtained after a propagation time of 10, 20 and 30 seconds to the *right*

stability constant as CFL= 0.14. Figure 2 shows that the simulated wave applying the quadratic vertical pressure profile keeps its shape as expected. In contrast, the linear vertical pressure leads to a too small dispersion and therefore to a steepening of the wave and the generation of a dispersive trail. This is in line with the results presented in [9].

5 Conclusion

We propose a novel discretization for the depth-averaged non-hydrostatic extension for shallow water equations to model dispersive fluid flow. Our specific version of this system of equations has a free parameter to choose the vertical profile of the non-hydrostatic pressure to be linear or quadratic. To the author's knowledge, it is the first Runge-Kutta discontinuous Galerkin discretization for this type of equations.

In the projection method applied, the predictor step consists of the solution of the hydrostatic shallow water equations in conservative variables. The corrector step involves the solution of a first order elliptic system for the non-hydrostatic pressure. The elliptic system is constructed to fulfill a divergence constraint for the velocity and it is also discretized using the discontinuous Galerkin approach.

Numerical tests confirm the stable and accurate behavior of the model, which are similar to the results presented in [9]. The discretization is able to correctly approximate analytical solutions of the non-hydrostatic shallow water model, and verified its correct dispersive behavior represented by the linear or the quadratic vertical pressure profile, respectively.

Acknowledgements The authors A.J. and J.B. want to thank the European Union, who funded this work within the project ASTARTE—Assessment, Strategy And Risk Reduction for Tsunamis in Europe—FP7-ENV2013 6.4-3, Grant 603839. The authors J.B. and S.V. acknowledge additional support through the ASCETE project, funded by the Volkswagen Foundation.

References

1. Bai, Y., Cheung, K.F.: Depth-integrated free-surface flow with parameterized non-hydrostatic pressure. Int. J. Numer. Methods Fluids **71**(4), 403–421 (2013). doi:10.1002/fld.3664
2. Casulli, V., Stelling, G.: Numerical simulation of 3D quasi-hydrostatic, free-surface flows. J. Hydraul. Eng. **124**(7), 678–686 (1998)
3. Cockburn, B.: Discontinuous galerkin methods. Zeitschrift fr Angewandte Mathematik und Mechanik **83**(11), 731–754 (2003). doi:10.1002/zamm.200310088
4. Cui, H., Pietrzak, J., Stelling, G.: Optimal dispersion with minimized poisson equations for non-hydrostatic free surface flows. Ocean Model. **81**, 1–12 (2014). doi:10.1016/j.ocemod.2014.06.004
5. Dumbser, M., Facchini, M.: A space-time discontinuous Galerkin method for Boussinesq-type equations. Appl. Math. Comput. Part 2 **272**, 336–346 (2016). doi:10.1016/j.amc.2015.06.052
6. Fringer, O., Gerritsen, M., Street, R.: An unstructured-grid, finite-volume, nonhydrostatic, parallel coastal ocean simulator. Ocean Model. **14**(3), 139–173 (2006)
7. Fuchs, A.: Effiziente parallele Verfahren zur Lösung verteilter, dünnbesetzter Gleichungssysteme eines nichthydrostatischen Tsunamimodells. Ph.D. thesis, AWI, Universität Bremen (2013). http://elib.suub.uni-bremen.de/edocs/00103439-1.pdf
8. Hesthaven, J.S., Warburton, T.: Nodal Discontinuous Galerkin Methods: Algorithms, Analysis, and Applications. Springer Publishing Company, Incorporated (2008). doi:10.1007/978-0-387-72067-8
9. Jeschke, A., Pedersen, G.K., Vater, S., Behrens, J.: Depth-averaged non-hydrostatic extension for shallow water equations with quadratic vertical pressure profile: Equivalence to boussinesq-type equations. Int. J. Numer. Methods Fluids (2017). doi:10.1002/fld.4361. http://dx.doi.org/10.1002/fld.4361. (In press)
10. Seabra-Santos, F.J., Renouard, D.P., Temperville, A.M.: Numerical and experimental study of the transformation of a solitary wave over a shelf or isolated obstacle. J. Fluid Mech. **176**, 117–134 (1987). doi:10.1017/S0022112087000594
11. Shu, C.W., Osher, S.: Efficient implementation of essentially non-oscillatory shock-capturing schemes. J. Comput. Phys. **77**(2), 439–471 (1988). doi:10.1016/0021-9991(88)90177-5
12. Stansby, P.K., Zhou, J.G.: Shallow-water flow solver with non-hydrostatic pressure: 2d vertical plane problems. Int. J. Numer. Methods Fluids **28**(3), 541–563 (1998). doi:10.1002/(SICI)1097-0363(19980915)28:3<541::AID-FLD738>3.0.CO;2-0
13. Stelling, G., Zijlema, M.: An accurate and efficient finite-difference algorithm for non-hydrostatic free-surface flow with application to wave propagation. Int. J. Numer. Methods Fluids **43**(1), 1–23 (2003). doi:10.1002/fld.595
14. Toro, E.F.: Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction, 3 edn. Springer (2009)

15. Ueckermann, M., Lermusiaux, P.: Hybridizable discontinuous Galerkin projection methods for NavierStokes and Boussinesq equations. J. Comput. Phys. **306**, 390–421 (2016). doi:10.1016/j.jcp.2015.11.028
16. Vater, S., Beisiegel, N., Behrens, J.: A limiter-based well-balanced discontinuous Galerkin method for shallow-water flows with wetting and drying: one-dimensional case. Adv. Water Resour. **85**, 1–13 (2015). doi:10.1016/j.advwatres.2015.08.008
17. Walters, R.A.: A semi-implicit finite element model for non-hydrostatic (dispersive) surface waves. Int. J. Numer. Methods Fluids **49**(7), 721–737 (2005). doi:10.1002/fld.1019

Design of a Second-Order Fully Explicit Residual Distribution Scheme for Compressible Multiphase Flows

Rémi Abgrall and Paola Bacigaluppi

Abstract The design of a fully explicit second-order scheme applied to the framework of non-conservative time dependent 1D hyperbolic problems, in the context of compressible multiphase flows with strong interacting discontinuities, is presented. The aim is to investigate an explicit second-order approximation for a non-conservative system, given by the five equation model of Kapila et al. (*Physics of Fluids* 2001). The discretization is based on a predictor-corrector scheme, which follows the concept of residual distributions in Ricchiuto and Abgrall (*J. Comp. Physics* 2010). The novelty of this work is the capability of the presented approximation to provide mesh convergence and to be easily extended to 2D and unstructured meshes. A benchmark on the two-phase compressible system for a stiffened gas verifies the robustness and convergence to the expected solution of the presented approximation.

Keywords Explicit scheme · Residual distributions · High-order methods · Compressible multiphase flows

MSC (2010): 65M60 · 76T99

1 Introduction

In recent years, the ability of models to approximate the interface of multiphase compressible flows has gained increasingly interest. In particular, one of the main issues coming along with the study of multiphase flows is the necessity of employing non-conservative schemes, due to the inability of conservative ones to preserve solutions from large numerical errors on the pressures and velocities [1, 5, 14]. The

R. Abgrall · P. Bacigaluppi (✉)
Institute of Mathematics, University of Zurich, Winterthurerstrasse 190,
8057 Zürich, Switzerland
e-mail: paola.bacigaluppi@math.uzh.ch

R. Abgrall
e-mail: remi.abgrall@math.uzh.ch

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_28

non-conservative character of the problem set has been found to be an open issue. Many difficulties arise when looking for compatible jump relations [3] to be able to guarantee a sufficiently accurate approximation and many methods, as in [15, 16], have been studied. The following paper treats a five equation model originally proposed by Kapila et al. [4], where they consider the formal limit [8] of the Baer and Nunziato seven equation model when the relaxation parameters tend simultaneously to infinity, and thus permitting to assume a single pressure and velocity. This is a well studied approach, among which many contributions, as [6, 9, 10, 13, 15–18], can be counted. During the last decades, among the wide framework of spatial discretizations, residual distributions have proven to be a very solid and easy-to-implement way to discretize PDEs. It is known that any finite volume and finite element scheme can be rewritten in terms of a distribution of the residual with a technical ease to achieve second order accuracy in space, while maintaining a compact stencil and, moreover, be easily put in parallel form.

Inspired by several works presented in recent years [11, 20], high-order schemes can be obtained by applying a first step, where an approximated result is initially predicted on the basis of the sole flux, and a second step that uses the predictive approximation to increase the accuracy of the solution.

The main goal of this work is to show how a non-conservative hyperbolic multiphase system can easily achieve a higher than one-order approximation with a prediction-correction method, following the work of Ricchiuto and Abgrall [11]. Adopting for the space discretization a finite element based residual distribution scheme, the second-order in space is achieved by applying a Lax Friedrichs scheme, while a limiter provides second-order accuracy across shocks, reducing, at least formally, the numerical oscillations coming along with discontinuities.

This paper is divided into four sections. In Sect. 2 we present the modeling equations and give some definitions and assumptions for the problem set. In Sect. 3 we describe the fully second-order explicit residual distribution scheme. Finally, we compare the results of the proposed approximation to an exact solution and present our conclusive remarks.

2 The Five Equation Model

To model a compressible two-phase flow, Kapila et al. [4] and Murrone et al. [8] rewrote the well-known Baer and Nunziato seven equation system [2] into a five equation model, assuming a stiff mechanical relaxation. To reduce further complexity, the proposed five equation model considers the absence of mass and heat transfer and is discussed for 1D only, reading

$$\begin{cases} \frac{\partial \alpha_1}{\partial t} + u\frac{\partial \alpha_1}{\partial x} - K\frac{\partial u}{\partial x} = 0 \\ \frac{\partial (\alpha_1 \rho_1)}{\partial t} + \frac{\partial (\alpha_1 \rho_1 u)}{\partial x} = 0 \\ \frac{\partial (\alpha_2 \rho_2)}{\partial t} + \frac{\partial (\alpha_2 \rho_2 u)}{\partial x} = 0 \\ \frac{\partial (\rho_{tot} u)}{\partial t} + \frac{\partial (\rho_{tot} u^2 + P)}{\partial x} = 0 \\ \frac{\partial E_{tot}}{\partial t} + \frac{\partial (E_{tot} + P)u}{\partial x} = 0 \end{cases} \tag{1}$$

The first equation of (1) describes the transport law for the volume fraction α_1. This first law displays the non-conservative behaviour of the chosen problem set, which, as mentioned in the introductory section, is a key issue for finding an accurate approximation to the problem.

The second and third equation represent the mass conservation law for each phase k, where ρ_k and u_k represent for each phase the density and the velocity respectively. Due to the assumed relaxation, the velocity is the same for each phase $u_1 = u_2 = u$ and same for the pressure $P_1 = P_2 = P$. This leads to the possibility to rewrite the momentum and energy conservation laws in terms of mixture terms, that obey to the relations $\sum_k \alpha_k = 1$ for the volume fractions, and $\sum_k \alpha_k \rho_k = \rho_{tot}$ for the total density. Further, the total energy E_{tot} is given by the sum of the internal energy $e = \sum_k \alpha_k e_k$ and kinetic energy $e_{kin} = \frac{1}{2}\rho_{tot} u^2$. The parameter K writes $K = \frac{\alpha_1 \alpha_2 (\rho_2 c_2^2 - \rho_1 c_1^2)}{\alpha_1 \rho_2 c_2^2 + \alpha_2 \rho_1 c_1^2}$.

The Wood velocity $\frac{1}{\rho_{tot} c_{tot}^2} = \frac{\alpha_1}{\rho_1 c_1^2} + \frac{\alpha_2}{\rho_2 c_2^2}$ describes the total sound speed, where $c_k = \left(\frac{\partial P_k}{\partial \rho_k}\right)\Big|^{\frac{1}{2}}_{Entropy\ in\ k}$ is the squared speed of sound for phase k.
The set of Eqs. (1) is closed by an equation of state for each phase k.

3 High-Order Explicit Residual Distribution Scheme

The five equation model (1) can be rewritten in the compact form

$$\begin{cases} \partial_t V + \nabla \cdot \mathscr{F}(V) + A \cdot \nabla V = 0 \quad \text{on} \quad \Omega \times [0, T] \\ V^0 \text{ initial solution at } t_0 = 0 \end{cases} \tag{2}$$

with $V = [\alpha_1 \rho_1,\ \alpha_2 \rho_2,\ u\rho_{tot},\ E_{tot}]^T$ and $\mathscr{F}(V)$ the corresponding fluxes. A reads $A_{1,s} = [u,\ \frac{K\,u}{\rho_{tot}},\ \frac{K\,u}{\rho_{tot}},\ -\frac{K}{\rho_{tot}},\ 0]^T$ and $A_{s,q} = 0$, where $s = 2, .., 5$ and $q = 1, .., 5$.

3.1 Residual Distribution Scheme

System (2) is discretized on a finite elements based residual distribution scheme, by considering the time domain $[0, T]$ split into N intervals $0 = t_0 < t_1 < ... < t_n < ... < t_N = T$. Each of these intervals contains further M sub-intervals $t_n = t_{n,0} <$

$t_{n,1} <...< t_{n,m} <...< t_{n,M} = t_{n+1}$, where the sub-timestep is $\Delta t_{n,m} = t_{n,m+1} - t_{n,m}$. Let Ω be the spatial domain discretised by Ω_h, and denote by Q a generic element of the mesh and h the generic characteristic mesh size, for simplicity of same length Δx. Consider piecewise linear polynomials V_h spanned by the basis functions φ_i on a node $i \in \Omega_h$. Denote by $\mathbb{P}^1(Q)$ the Lagrange approximation of a C^1 function, which verifies the standard conditions $\varphi_i(x_j, y_j) = \delta_{i,j}, \ \forall i, j \in \Omega_h$, and $\sum_{i \in Q} \varphi_i(x, y) = 1, \ \forall Q \in \Omega_h$. The approximation of V_h is denoted at a certain time $t_{n,m}$ for a single element by $V_h^{n,m}$ and is given by $V_h^{n,m} = \sum_{i \in \Omega_h} V_i^{n,m} \varphi_i = \sum_{Q \in \Omega_h} \sum_{i \in Q} V_i^{n,m} \varphi_i$.

The residual $r_h = \frac{V_h^{n,m+1} - V_h^{n,m}}{\Delta t_{n,m}} + \nabla \cdot \mathscr{F}_h(V_h^{n,m}, V_h^{n,m+1})$ is defined by a spatial and temporal contribution and is built for each single node composing one element, which, in the specific case of this work, is two for each cell. The integral over the residual is denoted as the fluctuation term $\Phi^Q = \int_Q r_h$. The fluctuation is distributed to a local nodal residual ϕ_i^Q in each node belonging to an element through a bounded distribution matrix coefficient β_i^Q, and such that $\phi_i^Q = \beta_i^Q \Phi^Q$, and $\Phi^Q(V_h^n) = \sum_{i \in Q} \phi_i^Q, \ \forall Q \in \Omega_h, \forall \Omega_h$.

According to [12], the key characterizing each residual distribution scheme from one another is the specific choice of the β_i^Q, under the consistency requirement $\sum_{i \in Q} \beta_i^Q = 1$.

3.2 A Second-Order Fully Explicit Scheme

Following [11], we choose the time-substeps to be $M = 2$. The method, also known as a predictor-corrector approximation, reads

$$|C_i| \frac{V_i^{n,m+1} - V_i^{n,m}}{\Delta t_{n,m}} + \sum_{Q \in Q_i} \phi_i^Q(V_h^{n,m}, V_h^{n,m+1}) = 0, \quad \forall i \in \Omega_h \tag{3}$$

where $|C_i|$ is the area of the median dual cell C_i obtained by joining the gravity centres of the cells with the midpoints of the edges meeting in i.
At the prediction step, $V_i^{n,1} = V_i^{n,0}$ is first order accurate and given by a flux difference. The second-order in time is then achieved with the correction step, which considers the obtained prediction approximation as the previous sub-timestep solution. The second-order in space is guaranteed by a correct choice of approximating the nodal residual term ϕ_i^Q.

Across strong interacting discontinuities, the approximation guarantees a second-order accuracy through the choice of limiting the distribution matrix coefficient β_i^Q. First, the local nodal residual ϕ_j^Q, for brevity ϕ_j, is computed (see Sect. "A Comment on the Nodal Residual" for more detail) on a cell, which in the specific case of a 1D stencil is given for $j = i, i+1$. Then, ϕ_j is rewritten in form of an eigen-decomposition $\phi_j = \sum_{s=1}^{5} l_s(\phi_j) R_s$, where $l_s(\phi_j)$ are the left eigenvectors corresponding to ϕ_j for each right eigenvector R_s.

Inspired by [11, 19], to enforce the invariance along the characteristics, the left eigenvector is changed to

$$(l_s(\phi_j))^\star = (1-\xi)\,\frac{\max\left(0,\ \dfrac{l_s(\phi_j)}{\sum_{z=i}^{i+1} l_s(\phi_z)}\right)}{\sum_{r=i}^{i+1}\max\left(0,\ \dfrac{l_s(\phi_r)}{\sum_{z=i}^{i+1} l_s(\phi_z)}\right)}\ \sum_{z=i}^{i+1} l_s(\phi_z) + \xi\ l_s(\phi_j) \tag{4}$$

for each $j = i, i+1$ and where $\xi = \dfrac{|\sum_{z=i}^{i+1} l_s(\phi_z)|}{\sum_{z=i}^{i+1}|l_s(\phi_z)|}$.

Finally, the explicit residual distribution scheme is updated by $\phi_j^\star = \sum_s (l_s(\phi))_j^\star R_s$. The second order in space and time, as well as the fact that the scheme is not (formally) oscillating, is guaranteed by ξ ranging between 0 and 1.

3.3 A Comment on the Nodal Residual

The reader might have noticed, that so far no details have been given on how actually the ϕ_j is approximated. The strength of this method is indeed its capability to be suited for all types of schemes, which makes it particularly portable. In this specific work, the validation is carried out by using a Lax Friedrich's scheme, such that

$$\phi_j = \phi_i^Q(V_h^{n,m}, V_h^{n,m+1}) = \frac{1}{N_Q}\int_Q (V_h^{n,m+1} - V_h^{n,m}) + \int_{\partial Q} \mathscr{F}(V_h^{n,m}, V_h^{n,m+1})\cdot\mathbf{n} + \theta_Q(\widetilde{V_h} - \overline{V}_Q) \tag{5}$$

for $j = i, i+1$, with N_Q the number of degrees of freedom in Q and $\widetilde{V}_h = \frac{V_h^{n,m+1}+V_h^{n,m}}{2}$, $\overline{V}_Q = \frac{1}{2}\Big(\sum_{j=i}^{i+1}\frac{1}{2}\big(V_j^n + V_j^\star\big)\Big)$. Note that θ_Q is in the form of a characteristic velocity. The surface integral for the flux is defined as

$$\psi = \int_{\partial Q} \mathscr{F}(V_h^{n,m}, V_h^{n,m+1})\cdot\mathbf{n} = \frac{1}{2}\Big[\big(F^{n,m} + F^{n,m+1}\big)_{i+1} - \big(F^{n,m} + F^{n,m+1}\big)_i\Big] \tag{6}$$

with the flux $F = [\alpha_1\rho_1 u,\ \ \alpha_2\rho_2 u,\ \ \rho_{tot}u^2 + P,\ \ (E_{tot}+P)u]^T$.

To treat the non-conservative transport equation for the volume fraction, (5) is applied with the only change of considering the space discretisation to be

$$\overline{u}\Big[\frac{1}{2}\big(\alpha_1^{n,0} + \alpha_1^{n,m}\big)_{i+1} - \frac{1}{2}\big(\alpha_1^{n,0} + \alpha_1^{n,m}\big)_i\Big] - \overline{K}\Big[\frac{1}{2}\big(u^{n,0} + u^{n,m}\big)_{i+1} - \frac{1}{2}\big(u^{n,0} + u^{n,m}\big)_i\Big] \tag{7}$$

where the overlined values are of the same form as $\overline{V}_Q$ and $m = 0, 1$ for $P = 2$.

We point out, that in this specific work we present the results given by the choice of using arithmetic averages, but also Roe averages can be applied successfully.

Moreover, despite the focus of this work on 1D elements, the extension to multidimensional elements is straightforward: none of the formula (3), (4), (5) and (6) depend on the element topology. Once a globally continuous approximation of the

data on the mesh is obtained, it is possible to compute the surface flux (6) and the non-conservative terms (7) since it depends only on the amount of quadrature points and dot products. This extension is currently being done.

4 Numerical Validation

The proposed benchmark consists in an epoxy-spinel shock tube problem taken from literature [7, 9, 10, 15, 16] and is chosen due to the severe pressure jump across the discontinuity located at $x = 0.6\,\text{m}$ on a total domain of length $L = 1\,\text{m}$. The equation of state to close system (1) has been chosen to be for a stiffened gas $P(\rho_k, e_k) = (\gamma_k - 1)e_k - \gamma_k P_{\infty,k}$ for each phase k, with $P_{\infty,k}$ the reference pressure.

On the left side of the domain the pressure measures $P_L = 2 \cdot 10^{11}$ Pa, while on the right $P_R = 1 \cdot 10^5$ Pa. Initially, $u = 0\ [\frac{m}{s}]$ and the results are displayed at $t = 29\mu\text{s}$. The CFL is set to 0.45. The epoxy has a volume fraction $\alpha_1 = 0.5954$, a density $\rho_1 = 1185\ [\frac{kg}{m^3}]$, the heat capacity ratio $\gamma_1 = 2.43$ and the reference pressure $P_{\infty,1} = 5.3 \cdot 10^9$ Pa. The spinel has $\alpha_2 = 0.4046$, $\rho_2 = 3622\ [\frac{kg}{m^3}]$, $\gamma_2 = 1.62$ and

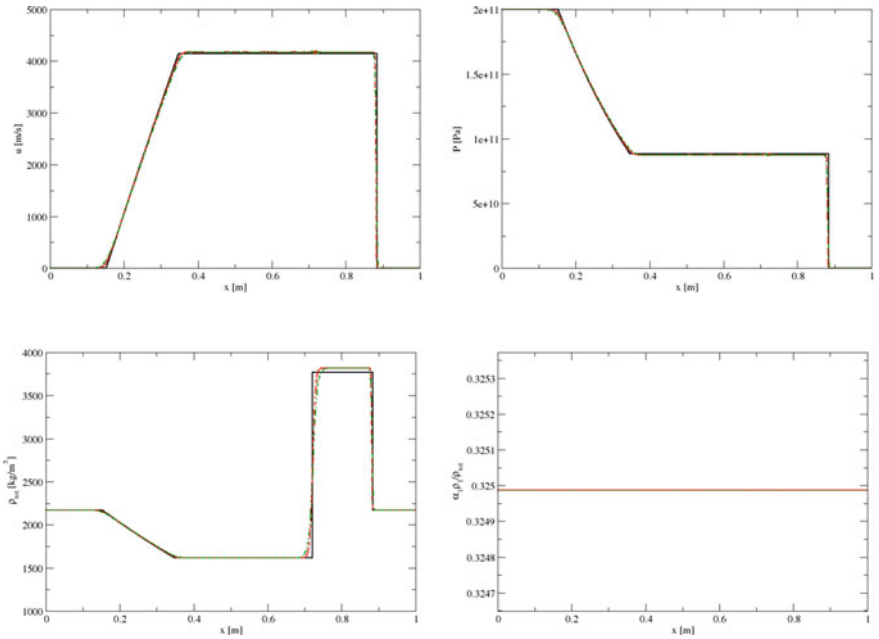


Fig. 1 Comparison between the exact solution (—), explicit residual distribution scheme for $\Delta x = 0.0001\,\text{m}$ (—) and the explicit residual distribution scheme for $\Delta x = 0.002\,\text{m}$ (- · ·-) with $\Delta x = 0.002\,\text{m}$. The velocity u, the total density ρ_{tot}, the pressure $P = P_1 = P_2$ and the mass fraction $\frac{\alpha_1 \rho_1}{\rho_{tot}}$ are represented along the shock tube

$P_{\infty,2} = 141 \cdot 10^9$ Pa. In Fig. 1, the fully explicit residual distribution scheme is tested on a coarse mesh with $\Delta x = 0.002$ m and on a finer mesh with $\Delta x = 0.0001$ m and the results are compared with the exact solution.

Both the coarse and fine mesh approximate very well the exact solution. It is possible to observe a slight overshoot of the total density. This is due to the non-conservative transport equation for the volume fraction, where the specific choice of the arithmetic averages, gives a slight numerical error. The existence of this small error does not result in relevant variations in the computation of the velocity and of the pressure, and it is possible to see that the mass fraction does also not suffer from the arithmetic averages induced numerical overshoot. A few oscillations in the contact discontinuity region can be observed, due to the specific choice of the limiter for this work. Comparing the coarser and finer mesh results, it is possible to appreciate the mesh convergence offered by the proposed scheme, even in case of very severe conditions.

5 Conclusions

In this work a fully explicit second-order scheme for the five equation model of Kapila et al. [4] has been presented. It is shown how a very easy-to-implement approximation can offer a second-order accuracy on a non-conservative multiphase compressible hyperbolic problem. The resulting scheme has been tested on a very severe benchmark, that has proven its robustness and its mesh convergence capability. The idea has been to choose a simple method, such as the well known Lax Friedrichs scheme, in order to show how easily an excellent approximation can be obtained. The extension of the presented fully explicit second-order scheme for compressible multiphase flows to 2D and to unstructured meshes results to be straightforward, and is currently being done.

Acknowledgements The SNF grant # 200021_153604/1 of the Swiss National Foundation has partly funded this work.

References

1. Abgrall, R.: How to prevent pressure oscillations in multicomponent flow calculations: a quasi conservative approach. J. Comput. Phys. **125**(1), 150–160 (1996)
2. Baer, M., Nunziato, J.: A two-phase mixture theory for the deflagration-to-detonation transition (ddt) in reactive granular materials. Int. J. Multiph. Flow **12**(6), 861–889 (1986)
3. Godlewski, E., Raviart, P.A.: Numerical approximation of hyperbolic systems of conservation laws, vol. 118. Springer Science & Business Media (2013)
4. Kapila, A., Menikoff, R., Bdzil, J., Son, S., Stewart, D.S.: Two-phase modeling of deflagration-to-detonation transition in granular materials: reduced equations. Phys. Fluids (1994-present) **13**(10), 3002–3024 (2001)

5. Karni, S.: Multicomponent flow calculations by a consistent primitive algorithm. J. Comput. Phys. **112**(1), 31–43 (1994)
6. LeMartelot, S., Nkonga, B., Saurel, R.: Liquid and liquidgas flows at all speeds. J. Comput. Phys. **255**, 53–82 (2013)
7. Marsh, S.: LASL Shock Hugoniot Data. Los Alamos Scientific Laboratory Series on Dynamic Material Properties, vol. 5. University of California Press (1980)
8. Murrone, A., Guillard, H.: A five equation reduced model for compressible two phase flow problems. J. Comput. Phys. **202**(2), 664–698 (2005)
9. Petitpas, F., Franquet, E., Saurel, R., Le Metayer, O.: A relaxation-projection method for compressible flows. part ii: artificial heat exchanges for multiphase shocks. J. Comput. Phys. **225**(2), 2214–2248 (2007)
10. Petitpas, F., Saurel, R., Franquet, E., Chinnayya, A.: Modelling detonation waves in condensed energetic materials: multiphase cj conditions and multidimensional computations. Shock Waves **19**(5), 377–401 (2009)
11. Ricchiuto, M., Abgrall, R.: Explicit runge-kutta residual distribution schemes for time dependent problems: second order case. J. Comput. Phys. **229**(16), 5653–5691 (2010)
12. Ricchiuto, M., Abgrall, R., Deconinck, H.: Application of conservative residual distribution schemes to the solution of the shallow water equations on unstructured meshes. J. Comput. Phys. **222**(1), 287–331 (2007)
13. Rodio, M.G., Abgrall, R.: An innovative phase transition modeling for reproducing cavitation through a five-equation model and theoretical generalization to six and seven-equation models. Int. J. Heat Mass Trans. **89**, 1386–1401 (2015)
14. Saurel, R., Abgrall, R.: A multiphase godunov method for compressible multifluid and multiphase flows. J. Comput. Phys. **150**(2), 425–467 (1999)
15. Saurel, R., Franquet, E., Daniel, E., Le Metayer, O.: A relaxation-projection method for compressible flows. part i: The numerical equation of state for the euler equations. J. Comput. Phys. **223**(2), 822–845 (2007)
16. Saurel, R., Le Métayer, O., Massoni, J., Gavrilyuk, S.: Shock jump relations for multiphase mixtures with stiff mechanical relaxation. Shock Waves **16**(3), 209–232 (2007)
17. Saurel, R., Petipas, F., Abgrall, R.: Modelling phase transition in metastable liquids: application to cavitating and flashing flows. J. Fluid Mech. **607**, 313–350 (2008)
18. Saurel, R., Petitpas, F., Berry, R.: Simple and efficient relaxation methods for interfaces separating compressible fluids, cavitating flows and shocks in multiphase mixtures. J. Comput. Phys. **228**(5), 1678–1712 (2009)
19. Struijs, R.: A multi-dimensional upwind discretization method for the euler equations on unstructured grids. Ph.D. thesis, TU Delft, Delft University of Technology (1994)
20. Xia, Y., Xu, Y., Shu, C.: Efficient time discretisation for local discontinuous galerkin methods. Discret. Contin. Dyn. Syst. Ser. B **8**(3), 677–693 (2007)

An Unstructured Forward-Backward Lagrangian Scheme for Transport Problems

Martin Campos Pinto

Abstract In a recent work Campos Pinto and Charles (2016), [3] From particle methods to forward-backward Lagrangian schemes, a novel method has been proposed and analyzed to reconstruct accurate backward transport flows, based on point markers pushed forward. When used in conjunction with a reliable particle code, this approach provides a simple tool to improve the accuracy of density approximations. In this article we report on an extension of the method to unstructured sets of markers, which are generic in particle codes. The resulting approximations have the same order of convergence, both in the a priori estimates and in the numerical simulations.

Keywords Particle methods · Semi-lagrangian methods · Transport equations · A priori error estimates · Remapped particle methods

MSC (2010): 76M28 · 35F10 · 65M12

1 Introduction

Consider a transport equation

$$\partial_t f(t,x) + u(t,x) \cdot \nabla f(t,x) = 0, \qquad t \in [0,T], \quad x \in \mathbb{R}^d \tag{1}$$

associated with an initial data $f^0 : \mathbb{R}^d \to \mathbb{R}$ and a velocity field $u : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$. If u is smooth, e.g. $L^\infty(0,T; W^{1;\infty}(\mathbb{R}^d))$ [13], we can define characteristic trajectories $X(t) = X(t;s,x)$ solutions to the ODEs $X'(t) = u(t, X(t))$, $X(s) = x$ on $[0,T]$, for all $x \in \mathbb{R}^d$ and $s \in [0,T]$. The corresponding flow $F_{s,t} : x \mapsto X(t)$

M. Campos Pinto (✉)
CNRS, Sorbonne Universités, UPMC Univ Paris 06, UMR 7598,
Laboratoire Jacques-Louis Lions, 4, Place Jussieu, 75005 Paris, France
e-mail: campos@ljll.math.upmc.fr

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_29

is then invertible and satisfies $(F_{s,t})^{-1} = F_{t,s}$. In particular, the transported density reads

$$f(t,x) = f^0((F_{0,t})^{-1}(x)) \qquad \text{for } t \in [0,T],\ x \in \mathbb{R}^d. \tag{2}$$

In general u depends on f through some self-consistent coupling. Following [3] we assume that we are given an accurate particle solver that (i) pushes forward arbitrary sets of markers along the characteristic flow, and (ii) computes reliable approximations to the velocity field u, at least for moderate simulation times $T' < T$. Using this solver as a black box we may then consider that u is given and that we can apply the exact forward flow $F_{\text{ex}}^{n,n+1}$ on small time steps of size Δt, on a given set of point markers.

The method that we describe below follows a series of works [1, 2, 5, 8, 10] where accurate approximations of transported densities are obtained through enhanced representations of the transport flow. In its most recent version [3], it implements the fundamental idea that local descriptions of the characteristic flow can be computed using a rather inexpensive method and then exploited to accurately reconstruct the transported density [6]. In practice the method studied in [3] combines key tools from the usual forward and backward lagrangian methods. It consists of

- pushing forward given markers along the characteristic trajectories, like in a standard particle method, and
- representing the density on a grid at given time steps, like in a backward semi-lagrangian method.

The crux of the method is then to use the markers pushed forward to approximate the backward flow between two time steps. The approximated density is then transported as in a standard Backward Semi-Lagrangian (BSL) method [14]. As explained in [2, 3], the strength of this approach over a standard particle method with smooth remappings (interpolations) is a lower diffusivity and higher convergence rate, and compared to the BSL method it has the advantage of avoiding a backward time integration of the trajectories. Owing to its hybrid nature we call it a *Forward-Backward Lagrangian* method.

2 The Forward-Backward Lagrangian Method

2.1 Backward Flow Reconstruction

The method relies on local approximations of the backward flow that are valid close to the marker positions x_k^n. For simplicity we restrict our presentation to first order flow approximations. Following [2] we define

$$B_{h,k}^{0,n} : x \mapsto x_k^0 + D_k^n (x - x_k^n) \tag{3}$$

with D_k^n a $d \times d$ matrix approximating the Jacobian $J_{B_{\text{ex}}^{0,n}}(x_k^n)$ of the backward flow, that can be computed from the position of the neighboring markers.

To compute a global approximation to the backward flow we subdivide the computational domain Ω in cartesian cells of size h, with centers denoted $\xi_i = ih$, $i \in \mathbb{Z}^d$, to avoid a confusion with the particle positions. To any ξ_i we associate a nearby marker, e.g. the closest one,

$$k^*(n,i) := \text{argmin}_{k\in\mathbb{Z}^d} \|x_k^n - \xi_i\|_\infty$$

and its corresponding backward flow (3). The global approximation to $B_{\text{ex}}^{0,n}$ is then obtained by smoothly patching these local approximations. Given a partition of unity $\sum_{i\in\mathbb{Z}^d} S(x-i) = 1$ involving a compactly supported, non-negative shape function S (e.g., a B-spline), we set

$$B_h^{0,n}(x) := \sum_{i\in\mathbb{Z}^d} B_{h,k^*(n,i)}^{0,n}(x) S\Big(\frac{x-\xi_i}{h}\Big). \tag{4}$$

2.2 Remapped FBL Method

Used in conjunction with a standard particle code, the above technique can be used in several ways to derive numerical schemes that improve the accuracy of the particle approximations.

For instance, the density can be approximated at any time step n by using the approximated flow (4) in the Lagrangian formula

$$f_h^{n,\text{fbl}}(x) := f^0\big(B_h^{0,n}(x)\big), \quad x \in \mathbb{R}^d. \tag{5}$$

Assuming that the underlying particle code pushes the markers along accurate trajectories, this reconstruction will be accurate as long as the associated characteristic flow remains smooth.

In many cases however, the regularity of the flow deteriorates over time and so does the accuracy of its approximations. To reduce this effect a simple method then consists of restarting the transport problem from time to time, namely before the approximated flow becomes too inaccurate. In the literature these restart time steps are often called *remappings*, and they essentially consist of re-initializing both the approximated density and the flow markers. After a restart indeed we must solve a new transport problem, where the characteristic flow has been reset to the identity mapping of $\mathbb{R}^d$. This comes at a price, which is the approximation error on the transported density. Formally the method reads as follows.

1. The two ingredients of the method are initialized: the positions of the markers x_k^0, $k \in \mathbb{Z}^d$, are computed, and the initial density f^0 is approximated on some grid of size h. This grid is a priori independent of the markers, and many methods can

be used. B-spline interpolations or quasi-interpolations are simple and efficient, see e.g. [2, 15]. We denote the corresponding approximation by $f_h^0 = A_h f^0$.

2. Letting $m_0 = 0 < m_1 < m_2 < \cdots$, denote the initial and subsequent remapping steps, on the r-th remapping cycle, $r = 0, 1, \cdots$, we do:
   a. For $n = m_r, \cdots, m_{r+1} - 1$, push all the markers $x_k^{n+1} = F^n(x_k^n)$, $k \in \mathbb{Z}^d$.
   b. Define the FBL approximation $f_h^{m_{r+1},\mathrm{fbl}} := f^{m_r} \circ B_h^{m_r,m_{r+1}}$ to $f(m_{r+1}\Delta t)$.
   c. Compute a new approximated density $f_h^{m_{r+1}} := A_h f_h^{m_{r+1},\mathrm{fbl}}$ for the next cycle.
   d. Re-initialize the markers to prepare the local flow approximations (3) between the present remapping time $m_{r+1}\Delta t$ and the future times $n\Delta t$.

To determine the method it thus remains to specify how the matrices D_k^n involved in the local flows (3) are computed from the markers positions, and how the latter must be initialized (and re-initialized) so that these matrices approximate well the Jacobian matrices of the backward flow.

3 Flow Reconstructions with Structured or Unstructured Markers

3.1 A Method Using Structured Markers

In the structured version proposed and studied in [3] following [2], the flow markers are initialized on a cartesian grid,

$$x_k^0 = hk, \qquad k \in \mathbb{Z}^d. \tag{6}$$

After pushing forward these markers over n time steps, one computes the deformation matrix D_k^n approximating the Jacobian matrix of the backward flow at the particle position x_k^n, namely

$$J_{B_{\mathrm{ex}}^{0,n}}(x_k^n) = \left(\partial_j (B_{\mathrm{ex}}^{0,n})_i (x_k^n)\right)_{1\le i,j\le d},$$

as follows. First one approximates the derivatives of the forward flow $F_{\mathrm{ex}}^{0,n}$ with finite differences involving the current particle positions $x_k^n = F_{\mathrm{ex}}^{0,n}(x_k^0)$. With a centered formula we define

$$J_k^n := \left(\frac{(x_{k+e_j}^n - x_{k-e_j}^n)_i}{2h} \right)_{1\le i,j\le d} \approx J_{F_{\mathrm{ex}}^{0,n}}(x_k^0) \tag{7}$$

and using the relation $J_{B_{\text{ex}}^{0,n}}(x_k^n) J_{F_{\text{ex}}^{0,n}}(x_k^0) = I_d$ which follows by differentiating the identity $x = B_{\text{ex}}^{0,n}(F_{\text{ex}}^{0,n}(x))$ at x_k^0, we approximate $J_{B_{\text{ex}}^{0,n}}(x_k^n)$ with

$$D_k^n := (J_k^n)^{-1}. \tag{8}$$

Re-initializing the markers on some time step n then simply consists by using a new set of markers located on the cartesian grid (6). We note that by doing this one forgets the previous marker positions x_k^n.

3.2 A New Method Using Unstructured Markers

Our extension of the above centered finite difference formulas to unstructured set of markers relies on the following notion of admissible simplices and parallelotopes.

Definition 1 Let $\alpha > 0$. An ordered simplex $(x_0, \ldots, x_d)$ of $\mathbb{R}^d$ is called *admissible* if the unit vectors

$$e_i = \frac{c - x_i}{\|c - x_i\|_2}, \quad i = 0, \ldots, d-1, \tag{9}$$

defined with $c = \frac{1}{2}(x_d + x_0)$ the center of the last edge, form a matrix satisfying

$$|\det(e_0, \ldots, e_{d-1})| \geq \alpha. \tag{10}$$

The associated *admissible parallelotope* is obtained by adding the vertices

$$x_{i+d} = x_0 + x_d - x_i, \quad i = 1, \ldots, d-1, \tag{11}$$

so that (x_0, x_i, x_d, x_{i+d}) forms a parallelogram with center c.

On every remapping step (including the initial step), the unstructured markers are then *prepared* in two steps. Below we consider a fixed value for $\alpha < 1$.

1. Inside every cell C_j of some cartesian mesh of resolution $\mathcal{O}(h)$, determine whether there exists an admissible simplex of markers $(x_0, \ldots, x_d)$. If not, insert new markers to form one (e.g. with a random algorithm).
2. Add d auxiliary markers corresponding to the $d-1$ remaining vertices (11) and center $x_{2d} = \frac{1}{2}(x_0 + x_d)$ of the associated parallelotope $X_j = (x_0, \ldots, x_{2d})$.

The backward flow close to a marker x_k^n is then approximated as follows. Denoting by $(\bar{x}_0, \ldots, \bar{x}_{2d})$ the position at time t^n of the pointers $X_j = (x_0, \ldots, x_{2d})$ corresponding to the cell C_j containing x_k^0, we first approximate the derivative of the forward flow $F^{0,n}$ along each unit vector $e_i = (x_{i+d} - x_i)/\|x_{i+d} - x_i\|_2$, see (9), by

$$\delta_i := \frac{\bar{x}_{i+d} - \bar{x}_i}{\|x_{i+d} - x_i\|_2} = \frac{F^{0,n}(x_{i+d}) - F^{0,n}(x_i)}{\|x_{i+d} - x_i\|_2}, \qquad i = 0, \ldots, d-1. \tag{12}$$

This is a second order formula for $J_{F_{\text{ex}}^{0,n}}(x_{2d})e_i$, where x_{2d} is the center of the parallelotope. The forward Jacobian matrix at x_{2d} is then approximated by

$$J_k^n := \big(\delta_0, \ldots, \delta_{d-1}\big) E^{-1}$$

where E is the matrix $\big(e_0, \ldots, e_{d-1}\big)$, and the backward Jacobian matrix by

$$D_k^n := (J_k^n)^{-1} = E\big(\delta_0, \ldots, \delta_{d-1}\big)^{-1}.$$

We observe that this matrix is actually determined by the parallelotope X_j and hence only depends on j. Moreover, since the center of the parallelotope is also a marker $x_{k(j)}^0 \in C_j$ that has been pushed forward, $D_k^n = D_{k(j)}^n$ approximates the backward Jacobian matrix at $x_{k(j)}^n$. The local backward flow (3) associated to any x_k^n that was initially (i.e. at the last remapping step) in the cell C_j is then be defined as

$$B_{h,k}^{0,n} = B_{h,k(j)}^{0,n} : x \mapsto x_{k(j)}^0 + D_{k(j)}^n (x - x_{k(j)}^n). \tag{13}$$

From the property (10) satisfied by the admissible parallelotope we can derive a priori estimates similar to the ones of the structured case [3].

Lemma 1 *Let x_k^n be a marker initially located in a cell C_j associated with an admissible parallelotope X_j with center $x_{k(j)}^0$. Then the approximated forward Jacobian defined above satisfies the a priori estimate*

$$\| J_k^n - J_{F_{\text{ex}}^{0,n}}\big(x_{k(j)}^0\big) \|_\infty \le C h^q |F_{\text{ex}}^{0,n}|_{q+1}, \quad q \in \{1, 2\} \tag{14}$$

with a constant depending only on d and α. In addition, there exists $h^* > 0$ such that for all $h < h^*$, J_k^n is invertible and the following a priori estimate holds

$$\| D_k^n - J_{B_{\text{ex}}^{0,n}}\big(x_{k(j)}^n\big) \|_\infty \le C \min_{q \in \{1,2\}} \big(h^q |F_{\text{ex}}^{0,n}|_{q+1}\big) |F_{\text{ex}}^{0,n}|_1^{2(d-1)}. \tag{15}$$

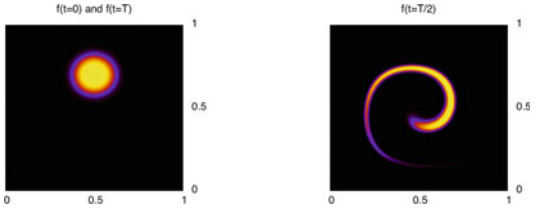


Fig. 1 Initial, intermediate and final profiles of the solution to the reversible test case in the text

4 Numerical Results

The efficiency of the structured method has been assessed on several transport problems in [3], see also [4] for a smooth particle approximation to Vlasov-Poisson plasmas based on similar flow reconstructions. To validate the unstructured version we use a passive transport problem of [11] which involves a swirling velocity field $u(t,x) := \cos\left(\frac{\pi t}{T}\right) \operatorname{curl} \phi(x)$ with $\phi(x) := -\frac{1}{\pi}\sin^2(\pi x_1)\sin^2(\pi x_2)$ and $T = 5$. The time symmetry yields a reversible problem: at $t = T/2$ the solutions reach a maximum stretching, and they revert to their initial value at $t = T$. For the initial data we consider a smooth hump centered on $\bar{x} = (0.5, 0.7)$ with approximate radius 0.2, $f^0(x) := \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{1}{3}(11 - 100\|x - \bar{x}\|_2)\right)\right)$. Figure 1 shows the profile of accurate solutions at initial, half and final times. L^2 convergence curves achieved by several particle methods are then plotted in Fig. 2. Here the particle pusher is a RK4 scheme with time step $\Delta t = T/100 = 0.05$ that has been taken small enough to have no significant effect on the final accuracy, and all the remappings are performed with cubic splines. As a reference, the top left panel shows results obtained with a standard Forward Semi-Lagrangian (FSL) scheme (i.e., a smooth particle method with periodic remappings), see e.g. [7, 9, 12]. The curves in the top right panel are obtained with the structured FBL version and confirm (i) the improved accuracy of this approach, and (ii) the need for much less remappings. In the two bottom panels we then show results corresponding to the unstructured FBL method which seem to exhibit similar accuracies. On the left the approximate flows are computed as

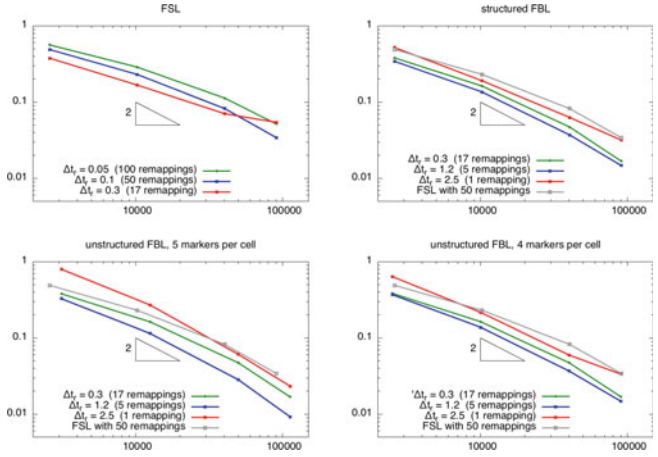


Fig. 2 L^2 convergence curves (errors at $t = T$ vs. number of particles) for the test case described in the text, using several particle methods with varying remappings periods Δt_r. Results obtained with a standard FSL scheme are shown for comparison on the *top left panel* and as a *gray curve* in the other panels. The other curves are obtained with a structured FBL method and an unstructured one using admissible simplices with parameter $\alpha = 0.5$. The *lower panels* show that similar results are obtained with or without using the center marker, see the text for details

described above, using 5 markers per cell corresponding to the vertices and centers of the admissible parallelograms. On the right a small variant is tested, where the center marker is discarded and replaced by the current marker x_k^n in the backward flow (13). This represents an error of order h on the initial and current positions of the center marker, but if the flow is $W^{2,\infty}$ it also corresponds to an error of order h on the Jacobian matrix, hence the resulting approximated flow is again of order h^2 which is confirmed by the numerical convergence rate.

5 Conclusion and Persectives

An unstructured version of a recent Forward-Backward particle method has been described an validated, showing second order accuracy on passive transport problems. These results represent an encouraging step towards the implementation of such reconstruction methods within standard particle codes. Further comparisons with standard advection methods should be performed to investigate the merits of this approach, and extensions to nonlinear transport problems should be adressed, including problems with discontinuous flows.

References

1. Alard, C., Colombi, S.: A cloudy Vlasov solution. Mon. Not. R. Astron. Soc. **359**(1), 123–163 (2005)
2. Campos Pinto, M.: Towards smooth particle methods without smoothing. J. Sci. Comput. (2014)
3. Campos Pinto, M., Charles, F.: From particle methods to forward-backward Lagrangian schemes. hal.archives-ouvertes.fr (2016). ⟨hal-01385676⟩
4. Campos Pinto, M., Sonnendrücker, E., Friedman, A., Grote, D., Lund, S.: Noiseless Vlasov–Poisson simulations with linearly transformed particles. J. Comput. Phys. **275**(C), 236–256 (2014)
5. Cohen, A., Perthame, B.: Optimal approximations of transport equations by particle and pseudoparticle methods. SIAM J. Math. Anal. **32**(3), 616–636 (2000)
6. Colombi, S., Alard, C.: A "metric" semi-Lagrangian Vlasov-Poisson solver (2016). Submitted
7. Cotter, C., Frank, J., Reich, S.: The remapped particle-mesh semi-Lagrangian advection scheme. Q. J. R. Meteorol. Soc. **133**(622), 251–260 (2007)
8. Cottet, G.H., Koumoutsakos, P., Salihi, M.: Vortex Methods with Spatially Varying Cores. J. Comput. Phys. **162**(1), 164–185 (2000)
9. Crouseilles, N., Respaud, T., Sonnendrücker, E.: A forward semi-Lagrangian method for the numerical solution of the Vlasov equation. Comput. Phys. Commun. **180**(10), 1730–1745 (2009)
10. Hou, T.: Convergence of a Variable Blob Vortex Method for the Euler and Navier-Stokes Equations. SIAM J. Numer. Anal. **27**(6), 1387–1404 (1990)
11. LeVeque, R.: High-resolution conservative algorithms for advection in incompressible flow. SIAM J. Numer. Anal. 627–665 (1996)
12. Nair, R., Scroggs, J., Semazzi, F.: A forward-trajectory global semi-Lagrangian transport scheme. J. Comput. Phys. **190**(1), 275–294 (2003)

13. Raviart, P.A.: An analysis of particle methods. In: Numerical methods in fluid dynamics (Como, 1983), Lecture Notes in Mathematics, pp. 243–324. Berlin (1985)
14. Sonnendrücker, E., Roche, J., Bertrand, P., Ghizzo, A.: The semi-Lagrangian method for the numerical resolution of the Vlasov equation. J. Comput. Phys. **149**(2), 201–220 (1999)
15. Unser, M., Daubechies, I.: On the approximation power of convolution-based least squares versus interpolation. IEEE Trans. Signal Process. **45**(7), 1697–1711 (1997)

A Godunov-Type Scheme for Shallow Water Equations Dedicated to Simulations of Overland Flows on Stepped Slopes

Nicole Goutal, Minh-Hoang Le and Philippe Ung

Abstract We introduce a new Godunov-type finite volume scheme for the Shallow Water equations based on a three-waves Approximate Riemann Solver. By linearizing the Bernoulli and consistency equations, the resulting scheme is positive, well-balanced and permits to improve the accuracy of numerical results compared with other methods. The proposed scheme is particularly suitable for simulations of overland flows on stepped slopes.

Keywords Shallow-water equations · Finite volume schemes · Source term approximations · Well–balanced schemes

MSC2010: 65M12 · 76M12 · 35L65

1 Introduction

In this work, we look for a numerical scheme for the well-known Shallow Water equations (SWE) given by:

$$\begin{cases} \partial_t h + \partial_x (hu) = 0, \\ \partial_t (hu) + \partial_x (hu^2 + gh^2/2) = -gh\partial_x b, \end{cases} \tag{1}$$

M.-H. Le (✉) · P. Ung
Laboratoire d'Hydraulique Saint Venant (LHSV), 6 quai Watier, 78401 Chatou, France
e-mail: minh-hoang.le@enpc.fr

P. Ung
e-mail: philippe.ung@enpc.fr

N. Goutal
LHSV and EDF R& D, 6 quai Watier, 78401 Chatou, France
e-mail: nicole.goutal@edf.fr

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_30

where g stands for the gravity constant, $w := (h, hu)^T$ is the conservative variable of the system; the water height $h \in \mathbb{R}^+$ and the velocity u depend on time t and space x. The spatial function $b := b(x)$ refers to a fixed topography. We recall hereafter some important properties which are useful when dealing with Approximate Riemann Solver (ARS).

Adding the trivial equation $\partial_t b = 0$ to (1) permits to view the SWE as a nonlinear hyperbolic system of variable $v := (h, hu, b)^T$ in nonconservative form $\partial_t v + A(v)\partial_x v = 0$ where the nonconservative product $A(v)\partial_x v$ can be defined as a Borel measure. The Jacobian matrix $A(v)$ admits three real eigenvalues, i.e. characteristic velocities

$$\lambda_1 = u - \sqrt{gh}, \quad \lambda_2 = u + \sqrt{gh}, \quad \lambda_3 = 0.$$

It can be checked that the first two characteristic fields associated with $\lambda_{1,2}$ are genuinely nonlinear while the last one related to λ_3 is linearly degenerate. As discussed in [11], the topography b remains constant along the wave curves $\mathscr{W}_{1,2}(w_0, b_0)$ related to $\lambda_{1,2}$ and constituted of all states (w, b) that can be connected to a given state (w_0, b_0) by a rarefaction wave or an admissible shock wave. The topography may change only across a stationary contact wave and satisfies the jump relations, also known as Bernoulli equations

$$[hu] = 0, \quad [u^2/2g + h + b] = 0. \tag{2}$$

where $[\bullet]$ denotes the jump operator. It is worth noticing that these stationary contact waves are nothing but the steady states of the SWE. Since the Eqs. (2) admit at least two solutions when the step $b - b_0$ is sufficiently small, one can use the *Monotonicity Criterion*, detailed later, in order to select the relevant one.

Derivation of positive and well-balanced schemes, i.e. those preserving the states verifying (2), was a very active research since the pioneer works of Bermudez-Vasquez [4], Greenberg-LeRoux [9]. Most of the existing methods only look for satisfying the trivial *lake at rest steady states* ($u = 0$)—a criteria which is nowadays a prerequisite for new approaches. Most of these methods might have difficulty to provide accurate results for *moving steady states* ($u \neq 0$). In this work, we look for designing a simple numerical scheme able to exactly restore the lake at rest steady state and providing more accurate results for moving steady states compared with existing methods, in particular the well-known hydrostatic reconstruction [2].

This paper is organized as follows. We recall first the definition of Godunov-type scheme and three-waves Approximate Riemann Solver. The expression of the intermediate states is presented in the next section following by an analysis on well-balancing property and entropic one of the scheme. Finally, several numerical benchmarks are performed to assert the interest of the method.

2 Godunov-Type Scheme Based on a Three-Wave ARS

We briefly recall the definition of a first-order Godunov-type finite volume scheme.

We introduce a space step Δx and a time step Δt both assumed to be constant for simplicity. The computational domain is discretized by a sequence of point $x_{j+1/2} := j\Delta x$ for $j \in \mathbb{Z}$. We define a piecewise constant approximation (w_j^0, b_j) as the initial condition and the topography on control cell $C_j := [x_{j-1/2}, x_{j+1/2})$.

Assuming that a piecewise constant approximation of the solution w_j^n at time t^n is known, the Godunov-type scheme computes the solution at the next time level $t^{n+1} := t^n + \Delta t$ in two steps. First, we build an approximate solution $w_{\mathscr{R}}(x, t)$ of the Riemann problem associated with (1) at each interface $x_{j+1/2}$ with the initial data $\{(w_j^n, b_j)\}_{j\in\mathbb{Z}}$. Next, a piecewise constant approximate solution at time t^{n+1} is obtained by averaging the solution $w_{\mathscr{R}}(x, t^{n+1})$ on each control cell C_j.

Given initial data $(w_L, w_R; b_L, b_R)$ of local Riemann problem, we adopt a simple approximate Riemann solver composed by three discontinuity waves propagating with velocities $\lambda_L < \lambda_3 = 0 < \lambda_R$ and two intermediate states w_L^* and w_R^* (Fig. 1).

Under the CFL condition $\Delta t \leq \frac{\Delta x}{2\max(-\lambda_L, \lambda_R)}$, the first order Godunov-type finite volume scheme described earlier applying to system (1) writes

$$w_j^{n+1} = w_j^n - \frac{\Delta t}{\Delta x}\left(f_{j+1/2}^L - f_{j-1/2}^R\right), \tag{3}$$

where the left- and right- numerical fluxes $f_{j+1/2}^{L,R} := f^{L,R}(w_j^n, w_{j+1}^n, b_j, b_{j+1})$ are defined by

$$\begin{aligned} f^L(w_L, w_R, b_L, b_R) &:= f(w_L) + \lambda_L(w_L^* - w_L), \\ f^R(w_L, w_R, b_L, b_R) &:= f(w_R) + \lambda_R(w_R^* - w_R), \end{aligned} \tag{4}$$

with $f(w) := (hu, hu^2 + gh^2/2)^T$ being the physical flux of the SWE. Hence, the construction of such a scheme consists in determinating the intermediate states $w_{L,R}^*$ which is subject of the next section.

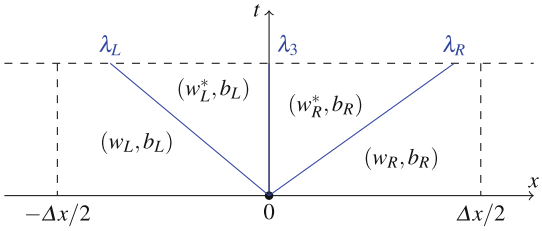


Fig. 1 A three-waves approximate Riemann problem

3 Expression of the Intermediate States

From [10], the approximate Riemann solver must be consistent with the exact solution in the integral sense. This condition provides that the intermediate states $w^*_{L,R}$ satisfy the following consistency relations

$$h_R u_R - h_L u_L = \lambda_L(h^*_L - h_L) + \lambda_R(h_R - h^*_R), \tag{5}$$

$$\left(h_R u_R^2 + \frac{g h_R^2}{2}\right) - \left(h_L u_L^2 + \frac{g h_L^2}{2}\right) + g\Delta x\{h\partial_x b\} = \lambda_L(h^*_L u^*_L - h_L u_L) + \lambda_R(h_R u_R - h^*_R u^*_R). \tag{6}$$

The approximation of the source term $\{h\partial_x b\}$ will be related to the well-balanced property and is precised later. Therefore, two additional relations are missing in order to close this system; these two ones are obtained with help of the well-balancing property characterized by (2).

As previously mentioned, the issue related to the choice of the solution of (2) is the keypoint of the method. Let recall that contact wave curve $\mathscr{W}_3(w_0, b_0)$ can be parametrized with h such that

$$u := u(h) = h_0 u_0/h, \quad b := b(h) = b_0 + (u_0^2 - u^2)/2g + h_0 - h. \tag{7}$$

It clearly appears that $b(h)$ increases if $0 < h < h_c := (h_0 u_0/\sqrt{g})^{2/3}$ and decreases elsewhere. For a given state (w_0, b_0) and a given $b < b_{max} := b(h_c)$, there exist two water heights $h_1 < h_c < h_2$ satisfying the jump relations (2) and corresponding to super- and sub-critical regime respectively. Therefore, the Riemann problem may admit more than one family of solutions, even if system (1) is strictly hyperbolic. Imposing an additional admissible condition called *Monotonicity Criterion* permits to select a unique solution [11]: *Along a wave curve* $\mathscr{W}_3(w_0, b_0)$, *the parametrized function* $b(h)$ *is monotone*. Since the state (w_0, b_0) itself belongs to $\mathscr{W}_3(w_0, b_0)$, we retain the solution h_1 (resp. h_2) if $h_0 < h_c$ (resp. $h_c < h_0$). In that sense, the well-balancing property can be reduce to impose $(w^*_R, b_R) \in \mathscr{W}_3(w^*_L, b_L)$ meaning that

$$h^*_L u^*_L = h^*_R u^*_R := q^*, \tag{8}$$

$$\varphi(h^*_L, h^*_R, q^*) := \frac{(q^*)^2}{2g}\left(\frac{1}{(h^*_R)^2} - \frac{1}{(h^*_L)^2}\right) + h^*_R - h^*_L = -\Delta b, \tag{9}$$

with $\Delta b := b_R - b_L$. With help of (8) and the definition of the intermediate state (h^{HLL}, q^{HLL}) associated with the HLL solver [10]

$$(\lambda_R - \lambda_L)h^{HLL} = \lambda_R h_R - \lambda_L h_L - h_R u_R + h_L u_L,$$

$$(\lambda_R - \lambda_L)q^{HLL} = \lambda_R h_R u_R - \lambda_L h_L u_L - \left(h_R u_R^2 + \frac{g h_R^2}{2}\right) + \left(h_L u_L^2 + \frac{g h_L^2}{2}\right),$$

the consistency Eqs. (5)–(6) can be written such as

$$\alpha h_R^* + (1-\alpha) h_L^* = h^{HLL}, \quad \alpha := \frac{\lambda_R}{\lambda_R - \lambda_L} > 0, \tag{10}$$

$$q^* = q^{HLL} - \frac{g\Delta x\{h\partial_x b\}}{\lambda_R - \lambda_L}. \tag{11}$$

Therefore, a given approximation $\{h\partial_x b\}$ permits to define the discharge q^*. Moreover, h^{HLL} is a convex combination of $h_{L,R}^*$, so $h^{HLL} \in [h_L^*, h_R^*]$. Next, we can consider φ as a function of one variable h_L^* by injecting (10) and (11) into (9). The Monotonicity criterion leads to $h_{L,R}^* \leq h_c^* := (q^*/\sqrt{g})^{2/3}$ (resp. $h_c^* \leq h_{L,R}^*$) if $h^{HLL} \leq h_c^*$ (resp. $h_c^* \leq h^{HLL}$). Accordingly, φ increases if $h^{HLL} \leq h_c^*$ and decreases elsewhere. Equation (9) admits a unique solution h_L^* if $-\Delta b \leq \varphi_{max} := \varphi(h_c^*)$.

It should be checked that we can always consider $-\lambda_L/\lambda_R >> 1$ or $-\lambda_R/\lambda_L >> 1$ [5] in order to ensure the existence condition $-\Delta b \leq \varphi_{max}$, so Eq. (9) has solution. In practice for a given velocities $\lambda_{L,R}$, we impose $h_L^* = h_c^*$ in the case where the existence condition fails. Furthermore, the fact that $\varphi(h^{HLL}) = 0$ suggests to replace $\varphi(h_L^*)$ in Eq. (9) by its linearization in a neighbourhood of h^{HLL}, especially when Δb is small enough, to obtain a simpler equation

$$\beta(h_R^* - h_L^*) = -\Delta b, \quad \beta := 1 - (h_c^*/h^{HLL})^3 \leq 1. \tag{12}$$

We find that the parameter β characterizes the regime of intermediate states, i.e. $w_{L,R}^*$ characterize the sub-critical if $0 < \beta \leq 1$ and super-critical one in the opposite case. Note that the scheme proposed in [3], relying on trivial steady states, is nothing but a particular case of (12) with $\beta = 1$. Using (12) both with (10) yields to

$$h_L^* = h^{HLL} + \alpha\frac{\Delta b}{\beta}, \quad h_R^* = h^{HLL} - (1-\alpha)\frac{\Delta b}{\beta}. \tag{13}$$

Solution (13), which can be seen as linear functions of $\Delta b/\beta$, needs a simple post-treatment in order to ensure the positivity as well as the monotonicity criteria. For example, in sub-critical case ($\beta > 0$), we modify $h_L^* := \max(h_L^*, h_c^*)$ (resp. $h_R^* := \max(h_R^*, h_c^*)$) if $\Delta b < 0$ (resp. $\Delta b > 0$) and use the consistency condition (5) to find h_R^* (resp. h_L^*). Alternatively, the linearized solution (13) is no longer accurate if β is close to 0, i.e. near sonic point. In this case, we could rather solve (numerically) the nonlinear equation (9).

4 Well-Balancing and Entropy

Now, we aim to present a suitable approximation of the source term and study the well-balancing property of the scheme. Assume that the data (w_L, b_L) and (w_R, b_R) are steady states and $q := h_L u_L = h_R u_R$. The Bernoulli equation (2) and the con-

sistency condition (6) yield

$$-\Delta b = \frac{q^2}{2g}\left(\frac{1}{h_R^2} - \frac{1}{h_L^2}\right) + h_R - h_L \approx \left(1 - \frac{q^2}{g(h^{HLL})^3}\right)(h_R - h_L),$$
$$-\Delta x\{h\partial_x b\} = \frac{q^2}{g}\left(\frac{1}{h_R} - \frac{1}{h_L}\right) + \frac{h_R^2 - h_L^2}{2} \approx h^{HLL}\left(1 - \frac{q^2}{g(h^{HLL})^3}\right)(h_R - h_L),$$

where we have linearized the equations near h^{HLL}. This result suggests to use

$$\Delta x\{h\partial_x b\} := h^{HLL}\Delta b, \tag{14}$$

which is a consistency approximation since $\{h\partial_x b\} = h^{HLL}\Delta b/\Delta x \to h\partial_x b$ when $\Delta x \to 0$, $w_{L,R} \to w$ and $b_{L,R} \to b$.

Let us show how a combination of (12) and (14) can preserve at least the lake at rest steady states. We use a classical choice of $\lambda_{L,R}$ writing

$$\lambda_L := \min_{w=w_L,w_R}(-\varepsilon, u - \sqrt{gh}), \quad \lambda_R := \max_{w=w_L,w_R}(\varepsilon, u + \sqrt{gh}) \tag{15}$$

with a small parameter ε to ensure $\lambda_L < 0 < \lambda_R$.

Considering a lake at rest steady state $(w_L, w_R; b_L, b_R)$, we have $\lambda_L = -\lambda_R$ since $u_L = u_R = 0$. As $h_R - h_L = -\Delta b$, straightforward calculations from (11)–(13) show that $q^* = 0$, $h_L^* = h_L$, $h_R^* = h_R$; hence $w_L^* = w_L$, $w_R^* = w_R$. Furthermore, the scheme should be able to handle the steady states of the wet-dry (resp. dry-wet) transition, i.e. the case where $h_L + b_L \le b_R$, $h_R = 0$ (resp. $h_R + b_R \le b_L$, $h_L = 0$). It is sufficient to operate a slight modification of $\lambda_{L,R}$ from (15). Indeed, for wet-dry transition steady state, we compute λ_L by (15) and impose $\lambda_R := \lambda_L(1 - 2\Delta b/h_L) \ge -\lambda_L$. The same idea holds for the case of dry-wet transition by setting $\lambda_L := \lambda_R(1 + 2\Delta b/h_R) \le -\lambda_R$.

It is worth to note that we have modified the velocities $\lambda_{L,R}$ in order to deal with wet-dry or dry-wet transition steady states and are able to preserve the approximation (14). This is not the case for other schemes. In [3], the authors use $\Delta x\{h\partial_x b\} := \bar{h}\Delta b$ with $\bar{h} := (h_L + h_R)/2$ for fully-wet case and replace, for example, Δb by h_L for the case of wet-dry transition. This kind of modification can lead to some difficulties to provide accurate results in the case of overlands flow on stepped slope as we will see in the next section. In [5], a more complex approximation consists to define $\Delta x\{h\partial_x b\} := (h_L h_R/\bar{h})\Delta b + (h_L h_R/\bar{h} - \bar{h})(h_R - h_L)$ which permits to handle the moving steady states, i.e. $q \neq 0$. However, as underlined by the authors, this approximation is no longer consistent with zero when the topography is flat.

About the entropy property, we can show that the scheme satisfies a discrete entropy inequality, with an error term such as

$$\mathscr{F}(w_R) - \mathscr{F}(w_L) - \Delta x\, \sigma(\Delta x, w_L, w_R, b_L, b_R) + \Delta x\, \varepsilon(\Delta x)$$
$$\leq \lambda_L(\mathscr{U}(w_L^*) - \mathscr{U}(w_L)) + \lambda_R(\mathscr{U}(w_R) - \mathscr{U}(w_R^*))$$
$$\text{with} \quad \mathscr{U}(w) = \frac{hu^2}{2} + \frac{gh^2}{2} \quad \text{and} \quad \mathscr{F}(w) = \left(\frac{u^2}{2} + gh\right) hu.$$

which proves that the scheme should converge with mesh refinement if the topography is at least continous. In our case, we can express $\sigma = -gq^{HLL}\Delta b/\Delta x$ which ensures the consistency with $-ghu\partial_x b$ when $\Delta x \to 0$, $w_{L,R} \to w$ and $b_{L,R} \to b$.

5 Numerical Results

We present in the following several tests to illustrate the behaviours of the proposed scheme. The reference solutions are given by SWASHES [8].

5.1 Accelerating Super-Critical Flow over a Downward Slope

This test was introduced by [7] to illustrate a subtle difficulty of the hydrostatic reconstruction (HR) to provide accurate results [2]. A slight modification of HR can be found in [6] allowing to improve the result. The bed profile consists in an inclined plane for different values of slope $S := -\partial_x b$ ranging from $S = 16\%$ to $S = 22\%$, i.e. stepped slopes. We impose a super-critical inflow boundary condition, at $x = 0$, with $h = 0.02$, $q = 0.01$ and perform the simulations until a steady state is reached. We used a moderate resolution $\Delta x = 0.05$ in order to get $|\Delta b| > \max(h_L, h_R)$ far from the left boundary. On Fig. 2, the advantages of the present scheme are clearly visible.

5.2 Transcritical Flow with Shock over a Bump

This well-known test case is focused on the way the scheme handles smooth transcritical solution together with a stationary shock. The topography is given by $b(x) = 0.02 - 0.05(x - 10)^2$ if $8 < x < 12$ and is flat elsewhere. Starting with a sub-critical regime at inlet ($x = 0$) with $q = 0.18$, the flow becomes super-critical by passing the bump and comes back to sub-critical regime after the stationary shock to reach $h = 0.33$ at outlet ($x = 25$). Figure 3, in particular the result on the discharge (right), reveals that this scheme is a gain the most accurate compared to others ones.

Other tests for transient cases have shown that the proposed scheme is as effective when dealing with dry/wet boundaries transition, e.g. Thacker's solution, as it is for complicated solutions of the Riemann problem such as dambreak over a step [1].

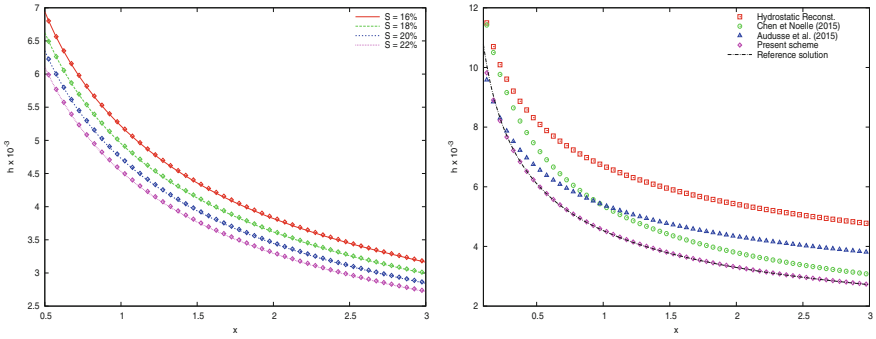


Fig. 2 Supercritical flow on inclined plane. *Left* reference solution (*line*) and numerical result (point) given by the present scheme with $\Delta x = 0.05$. *Right* numerical result with $S = 22\%$ given by different schemes compared with reference solution

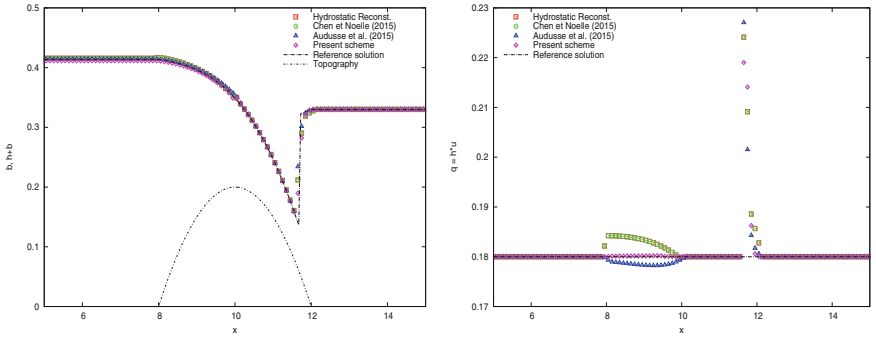


Fig. 3 Transtical flow with shock on a bump: water height (*left*) and unit discharge (*right*) given by different schemes, using $\Delta x = 0.1$, compared with reference solution

6 Conclusion

In this paper, we have proposed a Godunov-type finite volume scheme for the SWE based on three-waves ARS. By linearizing the Bernoulli and consistency equations, the resulting scheme is positive, well-balanced and permits to improve the accuracy of numerical results compared with other methods, in particular the well-known hydrostatic reconstruction.

References

1. Alcrudo, F., Benkhaldoun, F.: Exact solutions to the riemann problem of the shallow water equations with a bottom step. Comput. Fluids **30**(6), 643–671 (2001)
2. Audusse, E., Bouchut, F., Bristeau, M.O., Klein, R., Perthame, B.: A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water flows. SIAM J. Sci. Comput. **25**(6), 2050–2065 (2004)
3. Audusse, E., Chalons, C., Ung, P.: A simple well-balanced and positive numerical scheme for the shallow-water system. Commun. Math. Sci. **13**, 1317–1332 (2015)
4. Bermúdez, A., Vázquez, M.E.: Upwind methods for hyperbolic conservation laws with source terms. Comput. Fluids **23**(8), 1049–1071 (1994)
5. Berthon, C., Chalons, C.: A fully well-balanced, positive and entropy-satisfying godunov-type method for the shallow-water equations. Math. Comput. **85**, 1281–1307 (2016)
6. Chen, G., Noelle, S.: A new hydrostatic reconstruction scheme motivated by the wet-dry front **440** (2015). https://publications.rwth-aachen.de/record/565174
7. Delestre, O., Cordier, S., Darboux, F., James, F.: A limitation of the hydrostatic reconstruction technique for shallow water equations. C. R. Acad. Sci. Paris Ser. I **350**, 677–681 (2012)
8. Delestre, O., Lucas, C., Ksinant, P.A., Darboux, F., Laguerre, C., Vo, T.N.T., James, F., Cordier, S.: Swashes: a compilation of shallow water analytic solutions for hydraulic and environmental studies. Int. J. Numer. Methods Fluids **72**(3), 269–300 (2013)
9. Greenberg, J.M., LeRoux, A.Y.: A well-balanced scheme for the numerical processing of source terms in hyperbolic equations. SIAM J. Numer. Anal. **33**, 1–16 (1996)
10. Harten, A., Lax, P.D., van Leer, B.: On upstream differencing and Godunov-type schemes for hyperbolic conservation laws. SIAM Rev. **25**(1), 53–61 (1983)
11. LeFloch, P.G., Thanh, M.D.: The riemann problem for shallow water equations with discontinous topography. Commun. Math. Sci. **5**, 865–885 (2007)

Two Models for the Computation of Laminar Flames in Dust Clouds

Dionysios Grapsas, Raphaèle Herbin and Jean-Claude Latché

Abstract We address two models for the simulation of dust clouds premixed combustion: the first one consists in usual balance equations; to derive the second one, we suppose that the solution takes the form of a travelling combustion wave and track the location of the flame brush by a phase-field-like technique. We build a finite volume fractional step scheme for both models, which respects the natural physical bounds of the unknowns. Then we assess the consistency of both formulations.

Keywords Reactive flows · Low mach number flows · Finite volumes · Staggered discretizations

MSC (2010): 65N09 · 76M12

1 Introduction

We address in this paper two alternative models dedicated to the simulation of laminar flames in dust suspensions in a gaseous atmosphere, for which a one-dimensional representation, supposing a low Mach number flow, is sufficient. The combustible particules are supposed to be in mechanical and thermal equilibrium with the continuous phase (or, in other words, no drift nor temperature deviation between the gas and

D. Grapsas · R. Herbin
I2M UMR 7373, Aix-Marseille Université, CNRS, École Centrale de Marseille,
39 rue Joliot Curie, 13453 Marseille, France
e-mail: dyonisios.grapsas@univ-amu.fr

R. Herbin
e-mail: raphaele.herbin@univ-amu.fr

J.-C. Latché (✉)
Institut de Radioprotection et de Sûreté Nucléaire (IRSN),
13115 Saint-Paul-lez-Durance, France
e-mail: jean-claude.latche@irsn.fr

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_31

solid phases is taken into account). We consider two descriptions of the combustion phenomenon:

- the first one is obtained by collecting mass balance for the chemical species, the energy balance and the momentum balance for the mixture; the reaction term $\dot{\omega}$ is expressed by a closure law depending of the temperature, derived on the basis of physical arguments. This model will be refered to in the following as the *primitive formulation*.
- The second one relies on the assumption that the solution consists in a travelling reaction thin interface (the so-called flame front) separating a zone where the combustion is complete (the "burnt zone") from a zone where no combustion has yet occured (the "fresh zone"). This representation offers the possibility to reduce the problem to an explicit tracking of the front location, through the solution of a transport equation for a color function G ($G \in [0, 1]$, $G < 0.5$ in the burnt zone, $G \geq 0.5$ in the fresh atmosphere); the reaction term is governed by the value of G: $\dot{\omega} = 0$ if $G \geq 0.5$ and $\dot{\omega}$ is proportional to $1/\tau$ otherwise, where τ is a time-scale closely correlated to the flame front thickness. In the rest of this paper, we will call this model the *flame velocity formulation*.

The first option is standard for the computation of laminar flames. Variants of the flame velocity formulation are often chosen to compute turbulent deflagrations in industrial applications [4, 5], as in nuclear safety studies performed at the French Institut de Radioprotection et de Sûreté Nucléaire (IRSN). Indeed, this latter model seems easier to solve, in the sense that stable segregated algorithms may be designed for this purpose; in addition, the flame brush incorporates structures which are very small compared to the system scales, and the flame velocity approach allows an upscaling of this complex physical phenomenon through a single parameter (the turbulent flame velocity) which may be inferred from experimental data.

A finite volume fractional step numerical scheme was developed in [2] for the solution of the system of primitive equations; we shortly describe it here and review its main properties. The aim of this paper is then to assess the accuracy of the switch from the first model to the second one: first, we check that the solution to the primitive formulation, in conditions representative of the target physical reality, is indeed a flame front propagating through the medium; then, we compare such a solution with the one obtained with the corresponding flame velocity model.

2 The Primitive Formulation

2.1 The Governing Equations

The flow is supposed to be governed by the balance equations modelling a variable density flow in the asymptotic limit of low Mach number flows, namely the mass balance of the chemical species and of the mixture, the enthalpy balance, and the momen-

tum balance equations. For a one-dimensional flow in such a quasi-incompressible situation, the role played by the mass and momentum balance equations is quite different than in the multi-dimensional case: the velocity may be seen as the solution of the mass balance equation, and the momentum balance yields the dynamic pressure. Since this latter unknown does not appear in the other equations, its computation is of poor interest, and the momentum balance equation may be disregarded.

Except for this aspect, equations in this section are written in the usual multi-dimensional form. The computational domain is denoted by Ω, and its boundary $\partial\Omega$ is supposed to be split into an inflow part $\partial\Omega_I$ (where the flow enters the domain, *i.e.* $\boldsymbol{u}\cdot\boldsymbol{n}_{\partial\Omega} < 0$, with $\boldsymbol{u}$ the flow velocity and $\boldsymbol{n}_{\partial\Omega}$ the normal vector to $\partial\Omega$ outward Ω) and an outflow one $\partial\Omega_O$ (where the flow leaves the domain, *i.e.* $\boldsymbol{u}\cdot\boldsymbol{n}_{\partial\Omega} \geq 0$) of positive $(d-1)$-measure. The problem is posed over the time interval $(0, T)$.

Mass balance equations—The mass balance reads:

$$\partial_t \rho + \operatorname{div}(\rho \boldsymbol{u}) = 0, \tag{1}$$

where ρ stands for the fluid density. This equation must be complemented by an initial condition and a boundary condition on $\partial\Omega_I$ for the density; both functions are obtained by the data of the temperature and flow composition, thanks to the equation of state (see below).

Only four chemical species are supposed to be present in the flow, namely the dust, or fuel (denoted by F), the oxydant (O), the product (P) of the reaction, and a neutral gas (N). A one-step irreversible total chemical reaction is considered:

$$\nu_F F + \nu_O O + N \rightarrow \nu_P P + N,$$

where ν_F, ν_O and ν_P are the molar stoichiometric coefficients of the reaction. Chemical species other than the fuel are supposed to be gases. The system of the mass balance equations for the chemical species reads:

$$\partial_t(\rho y_i) + \operatorname{div}(\rho y_i \boldsymbol{u}) = \dot{\omega}_i, \qquad \text{for } 1 \leq i \leq N_s, \tag{2}$$

where y_i and $\dot{\omega}_i$ stand respectively for the mass fraction and the reaction rate of the species i. The number of species is denoted by N_s, with, by assumption, $N_s = 4$, and we indifferently use the notation $(y_i)_{1\leq i\leq N_s}$ or y_F, y_0, y_P and y_N for the fuel, oxydant, product and neutral gas mass fractions, respectively. To simplify the exposition, the mass diffusion fluxes have been supposed in the set (2) of equations to vanish. This system must be complemented by initial and Dirichlet boundary conditions for $(y_i)_{1\leq i\leq N_s}$ on the inflow part of the domain boundary $\partial\Omega_I$. The prescribed values of the mass fractions at the initial time and on the inflow boundary lie in the interval $[0, 1]$. The reaction rate of each chemical species may be written as:

$$\dot{\omega}_F = -\nu_F W_F \dot{\omega}, \quad \dot{\omega}_O = -\nu_O W_0 \dot{\omega}, \quad \dot{\omega}_P = \nu_P W_P \dot{\omega} \quad \text{and } \dot{\omega}_N = 0,$$

where W_F, W_0 and W_P stand for the molar masses of the fuel, oxydant and product respectively, and $\dot{\omega}$ is a non-negative reaction rate, which is supposed to vanish when either $y_F = 0$ or $y_O = 0$. Since $\nu_F W_F + \nu_O W_0 = \nu_P W_P$, we have $\dot{\omega}_F + \dot{\omega}_O + \dot{\omega}_P = 0$.

Energy balance—In the low Mach number approximation, the total enthalpy balance reads:

$$\sum_{i\in\mathscr{I}} c_{\mathrm{p},i}\left[\partial_t(\rho y_i\theta) + \mathrm{div}(\rho y_i\theta \boldsymbol{u})\right] - \mathrm{div}(\lambda\nabla\theta) = \dot{\omega}_\theta, \quad \dot{\omega}_\theta = -\sum_{i\in\mathscr{I}} \Delta h^0_{f,i}\dot{\omega}_i, \quad (3)$$

where θ stands for the temperature, $c_{\mathrm{p},i}$ for the specific heat of the species i (supposed to be constant), $\Delta h^0_{f,i}$ for the formation enthalpy at 0°K and λ the thermal conductivity. This equation is complemented by a total flux boundary condition at the inlet boundary, and we suppose that the diffusion flux vanishes at the outlet boundary.

Equation of state—We suppose that the gas phase is a mixture of perfect gases and that the density ρ_F of the solid phase (*i.e.* of the fuel) is constant, so:

$$\rho = \rho\big(\theta, (y_i)_{1\le 1\le N_s}\big) = \frac{1}{\dfrac{R\theta}{P_{th}}\displaystyle\sum_{i=O,P,N}\frac{y_i}{W_i} + \frac{y_F}{\rho_F}}, \quad (4)$$

where $R = 8.31451\,\mathrm{JK}^{-1}\mathrm{mol}^{-1}$ stands for the perfect gases constant. Since the computational domain is supposed not to be closed, the so-called thermodynamic pressure P_{th} is constant in time and space, and given by the initial state.

2.2 The Numerical Scheme

For the solution of the equations of the model, we define the variable z as follows:

$$z = \frac{s\,y_F + 1 - y_O}{1+s}, \quad \text{with } s = \frac{\nu_O W_O}{\nu_F W_F}.$$

Note that, combining the mass balance equation for the fuel and the oxydant, the variable z satisfies an homogeneous equation; for this reason, we replace the oxydant mass balance equation by the balance equation for z (since, given the values of z and y_F, we may deduce y_O).

Let us consider a partition $0 = t_0 < t_1 < \cdots < t_N = T$ of the time interval $(0, T)$, which we suppose uniform. Let $\delta t = t_{n+1} - t_n$ for $n = 0, 1, \ldots, N-1$ be the constant time step. We suppose that the interval Ω is split into a family of control volumes (sub-intervals of Ω) which realizes a partition of Ω; we denote these control volumes by $(K)_{K\in\mathscr{M}}$. The scalar unknowns, *i.e.* the density, mass fractions and temperature, are associated to the control volumes, and the corresponding unknowns read ρ^n_K, $(y_i)^n_K$, z^n_K and θ^n_K for $K \in \mathscr{M}$, $0 \le n \le N$ and $i \in \mathscr{I}$. The velocity is discretized at

the faces of the mesh, which we denote by $(\sigma)_{\sigma\in\mathscr{E}}$, so the corresponding unknowns are u_σ^n for $\sigma \in \mathscr{E}$ and $0 \le n \le N$. We implement a fractional-step algorithm, which consists in four steps, in order to calculate recursively the unknowns $(y_i)_{i\in\mathscr{I}}^{n+1}$, z^{n+1}, θ^{n+1}, ρ^{n+1} and u^{n+1} for $0 \le n < N$:

Chemistry step—Solve for $(y_N, z, y_F, y_P)^{n+1}$:
$\forall K \in \mathscr{M}$,

$$\frac{1}{\delta t}\left[\rho_K^n (y_N)_K^{n+1} - \rho_K^{n-1}(y_N)_K^n\right] + \operatorname{div}\left[\rho^n y_N^{n+1}\boldsymbol{u}^n\right]_K = 0, \tag{5a}$$

$$\frac{1}{\delta t}\left[\rho_K^n z_K^{n+1} - \rho_K^{n-1} z_K^n\right] + \operatorname{div}\left[\rho^n z^{n+1}\boldsymbol{u}^n\right]_K = 0, \tag{5b}$$

$$\frac{1}{\delta t}\left[\rho_K^n (y_F)_K^{n+1} - \rho_K^{n-1}(y_F)_K^n\right] + \operatorname{div}\left[\rho^n y_F^{n+1}\boldsymbol{u}^n\right]_K = (\dot\omega_F)_K^{n+1}, \tag{5c}$$

$$(y_F)_K^{n+1} + (y_O)_K^{n+1} + (y_N)_K^{n+1} + (y_P)_K^{n+1} = 1. \tag{5d}$$

Energy balance—Solve for θ^{n+1} :

$\forall K \in \mathscr{M}$,

$$\sum_{i\in\mathscr{I}} c_{\mathrm{p},i}\left[\frac{1}{\delta t}\left[\rho_K^n (y_i)_K^{n+1}\theta_K^{n+1} - \rho_K^{n-1}(y_i)_K^n\theta_K^n\right]\right] + \operatorname{div}\left[\rho^n y_i^{n+1}\theta^{n+1}\boldsymbol{u}^n\right]_K - \operatorname{div}(\lambda\nabla\theta^{n+1})_K = (\dot\omega_\theta)_K^{n+1}. \tag{5e}$$

Equation of state—$\rho_K^{n+1} = \varrho\left(\theta_K^{n+1}, ((y_i)_K^{n+1})_{1\le 1\le N_s}\right)$, for $K \in \mathscr{M}$. (5f)

Mass balance—Solve for $\boldsymbol{u}^{n+1}$:

$$\forall K \in \mathscr{M}, \qquad \frac{1}{\delta t}\left[\rho_K^{n+1} - \rho_K^n\right] + \operatorname{div}\left[\rho^{n+1}\boldsymbol{u}^{n+1}\right]_K = 0. \tag{5g}$$

The discrete operators appearing in these relations are approximated by finite-volume techniques. Thanks to a careful definition of the convection fluxes, derived to fulfill the conditions introduced in [3] to obtain a maximum-principle-preserving convection operators, the scheme is proven in [2] to preserve the physical bounds of the unknowns: mass fractions in the interval [0, 1], positivity of the temperature and so of the density.

3 A Model Based on an Explicit Tracking of the Flame Front

3.1 The Governing Equations

The physical model addressed in this section is based on an explicit computation of the flame brush location, by a phase-field-like technique. The "color function" is called G in this context; its transport equation is referred to as the G-equation,

and reads:

$$\partial_t(\rho G) + \mathrm{div}(\rho G \boldsymbol{u}) + \rho_u u_f |\nabla G| = 0. \tag{6}$$

Initial conditions are $G = 0$ at the location where the flame starts and $G = 1$ elsewhere. The quantity ρ_u is a constant density, which, from a physical point of view, stands for a characteristic value for the unburnt gases density, and u_f is the flame brush velocity. The reactive term $\dot{\omega}$ is given by:

$$\dot{\omega} = \frac{u_f}{\delta}\, \eta(y_F, y_O)\, (G - 0.5)^-, \quad \eta(y_F, y_O) = \min\left(\frac{y_F}{\nu_F W_F}, \frac{y_O}{\nu_O W_O}\right), \tag{7}$$

where δ is a quantity homogeneous to a length scale, which governs the thickness of the reaction zone.

The flame velocity model consists of Eq. (6), of the mixture mass balance equation (1), the chemical species mass balance equations (2) (with the modified expression (7) for the chemical reaction term $\dot{\omega}$) and of the energy balance (3). Note that, under some assumptions which are usually not valid in industrial applications, this model may be simplified: for instance, in perfectly premixed situations (*i.e.* constant in space initial data for the chemical mass fractions and the temperature) and supposing an infinitely fast chemical reaction (*i.e.*, in the present formalism, making δ tend to zero), the variable G may be indentified to a progress variable and all the other unknowns (*i.e.* the mass fractions and the temperature) may be deduced from G through an algebraic relation.

3.2 Numerical Scheme

The G function is discretized on the primal mesh, so the discrete unknowns are G_K^n, for $K \in \mathscr{M}$ and $0 \leq n \leq N$. The numerical algorithm differs from the scheme for the primitive formulation, *i.e.* System (5), by the insertion, as a first step, of a discrete counterpart to Eq. (6):

flame brush transport step – Solve for G^{n+1} :

$$\forall K \in \mathscr{M}, \qquad \frac{1}{\delta t}\left[\rho_K^n G_K^{n+1} - \rho_K^{n-1} G_K^n\right] + \mathrm{div}\left[\rho^n G^{n+1} \boldsymbol{u}^n\right]_K + \rho_u u_f |\nabla G|_K^{n+1} = 0.$$

For the discretization of the last term in this relation, we write:

$$|\nabla G| = \frac{\nabla G}{|\nabla G|} \cdot \nabla G, \qquad \text{so} \quad |\nabla G|_K^{n+1} = (\boldsymbol{N}_f^n \cdot \nabla G^{n+1})_K,$$

where $\boldsymbol{N}_f$ is an approximation of the advection field $\nabla G / |\nabla G|$ and we use an upwind finite volume formulation of the transport operator, *i.e.* the formulation obtained by writing $\boldsymbol{N}_f \cdot \nabla G = \mathrm{div}(G\, \boldsymbol{N}_f) - G\, \mathrm{div} \boldsymbol{N}_f$ and using an upwind finite volume (first

or second order) discretization of the convection operator. For the present solver, the convection operators are discretized by an explicit MUSCL-like technique [6].

4 Results

Computations presented in this section are performed with MATLAB for the primitive formulation and by the open-source CALIF^{3}S software developped at IRSN [1] for the flame-velocity model.

Data is chosen in order to allow to check the scheme properties (*i.e.* to avoid unrealistic simplifications, as, for instance, a same specific heat for all the chemical species), and to be in the range of practical applications. The mixture is initially at rest, homogeneous and with a uniform temperature:

$$(y_F)_0 = (y_O)_0 = 0.4, \quad (y_N)_0 = 0.2, \quad (y_P)_0 = 0, \quad \theta_0 = 300^\circ K.$$

In the primitive formulation, the reaction rate follows an Arrhenius law:

$$\dot{\omega}_K = 10^4 \, y_F \, y_O \, e^{-900/\theta}.$$

The molar masses of the chemical species are considered to be equal to 20 g mol^{-1} for all the species, so the combustion reaction reads $F + O + N \longrightarrow 2P + N$, and the initial atmosphere composition is stoichiometric. The temperature diffusion coefficient is $\lambda = 0.005$, the specific heat coefficients (J Kg^{-1} K^{-1}) are $c_{\mathrm{p},N} = 3.\,10^3$, $c_{\mathrm{p},F} = 1.\,10^3$, $c_{\mathrm{p},O} = 2.\,10^3$ and $c_{\mathrm{p},P} = 4.\,10^3$ and the formation enthalpies (J Kg^{-1}) are $\Delta h^0_{f,N} = 3.\,10^6$, $\Delta h^0_{f,F} = 1.\,10^6$, $\Delta h^0_{f,O} = -2.\,10^6$ and $\Delta h^0_{f,P} = -4.\,10^6$ (so the reaction is exothermic). The fuel density is equal to 100 Kg m^{-3}. Ignition is obtained in the primitive formulation by making $\dot{\omega}$ depend in a very thin zone on a fictitious elevated temperature, to trigger a reaction at the initial time. In the flame velocity model, G is imposed to zero in the same zone (while $G = 1$ elsewhere). Since the inflamation zone is very thin, the consequent initial burst is not too violent.

First of all, we observe that, for the primitive formulation, the solution tends after a transition period to a travelling combustion wave separating a fresh (or unburnt) zone from a burnt zone, where $y_F = y_O = 0$, $y_P = 0.8$ and the temperature is equal to the adiabatic combustion temperature. By construction of the scheme, the neutral gas mass fraction y_N and the reduced variable z are kept constant in time and space and equal to their initial value. Since the profile in the interface does not vary in space and time up to a translation velocity u_p (the velocity of the flame brush), we may write the jump conditions for the mixture mass balance equation, to obtain:

$$(\rho_u - \rho_b) \, u_p = \rho_u u_u - \rho_b u_b,$$

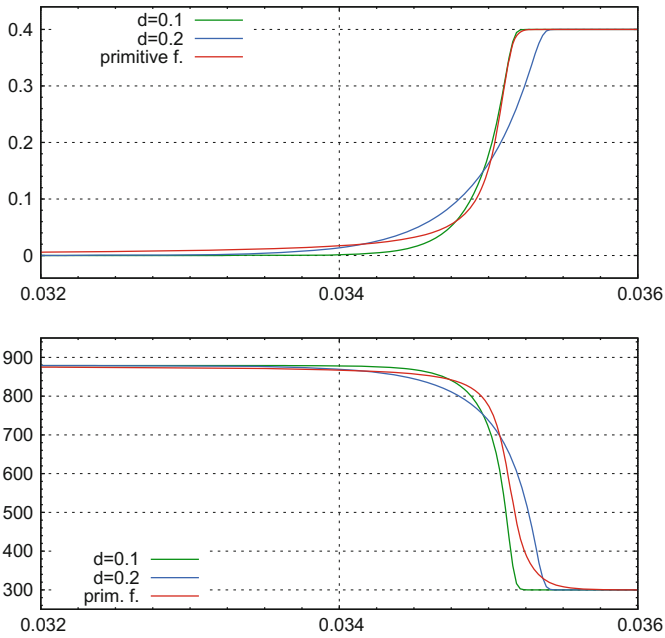


Fig. 1 Fuel mass fraction (*top*) and temperature (*bottom*) travelling profiles obtained with the primitive formulation of the equations (*red*) and with the flame velocity model, with $\delta = 0.1\,\text{mm}$ (*green*) and $\delta = 0.2\,\text{mm}$ (*blue*)

where ρ_b and u_b (resp. ρ_u and u_u) stand for the constant density and velocity in the burnt (resp. unburnt) zone. Thanks to symmetry conditions (due to the fact that the combustion takes place in an atmosphere initially at rest), $u_b = 0$ and we deduce from the previous relation that the flame velocity is given by: $u_f = u_p - u_u = u_u\,\rho_b/(\rho_u - \rho_b)$. The obtained value is injected in the flame velocity model, and we choose the length δ_f to fit as closely as possible the travelling profiles of the unknowns. Results for the fuel mass fraction and the temperature are given on Fig. 1. We observe that, as expected, the thickness of the combustion zone is scaled by δ_f and that a reasonable agreement is obtained with $\delta_f = 0.1\,\text{mm}$.

References

1. CALIF^{3}S: A software components library for the computation of reactive turbulent flows. https://gforge.irsn.fr/gf/project/calif3s
2. D'Amico, M., Dufaud, O., Grapsas, D., Latché, J.C.: A model and a numerical scheme to compute laminar flames in dust suspensions. In: Proceedings of ALGORITMY (2016)
3. Larrouturou, B.: How to preserve the mass fractions positivity when computing compressible multi-component flows. J. Comput. Phys. **95**, 59–84 (1991)

4. Lipatnikov, A., Chomiak, J.: Turbulent flame speed and thickness: phenomenology, evaluation, and application in multi-dimensional simulations. Prog. Energy Combust. Sci. **28**, 1–74 (2002)
5. Peters, N.: Turbulent Combustion. Cambridge University Press, Cambridge Monographs of Mechanics (2000)
6. Piar, L., Babik, F., Herbin, R., Latché, J.C.: A formally second order cell centered scheme for convection-diffusion equations on general grids. Int. J. Numer. Methods Fluids **71**, 873–890 (2013)

High Order Finite Volume Scheme and Conservative Grid Overlapping Technique for Complex Industrial Applications

Grégoire Pont and Pierre Brenner

Abstract The numerical foundation of the CFD solver FLUSEPA (French trademark N. 13400926) is presented. It is a Godunov's type unstructured finite volume method suitable for highly compressible turbulent scale-resolving simulations around 3D complex geometries and general non-Cartesian grids. First, a family of k-exact Godunov schemes is developed by recursively correcting the truncation error of the piecewise polynomial representation of the primitive variables. The keystone of the proposed approach is a quasi-Green gradient operator which ensures consistency on general meshes. In addition, a high-order single-point quadrature formula, based on high-order approximations of the successive derivatives of the solution, is developed for flux integration along curved cell faces. Then, a re-centering process is used to reduce as far as possible numerical diffusion. The proposed family of schemes is compact in the algorithmic sense, since it only involves communications between direct neighbors (cells which have common faces) of the mesh cells. To address complex geometries, a conservative grid intersection technique is used. Compressible numerical test cases are investigated to demonstrate the accuracy and the robustness of the presented numerical scheme, then, supersonic RANS/LES computations around the Ariane 5 space launcher are presented to shows the capability of the scheme to predict flows with shocks, vortical structures and complex geometries.

G. Pont (✉) · P. Brenner
Airbus Safran Launchers, Les Mureaux, France
e-mail: gregoire.pont@airbusafranlaunchers.com

P. Brenner
e-mail: pierre.brenner@airbusafranlaunchers.com

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_32

1 Introduction

The foundations of k-exact approaches can be found in the work of Barth and Frederickson [1] and recent developments can be found in [3]. The idea of improving non-linear flux integral first appears for Cartesian grids using high-order finite differences corrections [4]. The aim of the present paper is to give the outline of a high order k-exact finite volume scheme which keeps its formal accuracy on non-Cartesian grids made of non-planar faces. The proposed scheme corrects automatically the numerical errors in each step of the reconstruction process using a multiple-correction technique based on high order derivatives.

We look for numerical solutions of the compressible Reynolds-averaged or filtered Navier-Stokes equations. This system is approximated by means of a finite volume method on unstructured grids. The computational domain is discretized and contains N cells Ω_J. A_{JK} is a face between J and K cells. In each cell Ω_J, and for any field Ψ, we introduce the space average over one cell $\overline{\Psi}_J$ and the pointwise value at the cell center $\mathbf{x}_j$ of Ω_J, denoted Ψ_j, they are defined, respectively :

$$\overline{\Psi}_J = \frac{1}{\mid \Omega_J \mid} \iiint_{\Omega_J} \Psi \, d\Omega, \quad \text{and} \quad \Psi_j = \Psi(\mathbf{x}_j), \quad \text{with} \quad \mathbf{x}_j = \frac{1}{\mid \Omega_J \mid} \iiint_{\Omega_J} \mathbf{x} \, d\Omega \tag{1}$$

In the above, $\mid \Omega_J \mid = \iiint_{\Omega_J} d\Omega$ denotes the volume of Ω_J. In the present approach, the problem unknowns are represented by the average volume of the conservative variables over each cell, i.e., the values $\overline{\mathbf{w}}_J = \frac{1}{|\Omega_J|} \iiint_{\Omega_J} \mathbf{w} d\Omega$, where $\mathbf{w} = (\rho, \rho\mathbf{u}, \rho E_t)^T$ and where ρ is the density, $\mathbf{u}$ the velocity vector, and E_t the specific total energy. Let us introduce the mth order volume moment of Ω_J, $\mathscr{M}_J^{(m)}$, defined as $\mathscr{M}_J^{(m)} = \frac{1}{|\Omega_J|} \iiint_{\Omega_J} (\mathbf{x} - \mathbf{x}_j)^{\otimes m} d\Omega$, where, for any vector $\mathbf{v}$, $\mathbf{v}^{\otimes m} = \underbrace{\mathbf{v} \otimes \mathbf{v} \otimes \cdots \otimes \mathbf{v}}_{\text{m times}}$, with $\otimes$ the dyadic product between two vectors. We also introduce $\mathscr{S}_{A_{JK}}^{(m)}$, the mth-order moment of A_{JK}, defined as $\mathscr{S}_{A_{JK}}^{(m)} = \iint_{A_{JK}} (\mathbf{x} - \mathbf{x}_\Gamma)^{\otimes m} \cdot \mathbf{n} \, dS$. Note that $\mathbf{x}_\Gamma$ is the integration point of the numerical fluxes (see Eq. (3)). With the preceding notations, the system of conservation laws referred to a computational cell writes:

$$\mid \Omega_J \mid \frac{d\overline{\mathbf{w}}_J}{dt} + \sum_{K=1}^{P_f} \iint_{A_{JK}} \mathbf{F} \cdot \mathbf{n} \, dS = 0 \tag{2}$$

where P_f is the number of faces of cell Ω_J. In the following, we present a k-exact reconstruction of the solution that can be used to find high-order approximations of Eq. (2). For robustness reasons, the reconstruction is applied to the primitive variables vector $\mathbf{q} = (\mathbf{u}, P, T)^T$, (where P and T are respectively the pressure and the temperature), with, $\mathbf{q} = \mathbf{q}(\mathbf{w})$, for example, $\mathbf{u} = (\rho\mathbf{u})/\rho$. For further convenience,

we introduce for each cell J the vector $\widetilde{q}_J = \widetilde{q}(\overline{\mathbf{w}}_J)$ defined as $\widetilde{q}_J = (\widetilde{\mathbf{u}}_J, \widetilde{P}_J, \widetilde{T}_J)^T$, for example, $\widetilde{\mathbf{u}}_J = (\overline{\rho\mathbf{u}})_J/\overline{\rho}_J$. For each grid cell Ω_J let us consider a neighborhood $s(J)$ made by the union of cell Ω_J and of a set of surrounding cells. Specifically, we call $s_1(J)$ the set made of the current cell plus its first neighbors.

In the reconstruction process we start from the $\widetilde{\mathbf{q}}_J$ values. Using a high order correction term $\Delta\mathbf{q}$ we compute $\overline{\mathbf{q}}_J$ values (see Sect. "From $\widetilde{\mathbf{q}}_J$ to $\mathbf{q}_\Gamma$ Values", step 1). Then, we use a successive correction algorithm to compute high order derivatives base on the quasi-Green operator (see Sect. "Computation of Derivatives"). Then, $\mathbf{q}_j$ and $\mathbf{q}_\Gamma$ can be computed using high order formulas (see Sect. "From $\widetilde{\mathbf{q}}_J$ to $\mathbf{q}_\Gamma$ Values" step 2 and 3) and high order derivatives. Finally, fluxes can be integrated using a high order one point integration formula (see Sect. "Integration of the numerical fluxes").

1.1 Integration of the Numerical Fluxes

The construction of higher order schemes relies on the development of a high-accurate approximation formula for the flux term $\mathbf{F}$. This is achieved by expanding $\mathbf{F}$ in Taylor series around a suitable integration point Γ of A_{JK}:

$$\iint_{A_{JK}} \mathbf{F}\cdot\mathbf{n}\,dS = \mathbf{F}|_\Gamma \cdot \mathscr{S}^{(0)}_{A_{JK}} + \sum_{m=1}^{n-1} \frac{1}{m!}\mathbf{D}^{(m)}\,\mathbf{F}|_\Gamma \cdot \mathscr{S}^{(m)}_{A_{JK}} + |A_{JK}|\mathscr{O}(h^n) \qquad (3)$$

where $\mathbf{D}^{(l)}\,\mathbf{F}|_\Gamma$ is the lth tensor derivative of $\mathbf{F}$ at point Γ, and h a characteristic grid size. In the general case, Γ is the point that minimizes the first-order error term (see [5] or [6]). To achieve a nth-order approximation of the surface integral, the flux and its derivatives at point Γ have to be reconstructed from cell-centered averages with suitable accuracy: precisely, $\mathbf{D}^{(m)}\mathbf{F}\big|_\Gamma$ has to be reconstructed at order $n-m$. For this purpose, we choose to rewrite the mth derivatives of the used numerical flux in terms of the mth derivatives of the primitive variables $\mathbf{D}^{(m)}\mathbf{q}$. In turn, these are evaluated by means of a high-order k-exact reconstruction, along with the successive correction algorithm of Sect. "Computation of Derivatives".

1.2 From $\widetilde{\mathbf{q}}_J$ to $\mathbf{q}_\Gamma$ Values

1. $\overline{\mathbf{q}}_J$ **from** $\widetilde{\mathbf{q}}_J$: As a first step, we start writing an approximation for $\overline{\mathbf{q}}_J$. We know that $\overline{\mathbf{q}}_J = \widetilde{\mathbf{q}}_J + \mathscr{O}(h^2)$. To achieve a higher-order approximation for $\overline{\mathbf{q}}_J$, we need to correct the second order error term. Precisely, we look for a relation of the form $\overline{\mathbf{q}}_J = \widetilde{\mathbf{q}}_J + \Delta\mathbf{q} + \mathscr{O}(h^n)$. The correction term $\Delta\mathbf{q}$ depends on the specific primitive variable under consideration ($\mathbf{u}$, P or T). The full expressions for $\Delta\mathbf{q}$

are given in [5] or [6]. It can be shown that $\Delta\mathbf{q}$ depends only on derivatives of $\mathbf{q}$ at point $\mathbf{x}_j$ up to order $n-2$, and not $n-1$ since the error terms depending on $(n-1)$th derivatives are null (see [5] or [6] for more details). Moreover, getting an nth-order approximation for $\overline{\mathbf{q}}_J$ requires only an $(n-m-1)$th order approximation for $\mathbf{D}^{(m)}\mathbf{q}\big|_j$.
For example, if we consider a 2-exact reconstruction on $\mathbf{u}$ one gets:

$$\overline{\mathbf{u}}_J - \widetilde{\mathbf{u}}_J = \Delta\mathbf{u} = -\frac{1}{\overline{\rho}_J}\mathscr{M}_J^{(2)} : \left(\mathbf{D}^{(1)}\rho\big|_j \otimes \mathbf{D}^{(1)}\mathbf{u}\big|_j\right) + \mathscr{O}(h^3) \tag{4}$$

which requires a first order approximation of the gradients at point $\mathbf{x}_j$ to achieve a third-order approximation. Where ':' denotes the inner product between two mth-order tensors. The general formulation for $\Delta\mathbf{u}$, ΔP and ΔT are available for an ideal gaz in [5] or [6].

2. $\mathbf{q}_j$ **from** $\overline{\mathbf{q}}_J$: To calculate approximations of the solution at cell faces, high-order polynomials are constructed over each cell by using again Taylor-series expansions. The polynomial reconstruction is applied to the primitive variables by Taylor-series expanding the primitive variables $\mathbf{q}$ around the cell center $\mathbf{x}_j$, and by averaging over the cell using $\mathscr{M}_J^{(m)}$ and high order derivatives. An approximation of $\mathbf{q}_j$ at order $n > 2$ requires to find an $(n-m)$th-order accurate approximation for the mth derivative $\mathbf{D}^{(m)}\mathbf{q}\big|_j$ as well as an nth-order approximation for $\overline{\mathbf{q}}_J$.
3. $\mathbf{q}_\Gamma$ **from** $\mathbf{q}_j$: Equation (3) involves an nth-order approximation of the flux density at the integration point Γ, $\mathbf{F}|_\Gamma$, as well as $(n-m)$th-order approximations of its mth derivatives. To calculate $\mathbf{F}|_\Gamma$, we use high order extrapolation to obtain $\mathbf{q}_{\Gamma_L}$ and $\mathbf{q}_{\Gamma_R}$ on each side of interfaces. Once the reconstruction of the left and right values has been completed, a slope limiter is finally applied to prevent Gibbs instabilities while preserving high order of accuracy near smooth extrema. Then the 1D exact Riemann solver is used to calculate the numerical flux $\mathbf{F}_\Gamma$ from $\mathbf{q}_{\Gamma_L}$ and $\mathbf{q}_{\Gamma_R}$. The next step is to reconstruct the corrective terms at the right-hand side of Eq. (3). For any given cell face, the numerical flux is univocally defined, thus leading to an intrinsically conservative approximation scheme.

1.3 Computation of Derivatives

The high-order approximation of derivatives $\mathbf{D}^{(m)}\mathbf{q}\big|_j$ and cell average $\overline{q}_J$ required for the k-exact reconstruction and flux integration are obtained by means of a iterative procedure, based on successive corrections of the truncation error terms. At the beginning of the procedure, a first-order approximation of the first derivatives at point $\mathbf{x}_j$ is obtained by applying a 1-exact operator $\mathscr{D}_1^{(1)}$ to the nth order approximation

of $\overline{\mathbf{q}}$ calculated in the step 1 on the Sect. "From $\widetilde{\mathbf{q}}_J$ to $\mathbf{q}_\Gamma$ Values". The next step is to apply this 1-exact operator to itself and to correct the resulting approximations by using a set of correction matrices constructed thanks to k-exact condition on the neighborhood using canonical polynomial functions [3]. Hereafter, we present, the 1-exact operator used, namely, the quasi-Green operator [2], which allows to compute gradient at first order of accuracy regardless of the mesh quality. For a generic function Ψ, the gradient approximation $\mathscr{D}_{\mathbf{1}}^{(1)}$ given by:

$$(\mathscr{D}_{\mathbf{1}}^{(1)}\overline{\Psi})\Big|_J = \mathbf{M}_1^{-1} \sum_{K \in s1(J)} \left(\beta_K \overline{\Psi}_K + (1-\beta_K)\overline{\Psi}_J\right) \mathbf{A}_{JK} \tag{5}$$

is by construction 1-exact, and as such it is at least first-order accurate for any function Ψ and any general mesh. The matrix $\mathbf{M}_1 = \sum_{K \in s_1(J)} \beta_K (\mathbf{x}_k - \mathbf{x}_j) \otimes \mathbf{A}_{JK}$,which replaces the volume Ω_J of the cell J in the standard green approximation formula, is called the "simple correction" matrix (coefficients β_K are weight based on the relative distance of cell centers $\mathbf{x}_k$ to face A_{JK}). It is constructed using the $1 - exactness$ condition applied to the first derivative operator. Since the gradient operator uses only the direct neighborhood, the global successive correction algorithm is also algorithmically compact: the stencil enlargement is consecutive of the iterative process. More details about the algorithm and the corrections matrices can be found in [3, 5, 6].

1.4 Increasing the Accuracy to Resolve Turbulence

In an attempt to reach a good compromise between resolvability and computational cost, we restricted our attention to the third order scheme (2−exact), which provides low dispersion errors. Then, to reduce the numerical dissipation in vortex-dominated regions, a local re-centering process based on a local grid Reynolds number stability condition and also the numerical Ducros sensor is performed. The Vortex Centered (VC) scheme leads to a fourth-order accurate, non-dissipative scheme in vortex-dominated regions, while keeping a third-order accurate upwing scheme elsewhere. Note that this improvement is only valid for viscous flows and particularly well suited for Hybrid RANS/LES approaches where the remaining turbulent viscosity also participates to lower the local Reynolds number. More details about the spectral analysis of the previous k-exact schemes and the VC strategy can be found in [5, 6].

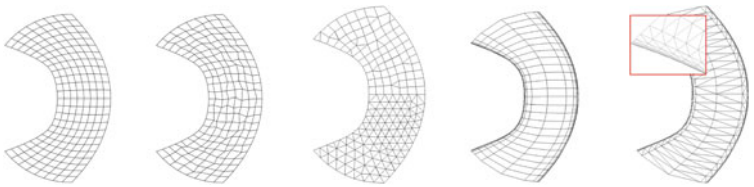
2 Conservative Overlapping Technique

The CHIMERA-like meshing strategy used in the solver FLUSEPA is based on a conservative 3D intersection process [2] which inlays geometrically the highest priority meshes in the others. In the case of a space launcher, for example, the user

meshes independently each part of the launcher, with their own boundary layers, and then, specifies the level of priority of each grid (see Fig. 3). In this study, most of the meshes are generated using this technique.

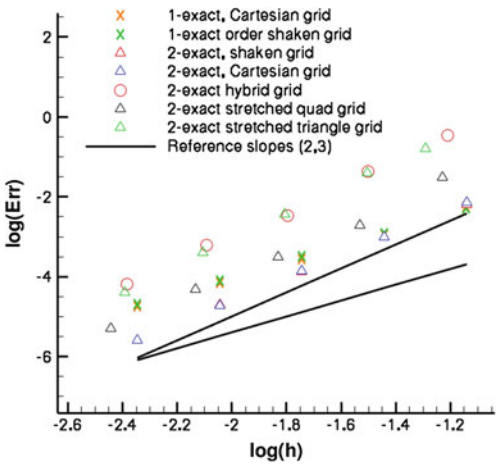
3 Numerical Test : The Ringleb Flow

To demonstrate the accuracy of the numerical schemes, we consider an inviscid transonic test case, the Ringleb flow. Initially, calculations are run using the second (1-exact) and third-order accurate (2-exact) upwind schemes on a set of smooth computational grids composed of 192, 768, 3072, 12288 and 49152 quadrangular cells, respectively (Fig. 1a). Afterwards, the capability of the scheme to preserve accuracy on highly deformed grids is demonstrated by using a set of grids obtained



(a) Cartesian (b) Shaken (c) Overlapped (d) Stretched (e) Stretched triangular

Fig. 1 Ringleb flow: different type of grids used for the calculations



(a) Grid convergence

Fig. 2 Grid convergence. $h = 1/\sqrt{nDOFs}$, where nDOFs is the number of degrees of freedom

by randomly shaking the nodes of the preceding ones (Fig. 1b). Finally, we perform calculation on hybrid grids involving both quadrangular and triangular cells (Fig. 1c), stretched quadrangular grids (Fig. 1d) and stretched triangular grids (Fig. 1e). The results are represented in Fig. 2 for the longitudinal velocity. The computed convergence orders are in fair agreement with the nominal ones, both for the second-order and the third-order upwind scheme. Using highly irregular grids does not affect the convergence order. For grids containing triangular cells, error slopes are in agreement with third-order accuracy, with somewhat higher errors levels than on quadrangular cells.

4 Industrial Applications : Ariane 5 and Unsteady Turbulence

The following application is a 3D 1/60th scale model of the Ariane 5 launcher. Precisely, for the present computations we perform a Delayed Detached Eddy Simulation [7]. The computational mesh is presented in Fig. 3. The total number of grid elements is 20 millions. Numerical experiments have shown that using the VC scheme (see Sect. "Increasing the Accuracy to Resolve Turbulence") is mandatory to correctly trigger convective Kelvin-Helmholtz instabilities seen in Fig. 4a. Figure 4b, c show respectively the comparison between the experiment and the calculation of the pressure and fluctuating pressure coefficients on the red ring located on the nozzle (see Fig. 3b): there is a good agreement for both unsteady and steady pressure coefficient.

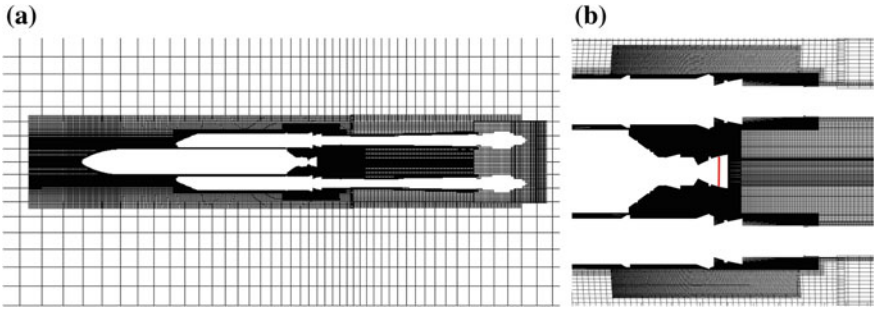
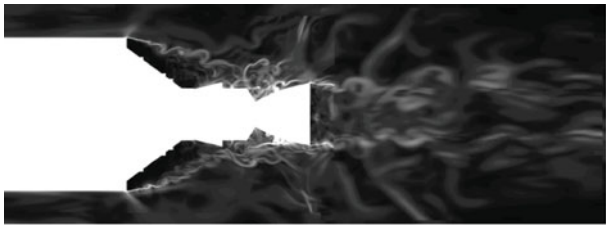
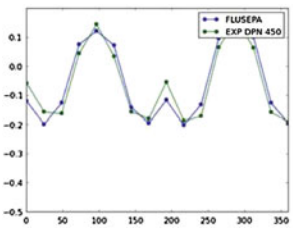


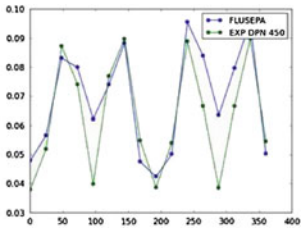
Fig. 3 3D Overlapped grid on the mock up of Ariane 5 spatial launcher (Cut in the boosters plane)



(a) Intantaneous Schlieren at Mach1.2



(b) Pressure coefficient



(c) Unsteady pressure coefficient

Fig. 4 Intantaneous Schlieren and pressure coefficient on a ring located on the nozzle

5 Conclusion

We presented the outlines of a methodology for constructing high-order Godunov-type finite volume methods on general unstructured grids. The procedure efficiently produces a k-exact reconstruction of the primitive variables by recursively correcting the truncation error of lower-order approximations. The proposed successive correction procedure makes use of numerical operators that warrant high accuracy on arbitrary unstructured grids. In the attempt of ensuring low numerical dissipation in vortex-dominated regions while preserving the shock-capturing capabilities of the method and keeping the computational cost as low as possible, a hybrid re-centered discretization is used. First, numerical results for the well-known Ringleb's inviscid flow prove the capability of the present k-exact schemes to keep its nominal order of accuracy on arbitrary grids. Finally, the coupling between the proposed scheme with a conservative overlapping algorithm and an hybrid RANS/LES model seems to be a good strategy for industrial aerodynamic applications.

References

1. Barth, T.J., Frederickson, P.: Higher order solution of the euler equations on unstructured grids using quadratic reconstruction, pp. 90–0013. Technical report. AIAA Paper (1990)
2. Brenner, P.: Unsteady flows about bodies in relative motion. In: 1st AFOSR Conference on Dynamic Motion CFD Proceedings. Rutgers University, New Jersey, USA (1996)
3. Haider, F., Brenner, P., Courbet, B., Croisille, J.P.: Efficient implementation of high order reconstruction in finite volume methods. In: Finite Volumes for Complex Application VI-Problem & Perspectives. Springer Proceedings in Mathematics, vol. 4, pp. 553–560 (2011)
4. McCorquodale, P., Colella, P.: A high-order finite-volume method for conservation laws on locally refined grids. Commun. Appl. Math. Comput. Sci. **6**, 1–25 (2011)
5. Pont, G.: Self adaptive turbulence models for unsteady compressible flows. Ph.D. thesis, Arts et Métiers ParisTech (2015)
6. Pont, G., Brenner, P., Cinnella, P., Robinet, J.C., Maugars, B.: Multiple-correction hybrid k-exact schemes for high-order compressible RANS- LES simulations on general unstructured grids. J. Comput. Phys. (Submmited in May 2016)
7. Spalart, P.R., Deck, S., Shur, S., Squires, M., Strelets, M., Travin, A.: A new version of detached-eddy simulation, resistant to ambiguous grid densities. Theor. Comput. Fluid Dyn. **20**, 181–195 (2006)

Part II
Elliptic and Parabolic Problems

Discontinuous Finite Volume Element Methods for the Optimal Control of Brinkman Equations

Sarvesh Kumar, Ricardo Ruiz-Baier and Ruchi Sandilya

Abstract We introduce and analyse a family of hybrid discretisations based on lowest order discontinuous finite volume elements for the approximation of optimal control problems constrained by the Brinkman equations. The classical optimise-then-discretise approach is employed to handle the control problem leading to a non-symmetric discrete formulation. An a priori error estimate is derived for the control variable in the L^2 − norm, and we exemplify the properties of the method with a numerical test in 3D.

Keywords Brinkman equations · Optimal control problems · Discontinuous finite volume element discretisation

MSC (2010): 49N05 · 49K20 · 65N30 · 76D07 · 76D55

1 Introduction

The numerical solution of optimal control problems constrained by equations of viscous incompressible flow (Stokes and Navier-Stokes problems) is encountered in many application problems arising in science and engineering. An abundant body of relevant literature is available, mainly in the context of finite element methods (see e.g. [3, 6, 7, 13, 14] and the references therein). Most of these contributions employ conforming discretisations for state, co-state and control variables, which typically

S. Kumar · R. Sandilya
Department of Mathematics, Indian Institute of Space Science and Technology, Thiruvananthapuram 695 547, Kerala, India
e-mail: sarvesh@iist.ac.in

R. Sandilya
e-mail: ruchisandilya.12@iist.ac.in

R. Ruiz-Baier (✉)
Mathematical Institute, University of Oxford, A. Wiles Building, Woodstock Road, Oxford OX2 6GG, UK
e-mail: ruizbaier@maths.ox.ac.uk

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_33

produce $\mathcal{O}(h)$ convergence rates for piecewise constant approximations of the control variables, where h is the meshsize. Here we propose a new discontinuous finite volume element (DFVE) method for the discretisation of optimal control problems constrained by the Brinkman equations. DFVE schemes are characterised by ability of writing local conservation equations as in classical finite volume methods, and through transformation maps between primal and dual meshes, they can be recast as discontinuous discretisations of Petrov-Galerkin type. A number of DFVE methods have been proposed for the primal formulation of Stokes and related flow problems in [8, 17] (see also their references). We consider the present method and its analysis as an extension of these contributions to the case of distributed optimal control, in combination with the ideas developed in [11, 12, 15, 16] for elliptic and parabolic optimal control problems. While here we will derive only an L^2- error bound for the control variable and motivate our findings with an example of optimal control in a porous cylinder, the corresponding error estimates in the energy norm for control, state, and co-state variables, as well as numerical verification of optimal convergence rates, will be presented in the forthcoming contribution [9].

2 The Optimal Control Problem

Let us consider the following distributed optimal control problem

$$\min_{\mathbf{u}\in\mathbf{U}_{\text{ad}}} J(\mathbf{u}) := \frac{1}{2}\,\|\mathbf{y}-\mathbf{y}_d\|^2_{0,\Omega} + \frac{\lambda}{2}\,\|\mathbf{u}\|^2_{0,\Omega}\,, \tag{1}$$

governed by the linear Brinkman equations

$$\mathbf{K}^{-1}\mathbf{y} - \mathbf{div}\,(\mu\boldsymbol{\varepsilon}(\mathbf{y}) - p\mathbf{I}) = \mathbf{u}+\mathbf{f} \quad \text{in } \Omega, \tag{2}$$

$$\text{div}\mathbf{y} = 0 \quad \text{in } \Omega, \tag{3}$$

$$\mathbf{y} = \mathbf{0} \quad \text{on } \partial\Omega, \tag{4}$$

where $\mathbf{U}_{\text{ad}}$ is the set of feasible controls (defined for $-\infty \le a_j < b_j \le \infty,\ j = 1, 2, 3$)

$$\mathbf{U}_{\text{ad}} = \{\mathbf{u}\in\mathbf{L}^2(\Omega) : a_j \le u_j \le b_j \text{ a.e. in } \Omega\}.$$

This model describes the motion of an incompressible viscous fluid within an array of porous particles, where $\mathbf{y}$ denotes the fluid velocity, p is the pressure field, $\mathbf{u}$ is the control variable, and $\lambda > 0$ is a given Tikhonov regularisation. The Cauchy stress is $\mu\boldsymbol{\varepsilon}(\mathbf{y}) - p\mathbf{I}$, where $\boldsymbol{\varepsilon}(\mathbf{y}) = \frac{1}{2}(\nabla\mathbf{y} + \nabla\mathbf{y}^T)$ is the infinitesimal rate of strain, $\mu = \mu(\mathbf{x})$ is the dynamic viscosity of the fluid, and $\mathbf{K} = \mathbf{K}(\mathbf{x})$ is for the permeability tensor of the medium (symmetric, uniformly bounded and positive definite). The desired velocity $\mathbf{y}_d$ and the applied body force $\mathbf{f}$ are known data in $\mathbf{L}^2(\Omega)$. The goal

is to identify an additional force $\mathbf{u}$ giving rise to a velocity $\mathbf{y}$ in order to match a given target velocity $\mathbf{y}_d$.

The standard weak formulation of the state equations (2)–(4) is given by: find $(\mathbf{y}, p) \in \mathbf{H}_0^1(\Omega) \times L_0^2(\Omega)$ such that

$$\begin{aligned} a(\mathbf{y}, \mathbf{v}) + c(\mathbf{y}, \mathbf{v}) + b(\mathbf{v}, p) &= (\mathbf{f} + \mathbf{u}, \mathbf{v})_{0,\Omega} \qquad \forall \mathbf{v} \in \mathbf{H}_0^1(\Omega), \\ b(\mathbf{y}, q) &= 0 \qquad \forall q \in L_0^2(\Omega), \end{aligned} \tag{5}$$

where the bilinear forms $a(\cdot,\cdot) : \mathbf{H}_0^1(\Omega) \times \mathbf{H}_0^1(\Omega) \to \mathbb{R}$, $c(\cdot,\cdot) : \mathbf{H}_0^1(\Omega) \times \mathbf{H}_0^1(\Omega) \to \mathbb{R}$ and $b(\cdot,\cdot) : \mathbf{H}_0^1(\Omega) \times L_0^2(\Omega) \to \mathbb{R}$ are defined as:

$$a(\mathbf{y}, \mathbf{v}) = \int_\Omega \mathbf{K}^{-1}\mathbf{y} \cdot \mathbf{v}\, \mathrm{d}\mathbf{x},\ c(\mathbf{y}, \mathbf{v}) = \int_\Omega \mu \boldsymbol{\varepsilon}(\mathbf{y}) : \boldsymbol{\varepsilon}(\mathbf{v})\, \mathrm{d}\mathbf{x},\ b(\mathbf{v}, q) = -\int_\Omega q \mathrm{div}\mathbf{v}\, \mathrm{d}\mathbf{x},$$

for all $\mathbf{y}, \mathbf{v} \in \mathbf{H}_0^1(\Omega)$ and $q \in L_0^2(\Omega)$. Problem (5) satisfies the Babuška-Brezzi condition: there exists $\xi > 0$ such that

$$\inf_{q \in L_0^2(\Omega)} \sup_{\mathbf{0} \neq \mathbf{v} \in \mathbf{H}_0^1(\Omega)} \frac{b(\mathbf{v}, q)}{\|\mathbf{v}\|_{1,\Omega}\, \|q\|_{0,\Omega}} \geq \xi,$$

and its unique solvability is therefore ensured. As the optimal control problem (1)–(4) is strictly convex, it admits a unique optimal solution [10], and the first order necessary conditions are also sufficient for optimality. Moreover, the optimality condition can be formulated as $J'(\mathbf{u})(\tilde{\mathbf{u}} - \mathbf{u}) \geq 0$ for all $\tilde{\mathbf{u}} \in \mathbf{U}_{\mathrm{ad}}$, or:

$$(\mathbf{w} + \lambda \mathbf{u}, \tilde{\mathbf{u}} - \mathbf{u})_{0,\Omega} \geq 0 \quad \forall \tilde{\mathbf{u}} \in \mathbf{U}_{\mathrm{ad}}, \tag{6}$$

where $\mathbf{w}$ is the velocity associated to the adjoint equation

$$\mathbf{K}^{-1}\mathbf{w} - \mathbf{div}(\mu \boldsymbol{\varepsilon}(\mathbf{w}) + r\mathbf{I}) = \mathbf{y} - \mathbf{y}_d \quad \text{in } \Omega, \tag{7}$$

$$\mathrm{div}\mathbf{w} = 0 \quad \text{in } \Omega, \tag{8}$$

$$\mathbf{w} = \mathbf{0} \quad \text{on } \partial\Omega. \tag{9}$$

The variational inequality (6) can be equivalently recast as

$$u_j(\mathbf{x}) = P_{[a_j, b_j]}\left(\frac{-1}{\lambda} w_j(\mathbf{x})\right) \quad \text{a.e. in } \Omega,\ j = 1, 2, 3,$$

where P denotes a projection defined for a generic scalar function f as

$$P_{[a,b]}(f(\mathbf{x})) = \max(a, \min(b, f(\mathbf{x}))), \quad \text{a.e. in } \Omega,$$

and if $f \in W^{1,\infty}(\Omega)$, it further satisfies $\left\|\nabla P_{[a,b]}(f)\right\|_{L^\infty(\Omega)} \leq \|\nabla f\|_{L^\infty(\Omega)}$.

3 Discontinuous Finite Volume Formulation

Let us consider a regular, quasi-uniform partition $\mathscr{T}_h$ of $\bar{\Omega}$ into closed tetrahedra, and referred to as *primal mesh*. By h_T we denote the diameter of a given element $T \in \mathscr{T}_h$, and the global meshsize by $h = \max_{T\in\mathscr{T}_h} h_T$; $\mathscr{E}_h$ and $\mathscr{E}_h^\Gamma$ will denote, respectively, the set of all faces and boundary faces in $\mathscr{T}_h$, and h_e is the area of the face e. In addition, each element $T \in \mathscr{T}_h$ is split into four sub-tetrahedra T_i^*, $i = 1, \ldots, 4$, by connecting the barycentre of the element to its corner nodes (cf. Fig. 1). The set of all these elements generated by barycentric subdivison will be denoted by $\mathscr{T}_h^*$ and will be called *dual partition* of Ω. The symbols $\{\cdot\}$ and $[\![\cdot]\!]$ will denote average and jump operators. A finite dimensional trial space (that will be used for the state and co-state velocity approximation) associated with $\mathscr{T}_h$ is

$$\mathbf{V}_h = \{\mathbf{v}_h \in \mathbf{L}^2(\Omega) : \mathbf{v}_h|_T \in \mathbf{P}_1(T),\ \forall T \in \mathscr{T}_h\},$$

the finite dimensional test space for velocities and corresponding to $\mathscr{T}_h^*$ is

$$\mathbf{V}_h^* = \{\mathbf{v}_h \in \mathbf{L}^2(\Omega) : \mathbf{v}_h|_{T^*} \in \mathbf{P}_0(T^*),\ \forall T^* \in \mathscr{T}_h^*\},$$

and the discrete space for state and co-state pressure approximation is defined as

$$Q_h = \{q_h \in L_0^2(\Omega) : q_h|_T \in P_0(T),\ \forall T \in \mathscr{T}_h\}.$$

In addition we define the higher-regularity space $\mathbf{V}(h) = \mathbf{V}_h + [\mathbf{H}^2(\Omega) \cap \mathbf{H}_0^1(\Omega)]$, and the connection between discrete spaces associated to the two different meshes is characterised by $\gamma : \mathbf{V}(h) \to \mathbf{V}_h^*$, defined from $\gamma\mathbf{v}|_{T^*} = \frac{1}{h_e}\int_e \mathbf{v}|_{T^*}\,\mathrm{d}s$, for $T^* \in \mathscr{T}_h^*$.

Let $\mathbf{v}_h \in \mathbf{V}_h$. We test (2) and (3) against $\gamma\mathbf{v}_h \in \mathbf{V}_h^*$ and $\phi_h \in Q_h$, respectively, and integrate by parts the momentum equation on each dual element and the mass equation on each primal element to obtain: find $(\mathbf{y}_h, p_h) \in \mathbf{V}_h \times Q_h$ such that

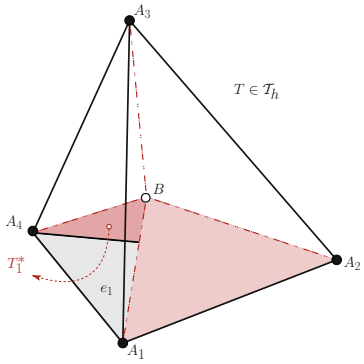


Fig. 1 Sketch of a single primal element T in $\mathscr{T}_h$, and sub-elements T_i^* belonging to the dual partition $\mathscr{T}_h^*$

$$A_h(\mathbf{y}_h, \mathbf{v}_h) + c_h(\mathbf{y}_h, \mathbf{v}_h) + C_h(\mathbf{v}_h, p_h) = (\mathbf{u}_h + \mathbf{f}, \gamma \mathbf{v}_h)_{0,\Omega} \quad \forall \mathbf{v}_h \in \mathbf{V}_h, \tag{10}$$

$$B_h(\mathbf{y}_h, \phi_h) = 0 \quad \forall \phi_h \in Q_h, \tag{11}$$

where the discrete bilinear forms $A_h(\cdot,\cdot)$ and $B_h(\cdot,\cdot)$ are defined as (see also [4]):

$$A_h(\mathbf{w}_h, \mathbf{v}_h) = (\mathbf{K}^{-1}\mathbf{w}_h, \gamma \mathbf{v}_h)_{0,\Omega}, \quad B_h(\mathbf{v}_h, q_h) = b(\mathbf{v}_h, q_h) - \sum_{e \in \mathscr{E}_h} \int_e \{q_h \mathbf{n}\}_e \cdot [\![\gamma \mathbf{v}_h]\!]_e \, \mathrm{d}s,$$

$$\begin{aligned} c_h(\mathbf{w}_h, \mathbf{v}_h) = &- \sum_{T \in \mathscr{T}_h} \sum_{j=1}^{4} \int_{A_{j+1}BA_j} \mu \boldsymbol{\varepsilon}(\mathbf{w}_h)\mathbf{n} \cdot \gamma \mathbf{v}_h \, \mathrm{d}s - \sum_{e \in \mathscr{E}_h} \int_e \{\mu \boldsymbol{\varepsilon}(\mathbf{w}_h)\mathbf{n}\}_e \cdot [\![\gamma \mathbf{v}_h]\!]_e \, \mathrm{d}s \\ &- \sum_{e \in \mathscr{E}_h} \int_e \{\mu \boldsymbol{\varepsilon}(\mathbf{v}_h)\mathbf{n}\}_e \cdot [\![\gamma \mathbf{w}_h]\!]_e \, \mathrm{d}s + \sum_{e \in \mathscr{E}_h} \int_e \frac{\alpha_d}{h_e^{\delta}} [\![\mathbf{w}_h]\!]_e \cdot [\![\mathbf{v}_h]\!]_e \, \mathrm{d}s, \end{aligned}$$

$$C_h(\mathbf{v}_h, q_h) = \sum_{T \in \mathscr{T}_h} \sum_{j=1}^{4} \int_{A_{j+1}BA_j} q_h \mathbf{n} \cdot \gamma \mathbf{v}_h \, \mathrm{d}s + \sum_{e \in \mathscr{E}_h} \int_e \{q_h \mathbf{n}\}_e \cdot [\![\gamma \mathbf{v}_h]\!]_e \, \mathrm{d}s,$$

for all $\mathbf{w}_h, \mathbf{v}_h \in \mathbf{V}_h$ and $q_h \in Q_h$. Here, α_d and δ are parameters independent of h. An appropriate inf-sup condition for B_h can be found in [17].

Analogously, we can state a DFVE formulation for the adjoint equation (7)–(9) as follows: find $(\mathbf{w}_h, r_h) \in \mathbf{V}_h \times Q_h$ such that

$$A_h(\mathbf{w}_h, \mathbf{z}_h) + c_h(\mathbf{w}_h, \mathbf{z}_h) - C_h(\mathbf{z}_h, r_h) = (\mathbf{y}_h - \mathbf{y}_d, \gamma \mathbf{z}_h) \quad \forall \mathbf{z}_h \in \mathbf{V}_h, \tag{12}$$

$$B_h(\mathbf{w}_h, \psi_h) = 0 \quad \forall \psi_h \in Q_h, \tag{13}$$

and introduce the following discrete norms in $\mathbf{V}(h)$:

$$|\!|\!| \mathbf{v}_h |\!|\!|_{1,h}^2 = \sum_{T \in \mathscr{T}_h} |\mathbf{v}_h|_{1,T}^2 + \sum_{e \in \mathscr{E}_h} h_e^{-\delta} \, \| [\![\mathbf{v}_h]\!]_e \|_{0,e}^2, \; |\!|\!| \mathbf{v}_h |\!|\!|_{2,h}^2 = |\!|\!| \mathbf{v}_h |\!|\!|_{1,h}^2 + \sum_{T \in \mathscr{T}_h} h_T^2 |\mathbf{v}_h|_{2,T}^2,$$

which are equivalent on $\mathbf{V}_h$. Next, the discrete counterpart of (6) reads

$$(\mathbf{w}_h + \lambda \mathbf{u}_h, \tilde{\mathbf{u}}_h - \mathbf{u}_h)_{0,\Omega} \geq 0 \qquad \forall \tilde{\mathbf{u}}_h \in \mathbf{U}_{h,\mathrm{ad}}. \tag{14}$$

Lemma 1 *There exist suitable constants $C_i = C_i(\alpha_d)$ independent of h, δ, such that*

$$|A_h(\mathbf{v}, \mathbf{w})| \leq C_1 \, |\!|\!| \mathbf{v} |\!|\!|_{0,\Omega} \, |\!|\!| \mathbf{w} |\!|\!|_{0,\Omega}, \text{ and } |c_h(\mathbf{v}, \mathbf{w})| \leq C_3 \, |\!|\!| \mathbf{v} |\!|\!|_{2,h} \, |\!|\!| \mathbf{w} |\!|\!|_{2,h} \quad \forall \mathbf{v}, \mathbf{w} \in \mathbf{V}(h),$$
$$A_h(\mathbf{v}_h, \mathbf{v}_h) \geq C_2 \, |\!|\!| \mathbf{v}_h |\!|\!|_{0,\Omega}^2 \quad \text{and} \quad c_h(\mathbf{v}_h, \mathbf{v}_h) \geq C_4 \, |\!|\!| \mathbf{v}_h |\!|\!|_{2,h}^2 \quad \forall \mathbf{v}_h \in \mathbf{V}_h.$$

We now turn to the $\mathbf{L}^2-$error analysis for the control field under element-wise constant discretisation, where the discrete control space is defined as

$$\mathbf{U}_h^0 = \{\mathbf{u}_h \in \mathbf{L}^2(\Omega) : \mathbf{u}_h|_T \in \mathbf{P}_0(T) \quad \forall T \in \mathscr{T}_h\}.$$

As in [5], the $\mathbf{L}^2-$projection $\Pi_0 : \mathbf{L}^2(\Omega) \to \mathbf{U}_{h,0}$ is such that there exists a positive constant C independent of h satisfying

$$\|\mathbf{u} - \Pi_0 \mathbf{u}\|_{0,\Omega} \leq Ch \, |\!|\!| \mathbf{u} |\!|\!|_{1,\Omega}, \qquad \mathbf{u} \in \mathbf{H}^1(\Omega). \tag{15}$$

Lemma 2 *Let* $\mathbf{u}$ *be the unique solution of* (1)–(4) *and* $\mathbf{u}_h$ *be the unique control solution of* (10)–(14) *under element-wise constant discretisation (to be verified in* [9]*). Then*

$$\|\mathbf{u} - \mathbf{u}_h\|_{0,\Omega} = \mathscr{O}(h).$$

Proof Since $\Pi_0 \mathbf{U}_{\text{ad}} \subset \mathbf{U}_{h,\text{ad}} := \mathbf{U}_h \cap \mathbf{U}_{\text{ad}}$, the continuous and discrete optimalities readily imply

$$(\mathbf{w} + \lambda \mathbf{u}, \mathbf{u}_h - \mathbf{u})_{0,\Omega} + (\mathbf{w}_h + \lambda \mathbf{u}_h, \Pi_0 \mathbf{u} - \mathbf{u}_h)_{0,\Omega} \geq 0.$$

Adding and subtracting $\mathbf{u}$ and rearranging terms we obtain

$$\lambda \, \|\mathbf{u} - \mathbf{u}_h\|_{0,\Omega}^2 \leq (\mathbf{w} - \mathbf{w}_h, \mathbf{u}_h - \mathbf{u})_{0,\Omega} + (\mathbf{w}_h + \lambda \mathbf{u}_h, \Pi_0 \mathbf{u} - \mathbf{u})_{0,\Omega},$$

and since Π_0 is an orthogonal projection and $\mathbf{u}_h \in \mathbf{U}_{h,\text{ad}}$, then the term $\lambda(\mathbf{u}_h, \Pi_0 \mathbf{u} - \mathbf{u})_{0,\Omega}$ vanishes to give

$$\lambda \, \|\mathbf{u} - \mathbf{u}_h\|_{0,\Omega}^2 \leq (\mathbf{w} - \mathbf{w}_h, \mathbf{u}_h - \mathbf{u})_{0,\Omega} + (\mathbf{w}_h, \Pi_0 \mathbf{u} - \mathbf{u})_{0,\Omega} =: I_1 + I_2. \tag{16}$$

For the first term, we use [11, Theorem 4.1] and arrive at

$$I_1 \leq Ch^2 \, \|\mathbf{u} - \mathbf{u}_h\|_{0,\Omega} + Ch \, \|\mathbf{u} - \mathbf{u}_h\|_{0,\Omega}^2 \,,$$

whereas a bound for I_2 follows from the orthogonality of Π_0:

$$I_2 \leq \|\mathbf{w}_h - \Pi_0 \mathbf{w}_h\|_{0,\Omega} \, \|\Pi_0 \mathbf{u} - \mathbf{u}\|_{0,\Omega} \leq Ch \, |||\mathbf{w}_h|||_{2,h} \, \|\Pi_0 \mathbf{u} - \mathbf{u}\|_{0,\Omega} \,.$$

It is left to show that $\mathbf{w}_h$ is uniformly bounded, which is a consequence of the coercivity of $A_h(\cdot,\cdot)$ and $c_h(\cdot,\cdot)$, and the uniform boundedness of $\mathbf{U}_{h,\text{ad}}$:

$$|||\mathbf{w}_h|||_{2,h} \leq C \left(\|\mathbf{u}_h\|_{0,\Omega} + \|\mathbf{f}\|_{0,\Omega} + \|\mathbf{y}_d\|_{0,\Omega} \right) \leq C.$$

Substituting the bounds for I_1 and I_2 in (16), and using (15), the result follows. □

4 A Numerical Test

We close with the numerical solution of a three-dimensional optimal control problem. The domain consists of a cylinder of height 4 and radius 1, aligned with the x_2 axis. The anisotropic permeability field is characterised by the tensor $\mathbf{K} = \text{diag}(0.1, 10^{-6}\chi_B + 0.1\chi_{B^c}, 0.1)$, where B is a ball of radius 1/4 located at the domain centre. A Poiseuille inflow profile is imposed for the state velocity at the bottom of

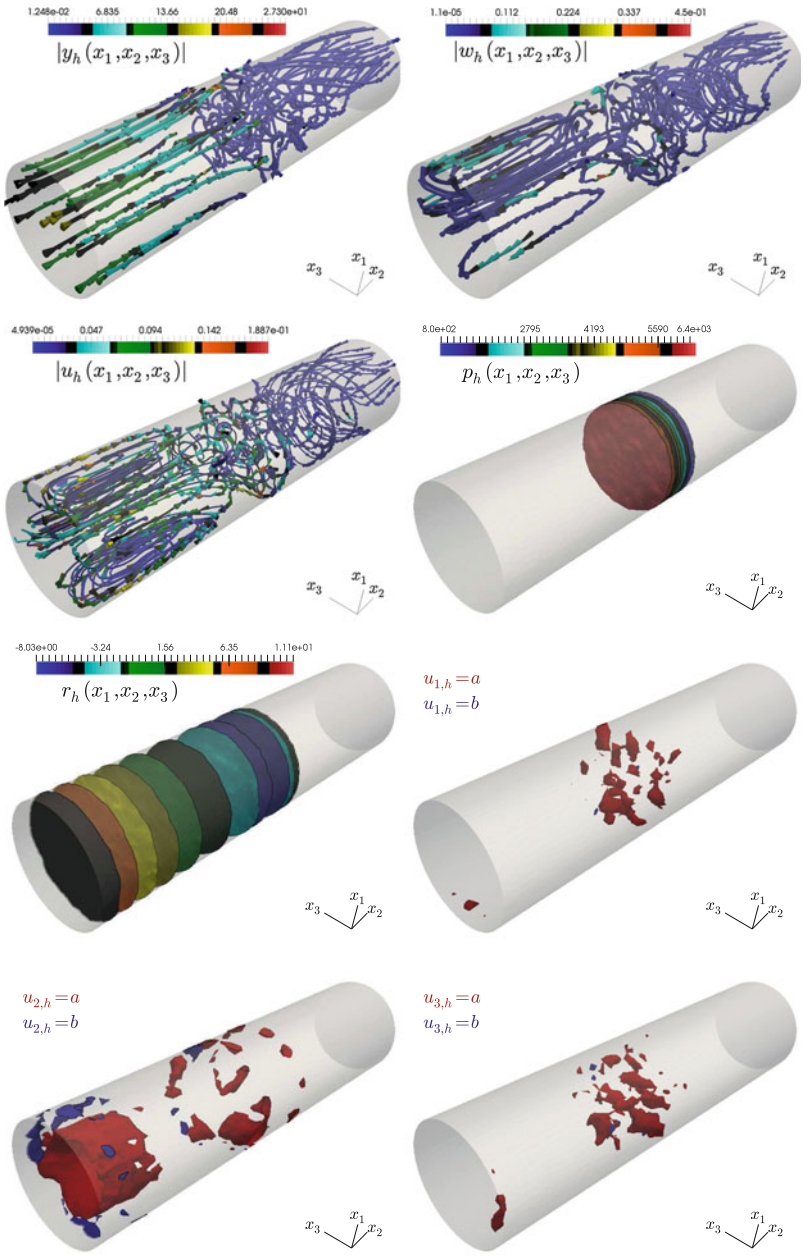


Fig. 2 Streamlines of the DFVE approximation of state and co-state velocities, along with control field, iso-surfaces of computed state and co-state pressures, and iso-surfaces of the control components associated to $a = a_1 = a_2 = a_3$ (in *red*) and $b = b_1 = b_2 = b_3$ (*blue*)

the cylinder (i.e. on $x_2 = 0$): $\mathbf{y} = (0, 10(1 - x_1^2 - (x_3 - 1/2)^2), 0)^T$, a zero-pressure is considered on $x_2 = 4$, whereas homogeneous Dirichlet data are enforced on the remainder of $\partial\Omega$. The viscosity is constant $\mu = 0.01$, the Tikhonov regularisation parameter is $\lambda = 1/2$, the desired velocity is set to zero $\mathbf{y}_d = \mathbf{0}$, the bounds for the control are $a_j = a = -0.1$ and $b_j = b = 0.2$, and a smooth body force is considered as the one in [1]: $\mathbf{f} = \mathbf{K}^{-1}(\exp(-x_2x_3) + x_1\exp(-x_2^2), \cos(\pi x_1)\cos(\pi x_3) - x_2\exp(-x_2^2), -x_1x_2x_3 - x_3\exp(-x_3^2))^T$. The primal mesh has 76766 internal tetrahedral elements and 13663 vertices. The solution is based on the active set strategy [2], involving primal and dual variables, and five iterations of that algorithm are required to reach an adequate stopping criterion. Snapshots of the resulting approximate fields are collected in Fig. 2. The iso-surface of the u_2 component of the control indicates that most of the controlling occurs near the domain centre.

Acknowledgements The authors gratefully acknowledge the support by the Indian National Program on Differential Equations: Theory, Computation and Applications (NPDE-TCA), and by the EPSRC through the Research Grant EP/R00207X/1.

References

1. Anaya, V., Mora, D., Ruiz-Baier, R.: Pure vorticity formulation and Galerkin discretization for the Brinkman equations. IMA J. Numer. Anal. (in press, 2016)
2. Bergounioux, M., Ito, K., Kunisch, K.: Primal-dual strategy for constrained optimal control problems. SIAM J. Control Optim. **37**, 1176–1194 (1999)
3. Braack, M.: Optimal control in fluid mechanics by finite elements with symmetric stabilization. SIAM J. Control Optim. **48**, 672–687 (2009)
4. Bürger, R., Kumar, S., Ruiz-Baier, R.: Discontinuous finite volume element discretization for coupled flow-transport problems arising in models of sedimentation. J. Comput. Phys. **299**, 446–471 (2015)
5. Casas, E., Tröltzsch, F.: Error estimates for linear-quadratic elliptic control problems. IFIP: Anal. Optim. Diff. Syst. **121**, 89–100 (2003)
6. Drăgănescu, A., Soane, A.M.: Multigrid solution of a distributed optimal control problem constrained by the Stokes equations. Appl. Math. Comput. **219**, 5622–5634 (2013)
7. Fourestey, G., Moubachir, M.: Solving inverse problems involving the Navier-Stokes equations discretized by a Lagrange-Galerkin method. Comput. Methods Appl. Mech. Engrg. **194**, 877–906 (2005)
8. Kumar, S., Ruiz-Baier, R.: Equal order discontinuous finite volume element methods for the Stokes problem. J. Sci. Comput. **65**, 956–978 (2015)
9. Kumar, S., Ruiz-Baier, R., Sandilya, R.: Error estimates for a DVFE discretization of the Brinkman optimal control problem (2016). http://infoscience.epfl.ch/record/215779
10. Lions, J.L.: Optimal Control of Systems Governed by Partial Differential Equations. Springer, Berlin (1971)
11. Luo, X., Chen, Y., Huang, Y.: Some error estimates of finite volume element approximation for elliptic optimal control problems. Int. J. Num. Anal. Model. **10**, 697–711 (2013)
12. Nicaise, S., Sirch, D.: Optimal control of the Stokes equations: conforming and non-conforming finite element methods under reduced regularity. Comput. Optim. Appl. **49**, 567–600 (2011)
13. Niu, H., Yuan, L., Yang, D.: Adaptive finite element method for an optimal control problem of Stokes flow with L^2-norm state constraint. Int. J. Numer. Meth. Fluids **69**, 534–549 (2012)
14. Rösch, A., Vexler, B.: Optimal control of the Stokes equations: a priori error analysis for finite element discretization with postprocessing. SIAM J. Numer. Anal. **44**, 1903–1920 (2006)

15. Sandilya, R., Kumar, S.: Convergence analysis of discontinuous finite volume methods for elliptic optimal control problems. Int. J. Comput. Methods **13**, 1640012–20 (2015)
16. Sandilya, R., Kumar, S.: On discontinuous finite volume approximations for semilinear parabolic optimal control problems. Inter. J. Numer. Anal. Model. **13**(4), 545–568 (2016)
17. Ye, X.: A discontinuous finite volume method for the Stokes problems. SIAM J. Numer. Anal. **44**, 183–198 (2006)

Non-isothermal Compositional Two-Phase Darcy Flow: Formulation and Outflow Boundary Condition

L. Beaude, K. Brenner, S. Lopez, R. Masson and F. Smai

Abstract This article deals with the modelling and formulation of compositional gas liquid Darcy flow. Our model includes an advanced boundary condition at the interface between the porous medium and the atmosphere accounting for convective mass and energy transfer, liquid evaporation, and liquid outflow. The formulation is based on a fixed set of unknowns whatever the set of present phases. The thermodynamical equilibrium is expressed as complementary constraints. The model and its formulation are applied to the simulation of the Bouillante high energy geothermal field in Guadeloupe characterized by a high temperature closed to the surface.

Keywords Non-isothermal compositional two-phase Darcy flow model · Geothermal energy · Boundary conditions for the interaction ground-atmosphere · Finite volume scheme

L. Beaude (✉) · K. Brenner · R. Masson
Université Côte d'Azur, Inria, CNRS, LJAD, UMR 7351 CNRS, Parc Valrose,
06108 Nice Cedex 02, France
e-mail: laurence.beaude@unice.fr

K. Brenner
e-mail: konstantin.brenner@unice.fr

R. Masson
e-mail: roland.masson@unice.fr

S. Lopez · F. Smai
BRGM Orléans France, 3 Avenue Claude-Guillemin, BP 36009,
45060 Orléans Cedex 2, France
e-mail: s.lopez@brgm.fr

F. Smai
e-mail: f.smai@brgm.fr

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_34

1 Non-isothermal Compositional Two-Phase Darcy Flow Model

We consider a non-isothermal compositional liquid gas Darcy flow model with $\mathscr{P} = \{g, l\}$ denoting the set of gas and liquid phases. The set of components is denoted by $\mathscr{C}$ including typically a water component which can vaporize in the gas phase and a set of gaseous components which can dissolve in the liquid phase. The thermodynamical properties of each phase $\alpha \in \mathscr{P}$ depend on its pressure P^α, the temperature T and its molar fractions $C^\alpha = \left(C_i^\alpha\right)_{i\in\mathscr{C}}$.

For each phase $\alpha \in \mathscr{P}$, we denote by $\zeta^\alpha(P^\alpha, T, C^\alpha)$ its molar density, by $\rho^\alpha(P^\alpha, T, C^\alpha)$ its mass density, by $\mu^\alpha(P^\alpha, T, C^\alpha)$ its dynamic viscosity, by $e^\alpha(P^\alpha, T, C^\alpha)$ its molar internal energy and by $h^\alpha(P^\alpha, T, C^\alpha)$ its molar enthalpy. Thermodynamical equilibrium between the gas and liquid phases will be assumed for each component and governed by the fugacity functions denoted by $f^\alpha(P^\alpha, T, C^\alpha) = (f_i^\alpha(P^\alpha, T, C^\alpha))_{i\in\mathscr{C}}$.

The rock porosity is denoted by $\phi(\mathbf{x})$ and the rock permeability tensor by $\mathbf{K}(\mathbf{x})$ where $\mathbf{x}$ denotes the spatial coordinates. The hydrodynamical Darcy laws are characterized by the relative permeabilities $k_r^\alpha(S^\alpha)$, function of the phase saturation S^α for each phase $\alpha \in \mathscr{P}$, and by the capillary pressure $P_c(S^g) = P^g - P^l$.

Our formulation of the model is based on the fixed set of unknowns defined by

$$X = \left(P^\alpha, T, S^\alpha, C^\alpha, \alpha \in \mathscr{P}\right). \tag{1}$$

Let $n_i(X)$ be the number of moles of the component $i \in \mathscr{C}$ per unit pore volume defined as

$$n_i(X) = \sum_{\alpha\in\mathscr{P}} \zeta^\alpha\, S^\alpha\, C_i^\alpha, \quad i \in \mathscr{C}.$$

We introduce the rock energy per unit rock volume defined by $E_r(P^\alpha, T)$ and the fluid energy per unit pore volume defined by

$$E(X) = \sum_{\alpha\in\mathscr{P}} \zeta^\alpha S^\alpha e^\alpha.$$

Let us denote by $\mathbf{g}$ the gravitational acceleration vector. The Darcy velocity of the phase $\alpha \in \mathscr{P}$ is then given by

$$\mathbf{V}^\alpha = -\frac{k_r^\alpha}{\mu^\alpha}\mathbf{K}(\mathbf{x})\left(\nabla P^\alpha - \rho^\alpha \mathbf{g}\right).$$

The total molar flux of the component $i \in \mathscr{C}$ is denoted by $\mathbf{q}_i$ and the energy flux by $\mathbf{q}_e$, with

$$\mathbf{q}_i = \sum_{\alpha \in \mathscr{P}} C_i^\alpha \zeta^\alpha \mathbf{V}^\alpha, \qquad \mathbf{q}_e = \sum_{\alpha \in \mathscr{P}} h^\alpha \zeta^\alpha \mathbf{V}^\alpha - \lambda \nabla T, \tag{2}$$

where λ stands for the thermal conductivity of the fluid and rock mixture.

The model accounts for the molar conservation of each component $i \in \mathscr{C}$ together with the energy conservation

$$\begin{aligned} \phi(\mathbf{x})\partial_t n_i + \mathrm{div}(\mathbf{q}_i) = 0, \quad i \in \mathscr{C}, \\ \phi(\mathbf{x})\partial_t E + (1 - \phi(\mathbf{x}))\partial_t E_r + \mathrm{div}(\mathbf{q}_e) = 0. \end{aligned} \tag{3}$$

It is complemented by the following capillary relation between the two phase pressures and the pore volume balance

$$\begin{cases} P_c(S^g) = P^g - P^l, \\ \sum\limits_{\alpha \in \mathscr{P}} S^\alpha = 1. \end{cases} \tag{4}$$

Due to change of phase reactions assumed to be at equilibrium, phases can appear or disappear. In our formulation the molar fractions C^α of an absent phase α are extended by the ones at equilibrium with the present phase. It results that the thermodynamical equilibrium can be expressed as the following complementary constraints for each phase $\alpha \in \mathscr{P}$ combined with the equality of the gas and liquid fugacities of each component [5]

$$\begin{cases} S^\alpha (1 - \sum\limits_{i \in \mathscr{C}} C_i^\alpha) = 0, \quad \alpha \in \mathscr{P}, \\ S^\alpha \geq 0, \quad 1 - \sum\limits_{i \in \mathscr{C}} C_i^\alpha \geq 0, \\ f_i^g(P^g, T, C^g) = f_i^l(P^l, T, C^l), \quad i \in \mathscr{C}. \end{cases} \tag{5}$$

Note that our formulation of the model leads to a fix set of unknowns and equations which is independent of the set of present phases and expresses the thermodynamical equilibrium as complementary constraints. This will allow the use of non-smooth Newton methods to solve the non-linear systems at each time step of the simulation as specified in the numerical section.

2 Boundary Condition at the Interface Between the Porous Medium and the Atmosphere

The fluid and energy transport in high energy geothermal systems is deeply governed by the conditions set at the boundary of the computational domain. In particular, it is well known that the modelling of the interaction between the porous medium model and the atmosphere plays an important role [6]. In this section we propose a boundary condition model taking into account the convective molar and energy transfer and

the vaporization of the liquid phase in the atmosphere as well as a liquid outflow condition.

The convective molar and energy boundary layers induced by the turbulent gas flow in the atmosphere are expressed using two boundary layer thicknesses denoted by δ_m for the molar convective transfer and by δ_T for the energy convective transfer. Let us also introduce the additional unknown $q^{g,atm}$ accounting for the gas molar flow rate at the interface on the atmosphere side oriented outward from the porous medium domain. The liquid phase is assumed to vaporize instantaneously when leaving the porous medium as long as the atmosphere is not saturated with water vapour. As soon as the atmosphere is vapour saturated at the interface, a liquid molar flow rate $q^{l,atm}$ is allowed to exit the porous medium. The prescribed far field atmospheric conditions are defined by the gas molar fractions $C_\infty^{g,atm}$, the temperature T_∞^{atm} and the gas pressure P^{atm}. The model assumes the continuity of the gas phase characterized by the continuity of the gas pressure $P^g = P^{atm}$, of the temperature T and of the gas molar fractions C^g at the interface. Let us recall that P^l is the liquid pressure and C^l the liquid molar fractions at the interface on the porous medium side. We introduce the liquid molar fractions $C^{l,atm} = (C_i^{l,atm})_{i\in\mathscr{C}}$ at the interface on the atmosphere side by the one at thermodynamical equilibrium with the gas phase. It is obtained by the equation $f^l(P^{atm}, T, C^{l,atm}) = f^g(P^g, T, C^g)$. Note that, due to the jump of the capillary pressure which vanishes on the atmosphere side, $C^{l,atm}$ does not match in general with C^l which satisfies $f^l(P^l, T, C^l) = f^g(P^g, T, C^g)$.

Let us denote by $(u)^+$ (resp. $(u)^-$) the positive part (the negative part) of the variable u such that $(u)^+ = max(0, u)$ (resp.$(u)^- = max(0, -u)$).

At the interface, on the atmosphere side, the component gas molar normal flux $q_i^{g,atm}$, $i \in \mathscr{C}$ and the gas energy normal flux $q_e^{g,atm}$ are defined by

$$q_i^{g,atm} = (q^{g,atm})^+ C_i^g - (q^{g,atm})^- C_{i,\infty}^{g,atm} + \frac{\zeta^g D^g}{\delta_m}\left(C_i^g - C_{i,\infty}^{g,atm}\right), \quad i \in \mathscr{C},$$
$$q_e^{g,atm} = (q^{g,atm})^+ h^g(P^g, T, C^g) - (q^{g,atm})^- h_\infty^{g,atm} + \frac{\lambda^g}{\delta_T}(T - T_\infty^{atm}),$$

where D^g is the gas molecular diffusion coefficient, λ^g is the gas thermal conductivity and $h_\infty^{g,atm} = h^g(P^{atm}, T_\infty^{atm}, C_\infty^{g,atm})$ is the far field atmospheric gas enthalpy.

The model prescribes the continuity at the interface of the molar and energy normal fluxes:

$$\begin{cases} \mathbf{q}_i \cdot \mathbf{n} = q_i^{g,atm} + C_i^{l,atm} q^{l,atm}, & i \in \mathscr{C}, \\ \mathbf{q}_e \cdot \mathbf{n} = q_e^{g,atm} + h^l(P^g, T, C^{l,atm}) q^{l,atm}, \end{cases} \tag{6}$$

where the unit normal vector $\mathbf{n}$ at the interface is oriented outward from the porous medium domain.

The liquid molar overflow rate $q^{l,atm}$ is determined by the following complementary constraints accounting for the thermodynamical equilibrium between the liquid and gas phases at the interface in the atmosphere:

$$\begin{cases} (1 - \sum_{i \in \mathscr{C}} C_i^{l,atm}) q^{l,atm} = 0, \\ 1 - \sum_{i \in \mathscr{C}} C_i^{l,atm} \geq 0, \quad q^{l,atm} \geq 0. \end{cases} \tag{7}$$

It remains to eliminate the liquid molar fractions $C^{l,atm}$ from Eqs. (6) and (7). Following [4], let us consider for $f \in \mathbb{R}^{\mathscr{C}}$ the function $\mathscr{C}^l(f, P^l, T) \in \mathbb{R}^{\mathscr{C}}$ defined as the unique solution of the equation $f^l(P^l, T, C^l) = f$.
From $f^g(P^g, T, C^g) = f^l(P^g, T, C^{l,atm}) = f^l(P^l, T, C^l)$ it results that

$$C^{l,atm} = \mathscr{C}^l(f^l(P^l, T, C^l), P^g, T).$$

On the one hand, if $S^l > 0$, it follows that

$$\begin{aligned} 1 - \sum_{i \in \mathscr{C}} C_i^{l,atm} &= \sum_{i \in \mathscr{C}} \left(C_i^l - C_i^{l,atm} \right) \\ &= \sum_{i \in \mathscr{C}} \left(\mathscr{C}_i^l(f^l(P^l, T, C^l), P^l, T) - \mathscr{C}_i^l(f^l(P^l, T, C^l), P^g, T) \right). \end{aligned} \tag{8}$$

Following [5], we can assume that the function $\sum_{i \in \mathscr{C}} \mathscr{C}_i^l(f, P, T)$ is strictly decreasing with respect to P, it results that the complementary constraints (7) is equivalent to

$$\begin{cases} (P^g - P^l) q^{l,atm} = 0, \\ P^g - P^l \geq 0, \; q^{l,atm} \geq 0. \end{cases} \tag{9}$$

On the other hand, if $S^l = 0$ then one has $P^g - P^l = P_c(1) > 0$ and $\sum_{i \in \mathscr{C}} C_i^{l,atm} < 1$. It results that both conditions (9) and (7) imply that $q^{l,atm} = 0$. Finally, let us remark that (9) and $C^l = \mathscr{C}^l(f^l(P^l, T, C^l), P^l, T)$ imply that $C^{l,atm}$ can be replaced by C^l in the normal flux continuity equations (6).

In order to account for a non zero entry pressure for the capillary function $P_c(S^g)$, let us choose P_c as primary unknown rather than S^g and denote by $\mathscr{S}^g(P_c)$ the inverse of the monotone graph extension of the capillary pressure. As detailed in [2], a switch of variable between S^g and P_c could also be used in order to account for non invertible capillary functions.

To conclude, our evaporation - overflow boundary condition model is defined at the interface by the set of unknowns $X_\Gamma = (q^{g,atm}, q^{l,atm}, T, P^\alpha, S^\alpha, C^\alpha, \alpha \in \mathscr{P})$ and the set of equations:

$$
\left\{
\begin{array}{l}
\mathbf{q}_i \cdot \mathbf{n} = (q^{g,atm})^+ C_i^g - (q^{g,atm})^- C_{i,\infty}^{g,atm} + \frac{\zeta^g D^g}{\delta_m} \left(C_i^g - C_{i,\infty}^{g,atm} \right) + C_i^l q^{l,atm}, \quad i \in \mathscr{C} \\
\mathbf{q}_e \cdot \mathbf{n} = (q^{g,atm})^+ h^g(P^g, T, C^g) - (q^{g,atm})^- h_\infty^{g,atm} + \frac{\lambda^g}{\delta_T}(T - T_\infty^{atm}) \\
\qquad\qquad + h^l(P^l, T, C^l) q^{l,atm}, \\
P^g = P^{atm}, \\
S^g = \mathscr{S}^g(P^g - P^l), \\
S^l + S^g = 1, \\
\sum\limits_{i \in \mathscr{C}} C_i^g = 1, \\
S^l (1 - \sum\limits_{i \in \mathscr{C}} C_i^l) = 0, \qquad S^l \geq 0, \qquad 1 - \sum\limits_{i \in \mathscr{C}} C_i^l \geq 0, \\
f_i^g(P^g, T, C^g) = f_i^l(P^l, T, C^l), \qquad i \in \mathscr{C} \\
(P^g - P^l) q^{l,atm} = 0, \qquad P^g - P^l \geq 0, \qquad q^{l,atm} \geq 0.
\end{array}
\right.
$$

3 Numerical Tests

The system of equations is discretized using a fully implicit Euler scheme in time and a finite volume discretization in space with a Two Point Flux Approximation (TPFA) [3]. The mobility terms of each phase are upwinded with respect to the sign of the phase Darcy flux. The non linear system is solved at each time step by a semi-smooth Newton algorithm (Newton-Min) adapted to complementary constraints [1]. In order to reduce the size of the linear systems to $\#\mathscr{C} + 1$ equations and unknowns in each degrees of freedom (the cells and boundary faces), the set of unknowns is splitted into $\#\mathscr{C} + 1$ primary unknowns and remaining secondary unknowns. This splitting is done for each degree of freedom in such a way that the Jacobian of the local closure laws with respect to the secondary unknowns is non singular (Table 1). Note that the non linear convergence criterion is prescribed on the maximum of the relative norms of the energy balance equation residual and of each component mole balance equation residual. This relative norm is defined as the ratio of the residual $l1$-norm by the initial residual $l1$-norm.

Table 1 Choices of the primary unknowns depending on the complementary constraints

Evaporation—overflow boundary		Interior cell and other boundaries	
$q^{l,atm} < P^g - P^l$ $1 - \sum\limits_{i \in \mathscr{C}} C_i^l < S^l$	$q^{g,atm}$, P_c, $(C_i^l)_{i=1,\#\mathscr{C}-1}$	$1 - \sum\limits_{i \in \mathscr{C}} C_i^g < S^g$ $1 - \sum\limits_{i \in \mathscr{C}} C_i^l < S^l$	P^g, S^g, $(C_i^l)_{i=1,\#\mathscr{C}-1}$
$P^g - P^l < q^{l,atm}$ $1 - \sum\limits_{i \in \mathscr{C}} C_i^l < S^l$	$q^{g,atm}$, $q^{l,atm}$, T, $(C_i^l)_{i=1,\#\mathscr{C}-2}$	$S^g < 1 - \sum\limits_{i \in \mathscr{C}} C_i^g$ $1 - \sum\limits_{i \in \mathscr{C}} C_i^l < S^l$	P^g, T, $(C_i^l)_{i=1,\#\mathscr{C}-1}$
$q^{l,atm} < P^g - P^l$ $S^l < 1 - \sum\limits_{i \in \mathscr{C}} C_i^l$	$q^{g,atm}$, T, $(C_i^g)_{i=1,\#\mathscr{C}-1}$	$1 - \sum\limits_{i \in \mathscr{C}} C_i^g < S^g$ $S^l < 1 - \sum\limits_{i \in \mathscr{C}} C_i^l$	P^g, T, $(C_i^g)_{i=1,\#\mathscr{C}-1}$

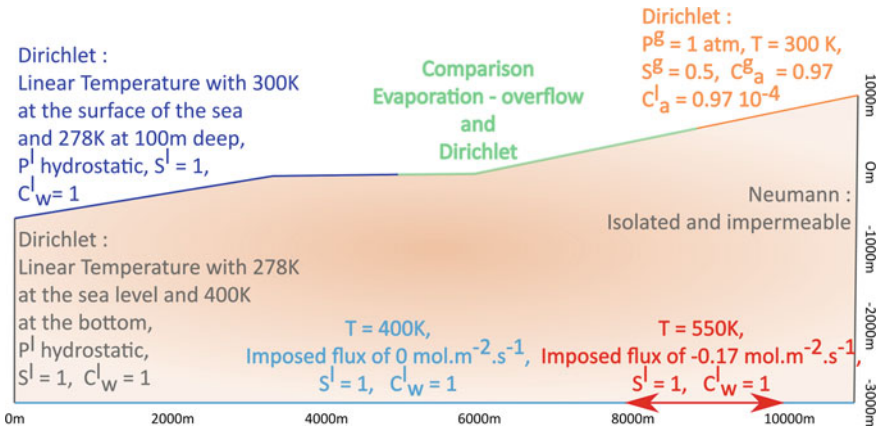


Fig. 1 Illustration of the 2D domain and the boundary conditions of the test case

The impact of the boundary condition is studied over a 2D dimensional test case representing a simplified domain of the Bouillante geothermal reservoir. A Voronoi mesh satisfying the admissibility condition of TPFA schemes at both inner and boundary faces is used [3]. We consider an homogeneous porous medium of porosity $\phi(\mathbf{x}) = 0.35$ and isotropic permeability $\mathbf{K}(\mathbf{x}) = K * I$ with $K = 1D$. The relative permeabilities are defined as $k_r^\alpha(S^\alpha) = (S^\alpha)^2$ for each phase $\alpha \in \mathscr{P}$. The capillary pressure function is given by the Corey law $P_c(S^g) = -b\ln(1 - S^g)$ for $S^g \in [0, s_1]$ and by $P_c(S^g) = -b\ln(1 - s_1) + \frac{b}{1-s_1}(S^g - s_1)$ for $S^g \in (s_1, 1]$ with $b = 2\ 10^5$ Pa and $s_1 = 0.99$. The capillary pressure is regularized to allow the disappearance of the liquid phase. The liquid and gas phases are a mixture of two components, the water denoted by w and the air denoted by a.

The gas thermodynamical laws are defined by the perfect gas molar density $\zeta^g = \frac{P^g}{RT}$, with $R = 8.314\ J.K^{-1}.mol^{-1}$ and the viscosity $\mu^g = (0.361T - 10.2)\ 10^{-7}$ $Pa.s$. The liquid molar density and viscosity as well as the liquid and gas enthalpies are taken from [7]. The vapour pressure $P_{sat}(T)$ is given by the Clausius-Clapeyron equation and the Henry constant of the air component is set to $H_a = 10^8\ Pa$. The molar internal energy of each phase is considered to be equal to its enthalpy. Finally, the fugacities are defined by

$$\begin{cases} f_i^g = C_i^g P^g, \quad i = a, w, \\ f_a^l = C_a^l H_a, \\ f_w^l = C_w^l P_{sat}(T) exp\left(-\frac{P_{sat}(T) - P^l}{1000RT/0.018}\right). \end{cases}$$

The simulation is run over 400 years, with an initial time step of 5 days and a maximum time step of 700 days. The mesh contains approximatively 3000 cells and is refined at the neighbourhood of the top boundary with a volume ratio of 29 between the smallest and the largest cells of the mesh.

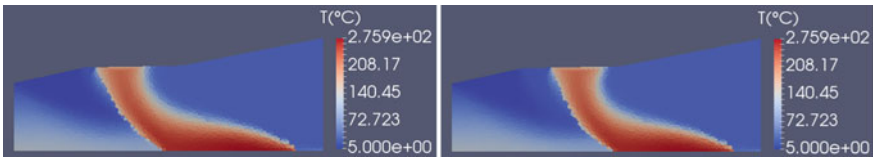


Fig. 2 Temperature at final time with the Evaporation—Overflow boundary condition (on the *left*) and with Dirichlet boundary condition (on the *right*)

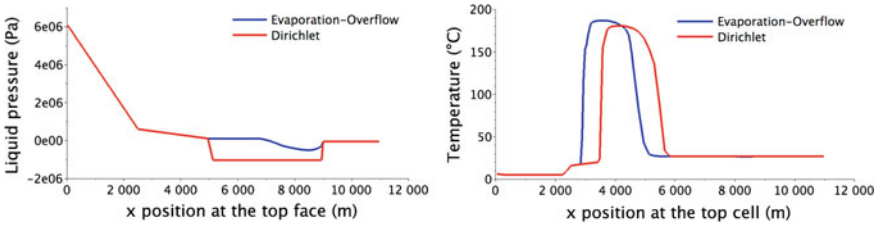


Fig. 3 Plots of the liquid pressure at the *top* boundary and of the temperature at the *top* cell at final time with the Evaporation—Overflow and Dirichlet boundary conditions

The convective molar and energy transfer layer thicknesses are fixed to $\delta_m = \delta_T = 10^{-1}\, m$. The far field atmospheric conditions are set to $C_{a,\infty}^{g,atm} = 0.98$, $C_{w,\infty}^{g,atm} = 0.02$, $T_\infty^{atm} = 300\ K$ and $P^{atm} = 1\ atm$.

The solution obtained using our evaporation - overflow boundary condition is compared with the solution obtained using a Dirichlet boundary condition prescribing directly the gas saturation $S^g = 1$, molar fractions $C_a^g = 1, C_w^g = 0$, pressure $P^g = 1\ atm$ and temperature $T = 300\ K$ (Fig. 1).

We observe in Figs. 2 and 3 that the evaporation—overflow condition favours the exit of the hot liquid flux in the sea (located between $x = 0$ m and $x = 5000$ m) and provides a better match with what happens in the Bouillante geothermal field. This can be explained by the lower liquid pressure $P^l = P^{atm} - P_c(1)$ provided at the top boundary by the gas Dirichlet condition than the one provided by the evaporation—overflow condition with in particular $P^l = P^g$ between say $x = 5000$ m and $x = 6800$ m as a result of the overflow condition.

Table 2 exhibits the good numerical behaviour of both test cases in terms of non linear and linear convergences. Note that the linear systems are solved using a GMRes iterative solver preconditioned by CPR-AMG.

Table 2 Numerical behaviour for both boundary conditions comparing the number of time steps $N_{\Delta t}$, the number of time step chops N_{chop}, the total number of Newton iterations N_{Newton}, the number of GMRes iterations by Newton iteration N_{GMRes} and the CPU time

	$N_{\Delta t}$	N_{chop}	N_{Newton}	N_{GMRes}	CPU time (s)
Dirichlet boundary condition	333	7	1835	22.3	436
Evaporation—overflow boundary condition	344	3	2072	21.1	404

Acknowledgements We would like to thank the BRGM and the Provence-Alpes-Côte d'Azur Region for the co-funding of the PhD of Laurence Beaude as well as the support of the CHARMS ANR project (ANR-16-CE06-0009).

References

1. Ben Gharbia, I., Jaffré, J.: Gas phase appearance and disappearance as a problem with complementarity constraints. Math. Comput. Simul. (2013)
2. Brenner, K., Groza, M., Jeannin, L., Masson, R., Pellerin, J.: Immiscible two-phase Darcy flow model accounting for vanishing and discontinuous capillary pressures: application to the flow in fractured porous media (Working paper or preprint, 2016)
3. Droniou, J.: Finite volume schemes for diffusion equations: Introduction to and review of modern methods. Math. Models Methods Appl. Sci. **24**(8), 1575–1619 (2014)
4. Lauser, A., Hager, C., Helmig, R., Wohlmuth, B.: A new approach for phase transitions in miscible multi-phase flow in porous media. Adv. Water Resour. **34**, 957–966 (2011)
5. Masson, R., Trenty, L., Zhang, Y.: Formulations of two phase liquid gas compositional Darcy flows with phase transitions. Int. J. Finite Vol. **11**, 34 (2014)
6. O'Sullivan, M.J., Pruess, K., Lippmann, M.J.: Geothermal reservoir simulation: the state-of-practice and emerging trends. Geothermics **30**(4), 395–429 (2001)
7. Schmidt, E.: Properties of Water and Steam in S.I. Units. Springer-Verlag (1969)

Numerical Scheme for a Stratigraphic Model with Erosion Constraint and Nonlinear Gravity Flux

Clément Cancès, Didier Granjeon, Nicolas Peton, Quang Huy Tran and Sylvie Wolf

Abstract In this work, we study an extension of the model introduced by Eymard et al. [*Int. J. Numer. Methods Engrg.* **60**, 527–248 (2004)] for the simulation of large scale transport processes of sediments, subject to an erosion constraint. The novelty we consider lies in the diffusion law relating the flux of sediments and the slope of the topography, that now involves a p-Laplacian with $p > 2$ in order to get more realistic landscape evolutions. This physical sophistication entails the construction of an entirely new numerical scheme, the details of which shall be supplied.

Keywords Stratigraphic forward modeling · Gravity-driven sediment transport · Weather-limited erosion · Evolutionary p-laplacian · Complementarity problem

MSC (2010): 35K85 · 65M08 · 86-08

1 A Constrained Model Arising in Stratigraphic Modeling

We are interested in the evolution of the sediment height $h : \Omega \times \mathbb{R}_+$ where $\Omega = (0, L_x) \times (0, L_y) \subset \mathbb{R}^2$ is a rectangular computational domain, the sea level being fixed to $h = 0$. The sediments are transported from the top to the bottom, due to

C. Cancès
Team RAPSODI, Inria Lille – Nord Europe, 40 avenue Halley, 59650 Villeneuve d'Ascq, France
e-mail: clement.cances@inria.fr

D. Granjeon · N. Peton (✉) · Q.H. Tran · S. Wolf
IFP Energies nouvelles, 1 & 4 avenue de Bois Préau, 92852 Rueil-Malmaison Cedex, France
e-mail: nicolas.peton@ifpen.fr

D. Granjeon
e-mail: didier.granjeon@ifpen.fr

Q.H. Tran
e-mail: quang-huy.tran@ifpen.fr

S. Wolf
e-mail: sylvie.wolf@ifpen.fr

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_35

gravity. The "natural" sediment flux $\mathbf{F} : \Omega \times (0, T) \to \mathbb{R}^2$ is given by

$$\mathbf{F} = -K(h)|\nabla h|^{p-2}\nabla h = -|\nabla h|^{p-2}\nabla \psi(h), \tag{1}$$

where the diffusion coefficient takes the form $K(h) = K_c$ if $h \geq 0$ and $K(h) = K_m$ if $h \leq 0$ (c stands for *continental* and m for *maritime*), and where $\psi(h) = \int_0^h K(a)\mathrm{d}a$. It is known by geomorphologists that sedimentation ($\partial_t h \geq 0$) and erosion ($\partial_t h < 0$) processes are non-symmetric: soil material must first be produced *in situ* by weathering processes prior to being transported by diffusion. As a consequence, it is postulated that the erosion is limited from below by $-E$, where the quantity $E > 0$ corresponds to a known maximal production rate of sediments, i.e.,

$$\partial_t h + E \geq 0 \quad \text{in } \Omega \times \mathbb{R}_+. \tag{2}$$

In order to incorporate this constraint in the problem, we follow the approach of [3] that consists in introducing a multiplier $\lambda : \Omega \times \mathbb{R}_+ \to [0, 1]$ to reduce the flux in a conservative way. More precisely, we impose that

$$\partial_t h + \nabla \cdot (\lambda \mathbf{F}) = 0 \quad \text{in } \Omega \times \mathbb{R}_+, \tag{3a}$$
$$(1 - \lambda)(\partial_t h + E) = 0 \quad \text{in } \Omega \times \mathbb{R}_+, \tag{3b}$$

where (3b) expresses that locally either the erosion constraint is saturated ($\partial_t h = -E$) or the flux is unlimited ($\lambda = 1$). Combining the inequality (2), the reduction assumption $\lambda \leq 1$ with (3), we end up with the synthetic system

$$\partial_t h + \nabla \cdot (\lambda \mathbf{F}) = 0, \qquad \text{in } \Omega \times \mathbb{R}_+, \tag{4a}$$
$$\min\{1 - \lambda,\ \gamma[E - \nabla \cdot (\lambda \mathbf{F})]\} = 0 \qquad \text{in } \Omega \times \mathbb{R}_+, \tag{4b}$$

in which $\mathbf{F}$ is given by (1) and in which the complementarity equation (4b) involves a scaling parameter $\gamma > 0$ whose role is to make the two arguments of the min function homogeneous. We impose the inflow of sediment across the boundary, i.e.,

$$\mathbf{F} \cdot \mathbf{n} = \phi \leq 0 \quad \text{on } \partial\Omega \times \mathbb{R}_+ \tag{5}$$

where $\mathbf{n}$ is the outward normal to $\partial\Omega$. Finally, we prescribe the initial condition

$$h_{|_{t=0}} = h^0 \quad \text{in } \Omega, \quad \text{with} \quad h_\star \leq h^0 \leq h^\star \tag{6}$$

for some $h_\star, h^\star \in \mathbb{R}$. The goal of this contribution is to propose a numerical scheme to approximate the solutions (h, λ) of (1), (4)–(6).

In comparison with the previous contributions [3–5], here our attention is restricted to the case of a single lithology but we lay emphasis on the nonlinearity $p > 2$ in the definition (1) of the flux $\mathbf{F}$. The reason why such a nonlinearity should be incorporated into the model comes from the experimental observation the "linear" diffusion law

$\mathbf{F} = -\nabla\psi(h)$ do not hold for most sedimentary systems of interest. In particular, the linear ($p = 2$) gravity flux implies that sediment propagation occurs at infinite speed. Thus, one of the motivation for considering $p > 2$ is to recover propagation at finite speeds, which in turn enable geologists to track down *knickpoints*.

2 A Cell-Centered Discretization on Cartesian Grids

Roughly speaking, the problem to be solved numerically consists of a constrained evolutionary p-Laplacian, complemented with Neumann boundary conditions. In order to discretize the natural fluxes (1), we use a semi-explicit Finite Volume scheme inspired from [1].

Let N_x, N_y be two positive integers, then $\Omega = (0, L_x) \times (0, L_y)$ is discretized into inner cells $C_{i,j} = ((i-1)\Delta x, i\Delta x) \times ((j-1)\Delta y, j\Delta y)$ where $\Delta x = L_x/N_x$ and $\Delta y = L_y/N_y$. The center of $C_{i,j}$ is denoted by $\mathbf{x}_{i,j} = ((i-1/2)\Delta x, (j-1/2)\Delta y)$. In order to impose the boundary condition (5), we extend the grid with ghost cells. Let

$$\mathscr{L} = \{1, \ldots, N_x\} \times \{0, N_y + 1\} \cup \{0, N_x + 1\} \times \{1, \ldots, N_y\} \tag{7a}$$

$$\mathscr{I} = \{1, \ldots, N_x\} \times \{1, \ldots, N_y\} \tag{7b}$$

be respectively the set of ghost cells (green dots in Fig. 1) and that of inner primal cells. The set of the edges between the primal cells is denoted by $\mathscr{E}$. Two particular subsets of $\mathscr{E}$ will be used in the sequel: the subset $\mathscr{E}_{\text{int}}$ of the inner edges (between two inner cells) and the subset $\mathscr{E}_{\text{ext}}$ of the boundary edges (between an inner cell and a ghost cell), as depicted in Fig. 1. Time is discretized by $0 = t_0 < t_1 < \ldots < t_n < \ldots$, in which the time-step is denoted by $\Delta t_n = t_{n+1} - t_n$.

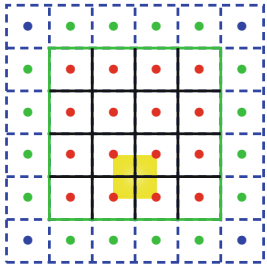


Fig. 1 The original Cartesian grid (*red dots at center*) is surrounded by ghost cells (*dashed blue*) to be used to impose the boundary conditions. We distinguish lateral ghost cells (*green dots at center*) and corner ghost cells (*blue dots at center*). The dual cells (*shaded yellow*) admit the primal cell centers as vertices. Concerning the edges, the inner edges $\mathscr{E}_{\text{int}}$ (*solid black*) and the boundary edges $\mathscr{E}_{\text{ext}}$ (*solid green*) are treated in a different way

The primal unknowns $(h^n_{i,j}, \lambda^n_{i,j})$, for $(i,j) \in \mathscr{I}$ and $n \geq 1$, are located at the centers of the inner cells and of the lateral ghost cells (cf. Fig. 1). The initial data h^0 is discretized into a piecewise-constant function. For $(i,j) \in \mathscr{I}$, we set

$$h^0_{i,j} = \frac{1}{\Delta x \Delta y} \int_{C_{i,j}} h^0(\mathbf{x}) \, \mathrm{d}\mathbf{x}, \tag{8}$$

while for $(i,j) \in \mathscr{L}$, a simple extrapolation is used to obtain $h^0_{i,j}$. To approximate the unconstrained flux $\mathbf{F} \cdot \mathbf{n}$, we first approximate $|\nabla h|^{p-2}$ on the dual cells by

$$\begin{aligned} B^n_{i+1/2,j+1/2} = \Bigg\{ & \frac{1}{2}\left(\frac{h^n_{i+1,j} - h^n_{i,j}}{\Delta x}\right)^2 + \frac{1}{2}\left(\frac{h^n_{i+1,j+1} - h^n_{i,j+1}}{\Delta x}\right)^2 \\ & + \frac{1}{2}\left(\frac{h^n_{i,j+1} - h^n_{i,j}}{\Delta y}\right)^2 + \frac{1}{2}\left(\frac{h^n_{i+1,j+1} - h^n_{i+1,j}}{\Delta y}\right)^2 \Bigg\}^{p/2-1}, \end{aligned} \tag{9}$$

which can be seen as an approximation of $|\nabla h|^2$ raised to the power $p/2 - 1$. It is worth noting that this approximation for $|\nabla h|^2$ is coercive: it cannot vanish unless the four values on the dual cell are identical. The unconstrained flux $\mathbf{F} \cdot \mathbf{n}$ across the inner edges of $\mathscr{E}_{\mathrm{int}}$ at time t_{n+1} is then computed thanks to the semi-implicit formulae

$$F^{n+1}_{i+1/2,j} = \frac{B^n_{i+1/2,j-1/2} + B^n_{i+1/2,j+1/2}}{2} \cdot \frac{\psi(h^{n+1}_{i,j}) - \psi(h^{n+1}_{i+1,j})}{\Delta x}, \tag{10a}$$

$$F^{n+1}_{i,j+1/2} = \frac{B^n_{i-1/2,j+1/2} + B^n_{i+1/2,j+1/2}}{2} \cdot \frac{\psi(h^{n+1}_{i,j}) - \psi(h^{n+1}_{i,j+1})}{\Delta y}, \tag{10b}$$

whereas the boundary fluxes are prescribed by (5), that is,

$$F^{n+1}_{1/2,j} = -\frac{1}{\Delta t_n \Delta y} \int_{t_n}^{t_{n+1}} \int_{(j-1)\Delta y}^{j\Delta y} \phi(x = 0, y) \, \mathrm{d}y \, \mathrm{d}t \tag{11}$$

and similar relations for $F^{n+1}_{N_x+1/2,j}$, $F^{n+1}_{i,1/2}$ and $F^{n+1}_{i,N_y+1/2}$. For $(i,j) \in \mathscr{I}$ and $n \geq 0$, the first equation of (4) is discretized into

$$\frac{h^{n+1}_{i,j} - h^n_{i,j}}{\Delta t_n} + D^{n+1}_{i,j} = 0, \tag{12}$$

where the discrete divergence

$$D_{i,j}^{n+1} = \frac{(\lambda F)_{i+1/2,j}^{n+1} - (\lambda F)_{i-1/2,j}^{n+1}}{\Delta x} + \frac{(\lambda F)_{i,j+1/2}^{n+1} - (\lambda F)_{i,j-1/2}^{n+1}}{\Delta y} \tag{13}$$

involves the upwinded flux

$$(\lambda F)_{i+1/2,j}^{n+1} = \lambda_{i,j}^{n+1}(F_{i+1/2,j}^{n+1})^+ - \lambda_{i+1,j}^{n+1}(F_{i+1/2,j}^{n+1})^-, \tag{14a}$$

$$(\lambda F)_{i,j+1/2}^{n+1} = \lambda_{i,j}^{n+1}(F_{i,j+1/2}^{n+1})^+ - \lambda_{i,j+1}^{n+1}(F_{i,j+1/2}^{n+1})^-, \tag{14b}$$

in which $a^+ = \max\{a, 0\}$ and $a^- = -\min\{a, 0\}$ are the positive and negative parts of the real number a. The boundary fluxes are not limited, so we impose that $\lambda_{i,j}^{n+1} = 1$ for $(i, j) \in \mathscr{L}$. As for the complementarity equation (4b), it is discretized by

$$\min\left\{1 - \lambda_{i,j}^{n+1},\ \frac{\Delta x \Delta y}{\langle F\rangle_{i,j}^{n+1}}[E - D_{i,j}^{n+1}]\right\} = 0, \tag{15}$$

where $\langle F\rangle_{i,j}^{n+1} = \Delta y[(F_{i+1/2,j}^{n+1})^+ + (F_{i-1/2,j}^{n+1})^-] + \Delta x[(F_{i,j+1/2}^{n+1})^+ + (F_{i,j-1/2}^{n+1})^-]$ represents the total unlimited outgoing flux from cell (i, j). The choice of the local weight $\gamma_{i,j}^{n+1} = \Delta x \Delta y / \langle F\rangle_{i,j}^{n+1}$ is justified by the following property.

Lemma 1 *Equation* (15) *is equivalent to*

$$\lambda_{i,j}^{n+1} = \min\left\{1,\ \frac{\Delta x \Delta y E + \rangle \lambda F \langle_{i,j}^{n+1}}{\langle F\rangle_{i,j}^{n+1}}\right\} \tag{16}$$

where $\rangle \lambda F \langle_{i,j}^{n+1} = \Delta y[\lambda_{i-1,j}^{n+1}(F_{i-1/2,j}^{n+1})^+ + \lambda_{i+1,j}^{n+1}(F_{i+1/2,j}^{n+1})^-] + \Delta x[\lambda_{i,j-1}^{n+1}(F_{i,j-1/2}^{n+1})^+ + \lambda_{i,j+1}^{n+1}(F_{i,j+1/2}^{n+1})^-]$ *is the total limited incoming flux into cell* (i, j). *This implies, in particular, that*

$$0 < \lambda_{i,j}^{n+1} \le 1. \tag{17}$$

Proof The discrete divergence (13)–(14) can be easily transformed into

$$D_{i,j}^{n+1} = \frac{\langle F\rangle_{i,j}^{n+1}}{\Delta x \Delta y}\lambda_{i,j}^{n+1} - \frac{\rangle \lambda F \langle_{i,j}^{n+1}}{\Delta x \Delta y}. \tag{18}$$

Multiplication by $-\gamma_{i,j}^{n+1}$ makes $-\lambda_{i,j}^{n+1}$ appear alone in the second argument of the min in (15). As a result, we can extract $-\lambda_{i,j}^{n+1}$ out of the min to obtain (16). We infer from (16) that $\lambda_{i,j}^{n+1} \le 1$, and from (15) that $D_{i,j}^{n+1} \le E$. It follows from (18) that $E\Delta x \Delta y + \rangle \lambda F \langle_{i,j}^{n+1} \ge \langle F\rangle_{i,j}^{n+1} \Delta x \Delta y \ge 0$. From (16), we infer that $\lambda_{i,j}^{n+1} > 0$. □

Contrary to [3–5], we advocate mounting the whole system (12)–(14), (16) in the unknowns $(h_{i,j}^{n+1}, \lambda_{i,j}^{n+1})$. This avoids the task of switching variables according to whether or not the constraint is saturated.

Lemma 2 *For all $n \geq 0$, one has*

$$\min_{(i,j)\in\mathscr{I}} h_{i,j}^n \geq h_\star, \tag{19a}$$

$$\sum_{(i,j)\in\mathscr{I}} h_{i,j}^n \, \Delta x \Delta y = \int_\Omega h^0 \, \mathrm{d}\mathbf{x} - \int_0^{t_n} \int_{\partial\Omega} \phi \, \mathrm{d}\gamma \, \mathrm{d}t. \tag{19b}$$

Proof The above estimates rely on induction. The mass balance (19b) is obtained by summing (12) over $(i, j) \in \mathscr{I}$. To derive (19a), let $(i_\star, j_\star) \in \mathscr{I}$ such that $h_{i_\star,j_\star}^{n+1} = \min_{(i,j)\in\mathscr{I}} h_{i,j}^{n+1}$. Then, $D_{i_\star,j_\star}^{n+1} \leq 0$, hence $h_{i_\star,j_\star}^{n+1} \geq h_{i_\star,j_\star}^n \geq h_\star$. □

As a consequence of Lemma 2, one gets a $L^\infty_{\mathrm{loc}}(\mathbb{R}_+; L^1(\Omega))$ estimate on the discrete sediment height, namely,

$$\sum_i \sum_j \left|h_{i,j}^n\right| \Delta x \Delta y \leq 2h_\star^- |\Omega| + \int_\Omega h^0 \, \mathrm{d}\mathbf{x} - \int_0^{t_n} \int_{\partial\Omega} \phi \, \mathrm{d}\gamma \, \mathrm{d}t. \tag{20}$$

From this and thanks to a topological degree argument [2], we can prove that the scheme admits at least one solution, as claimed in the following Proposition.

Proposition 1 *Let $h_{i,j}^n$, $(i, j) \in \mathscr{I}$, be such that (19) hold. Then, for all $\Delta t_n > 0$, there exists at least one solution $(h_{i,j}^{n+1}, \lambda_{i,j}^{n+1})$ to the nonlinear system (12)–(15) satisfying (17)–(19).*

3 Numerical Results

In order to illustrate the capabilities of the model and the numerical scheme, we show two test cases. We consider a basin of size 180 km × 180 km where the topography represents a continental domain made of mountains ($K_c = 500$ km^2/My) and a marine domain ($K_m = 10$ km^2/My). The sea level is $h = 0$ km. Two incoming fluxes constant in time are prescribed on the left and right borders as

$$-\int_0^T \int_0^{L_y} \phi(0, y, t) \, \mathrm{d}y \, \mathrm{d}t = 24.0T \text{ km}^2, \quad -\int_0^T \int_0^{L_y} \phi(L_x, y, t) \, \mathrm{d}y \, \mathrm{d}t = 26.6T \text{ km}^2.$$

To see the influence of the constraint upon the erosion rate, we consider no flux limitation in the first case (Fig. 2). To this end, we take $E \gg 1$ km/My, such that λ keeps the constant value 1 in the domain. In the second test case (Figs. 3–4), we activate the constraint by taking $E = 0.04$ km/My.

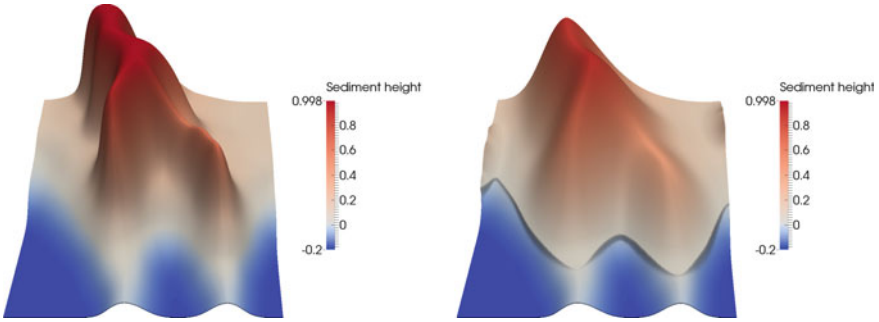


Fig. 2 Case without constraint: initial state (*left*) and final state (*right*) of h (km)

Numerically, the domain is made up of 361×361 cells. The exponent for the p-Laplacian in (1) is set at $p = 2.5$. Simulations are run until $T = 1$ My. At each time step, the nonlinear system is solved within a threshold of 10^{-5} km by Newton's method. The linear system arising at each Newton iteration is solved by means of PETSc routines and using the BiCGSTAB method with the ILU(0) preconditioner. The time steps vary dynamically as follows. We start with $\Delta t_0 = 10^{-4}$ My. If the time iteration is accepted, we set $\Delta t_{n+1} = 1.1\Delta t_n$ subject to the (commonly accepted) maximum value $\Delta t_{n+1} \leq 10^{-3}$ My. If Newton's method fails after 10 iterations, the time step is rejected and we restart the iteration with $\Delta t_n := 0.5\Delta t_n$.

In the test case with no erosion constraint (Fig. 2), we observe that after 1 My the diffusion has notably smoothed the global structure of the mountains, especially in steep areas. We can also distinguish the shoreline between the continental and the marine domains. This is due to the contrast between the diffusion coefficients K_c and K_m. In Table 1, we summarize some numerical data associated with the simulation. We can see that this test case is an "easy" one as no time steps were refused. The mean number of required Newton iterations is rather low, as well as the mean number of solver iterations. However, larger values of the diffusion coefficients may cause more severe difficulties, implying much smaller time steps.

In the second test case (Fig. 3) where the erosion constraint enters into play, we observe a different behavior. After 1 My, the mountains underwent less erosion and their structure is still recognizable. We can visualize the areas where the constraint is effective by looking at the values of the flux limiter λ in Fig. 4. We can notice

Table 1 Numerical data for the case without constraint

Accepted time steps	1016
Refused time steps	0
Mean Newton iterations per accepted time step	1.98
Mean solver iterations per Newton iteration	1.99
Computing time (s)	783

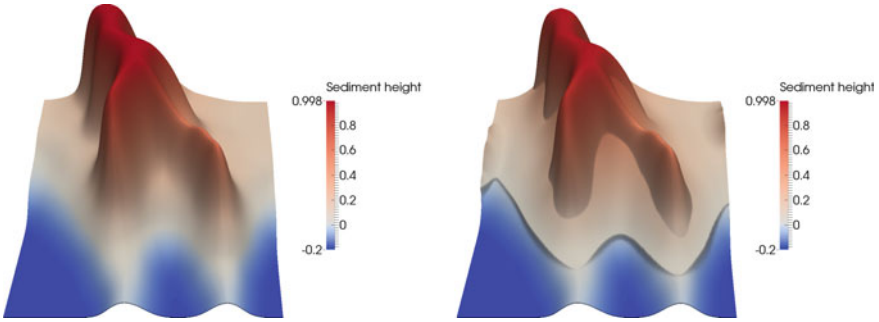


Fig. 3 Case with constraint: initial state (*left*) and final state (*right*) of h (km)

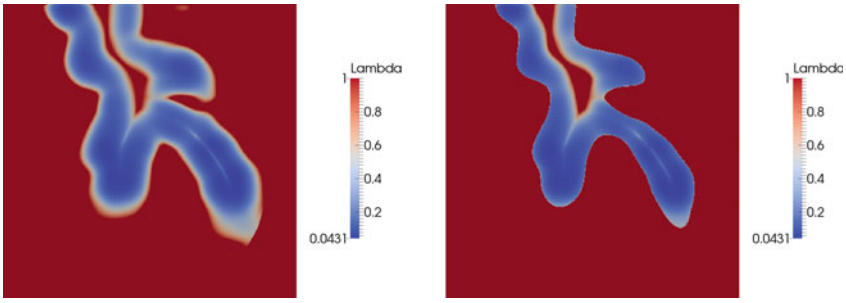


Fig. 4 Case with constraint: initial state (*left*) and final state (*right*) of λ

that λ takes values close to zero in the areas corresponding to the mountain's flanks, where the diffusion is the most important. The shoreline between the continental and marine domains is still present, for the same reason as before. By looking at Table 2 we also notice some differences. With the same management of the time steps as in the previous case, Newton's method sometimes fails to converge and some time steps are rejected. The model is more difficult to solve numerically: the mean number of Newton iterations and mean number of solver iterations are higher than in the case without constraint. This accounts for the rise in the computing time between the two

Table 2 Numerical results for the case with constraint

Accepted time steps	1235
Refused time steps	95
Mean Newton iterations per accepted time step	2.29
Mean solver iterations per Newton iteration	12.14
Computing time (s)	1809

simulations. Furthermore, it has been observed that the cases without constraint and with a very strong constraint were relatively easy to compute. The difficulties are most serious for "intermediate" values of the maximum erosion rate E.

4 Conclusion

This extension of the model [3] to a p-Laplacian diffusion law is the first step of a broader program whose objective is to enrich the physics of the industrial simulator Dionisos Flow^{TM}, developed by IFP Energies nouvelles. The next steps include usual features such as multi-lithology and variable bathymetry, but also a coupling of the sediment flow with water effects such as rains and rivers.

References

1. Andreianov, B., Boyer, F., Hubert, F.: Finite volume schemes for the p-Laplacian on Cartesian meshes. M2AN. Math. Model. Numer. Anal. **38**(6), 931–959 (2004)
2. Eymard, R., Gallouët, T., Ghilani, M., Herbin, R.: Error estimates for the approximate solutions of a nonlinear hyperbolic equation given by finite volume schemes. IMA J. Numer. Anal. **18**(4), 563–594 (1998)
3. Eymard, R., Gallouët, T., Granjeon, D., Masson, R., Tran, Q.H.: Multi-lithology stratigraphic model under maximum erosion rate constraint. Int. J. Numer. Methods Eng. **60**(2), 527–548 (2004)
4. Gervais, V.: Étude et simulation d'un modèle stratigraphique multi-lithologique sous contrainte de taux d'érosion maximal. Ph.D. thesis, Université de Provence, Aix-Marseille I (2004)
5. Gervais, V., Masson, R.: Numerical simulation of a stratigraphic model. Comput. Geosci. **12**(2), 163–179 (2008)

Comparison of Adaptive Non-symmetric and Three-Field FVM-BEM Coupling

Christoph Erath and Robert Schorr

Abstract The prototype for flow and transport in porous media in an interior domain is coupled to the Laplace equation on the complement, an unbounded domain. We approximate the solution of this interface problem either by the *non-symmetric* or the *three-field* coupling of the Finite Volume Method (FVM) and the Boundary Element Method (BEM). For these two coupling methods we introduce (semi-) robust a posteriori error estimators and use them in an adaptive algorithm to improve the convergence. Numerical experiments compare these two adaptive methods in terms of effectivity index, errors and mesh refinement.

Keywords Finite volume method · Boundary element method · Non-symmetric coupling · Three-field coupling · Robust a posteriori error estimates · Adaptive mesh refinement

1 Introduction and Model Problem

The finite volume method (FVM) is the method of choice for problems coming from fluid mechanics applications because of its direct flux conservation and the possibility to solve convection dominated problems via a simple upwind stabilization. When such a flow problem is coupled with a problem on an unbounded domain (e.g., to replace unknown boundary conditions) it is useful to reduce the exterior problem to a problem on the boundary. This leads to a formulation as an integral equation and its discretization to the boundary element method (BEM). There are several possibilities to couple FVM with BEM, in this work we compare the adaptive non-symmetric [3, 4] and the adaptive three-field FVM-BEM coupling approach [1, 2]. Both cou-

C. Erath
TU Darmstadt, Department of Mathematics, Dolivostraße 15, 64293 Darmstadt, Germany
e-mail: erath@mathematik.tu-darmstadt.de

R. Schorr (✉)
TU Darmstadt, Department of Mathematics/GSC CE, Dolivostraße 15,
64293 Darmstadt, Germany
e-mail: schorr@gsc.tu-darmstadt.de

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_36

plings have been analyzed for 2D and 3D cases. For simplicity we only consider the 2D case here.

Let $\Omega \subset \mathbb{R}^2$ be a bounded domain with connected polygonal Lipschitz boundary Γ with $\operatorname{diam}(\Omega) < 1$ (possible by scaling) to ensure $H^{-1/2}(\Gamma)$ ellipticity of the single layer operator defined below. The corresponding unbounded exterior domain is $\Omega_e = \mathbb{R}^2 \backslash \overline{\Omega}$. The coupling boundary $\Gamma = \partial\Omega = \partial\Omega_e$ is divided in an inflow and outflow part, namely $\Gamma^{in} := \left\{ x \in \Gamma \,\middle|\, \mathbf{b}(x) \cdot \mathbf{n}(x) < 0 \right\}$ and $\Gamma^{out} := \left\{ x \in \Gamma \,\middle|\, \mathbf{b}(x) \cdot \mathbf{n}(x) \geq 0 \right\}$, respectively, where $\mathbf{n}$ is the normal vector on Γ pointing outward with respect to Ω. Then the model problem reads, (see also [1, 3]): Find $u \in H^1(\Omega)$ and $u_e \in H^1_{\ell oc}(\Omega_e)$ such that

$$\operatorname{div}(-\mathbf{A}\nabla u + \mathbf{b}u) + cu = f \qquad \text{in } \Omega, \tag{1a}$$

$$-\Delta u_e = 0 \qquad \text{in } \Omega_e, \tag{1b}$$

$$u_e(x) = C_\infty \log|x| + \mathcal{O}(1/|x|) \qquad \text{for } |x| \to \infty, \tag{1c}$$

$$u = u_e + u_0 \qquad \text{on } \Gamma, \tag{1d}$$

$$(\mathbf{A}\nabla u - \mathbf{b}u) \cdot \mathbf{n} = \frac{\partial u_e}{\partial \mathbf{n}} + t_0 \qquad \text{on } \Gamma^{in}, \tag{1e}$$

$$(\mathbf{A}\nabla u) \cdot \mathbf{n} = \frac{\partial u_e}{\partial \mathbf{n}} + t_0 \qquad \text{on } \Gamma^{out}. \tag{1f}$$

Here, $L^m(\cdot)$ and $H^m(\cdot)$, $m > 0$ denote the standard Lebesgue and Sobolev spaces equipped with the corresponding norms $\|\cdot\|_{L^m(\cdot)}$ and $\|\cdot\|_{H^m(\cdot)}$. We will use $(\cdot,\cdot)_\omega$ for the L^2 scalar product for $\omega \subset \Omega$. The duality between $H^m(\Gamma)$ and $H^{-m}(\Gamma)$ is given by the extended L^2-scalar product $\langle\cdot,\cdot\rangle_\Gamma$. We collect all functions with local H^1 behavior in $H^1_{\ell oc}(\Omega)$ and the Lipschitz continuous functions in $W^{1,\infty}$.

The diffusion matrix $\mathbf{A} : \Omega \to \mathbb{R}^{2\times 2}$ has entries in $W^{1,\infty}(T)$ for every $T \in \mathcal{T}$, where $\mathcal{T}$ is a mesh of Ω introduced below in Sect. 2. Additionally, $\mathbf{A}$ is bounded, symmetric and uniformly positive definite. Furthermore, $\mathbf{b} \in W^{1,\infty}(\Omega)^2$ and $c \in L^\infty(\Omega)$ satisfy the coerciveness assumption $(\operatorname{div}\mathbf{b}(x))/2 + c(x) \geq 0$ for almost every $x \in \Omega$. For the a posteriori estimators we assume slightly more regularity on the data than usual; $f \in L^2(\Omega)$, $u_0 \in H^1(\Gamma)$ and $t_0 \in L^2(\Gamma)$. The constant C_∞ is unknown; see [1, 3] for possible different radiation conditions. Note that we can rewrite the exterior problem (1b)–(1c) with the aid of the Calderón system and the Cauchy data $\xi := u_e|_\Gamma \in H^{1/2}(\Gamma)$ and $\phi := \partial u_e/\partial\mathbf{n}|_\Gamma \in H^{-1/2}(\Gamma)$ into an equivalent integral equation. The model problem and the weak form are equivalent. There exists a unique weak solution $(u, u_e) \in H^1(\Omega) \times H^1_{\ell oc}(\Omega)$; see [1, 3].

2 Non-symmetric and Three-Field FVM-BEM Coupling

This section introduces two different types of FVM-BEM couplings. In order to do this we first fix some notation.

Triangulations and discrete function spaces: With $\mathcal{T}$ we denote a regular triangulation of Ω which consists of non-degenerate closed triangles. We assume that

$\mathscr{T}$ is *shape-regular*, i.e., $\max_{T\in\mathscr{T}} h_T^2/|T| \le \sigma < \infty$ with $h_T := \sup_{x,y\in T} |x-y|$ and that (possible) discontinuities of the known data $\mathbf{A}$, $\mathbf{b}$, c, f, u_0, and t_0 are aligned with $\mathscr{T}$. Then the sets $\mathscr{N}$ and $\mathscr{E}$ are the nodes and edges of $\mathscr{T}$, respectively. We denote by $\mathscr{E}_T \subset \mathscr{E}$ the set of all edges of T, i.e., $\mathscr{E}_T := \{E \in \mathscr{E} \mid E \subset \partial T\}$, by $\mathscr{E}_\Gamma := \{E \in \mathscr{E} \mid E \subset \Gamma\}$ the set of all edges on the boundary Γ, and by $\mathscr{E}_I = \mathscr{E}\backslash\mathscr{E}_\Gamma$ all interior edges. Furthermore, h_E is the length of an edge $E \in \mathscr{E}$ and the unit normal vector $\mathbf{n}$ on a boundary always points outwards with respect to the domain.

For a vertex-centered FVM formulation we need a dual mesh $\mathscr{T}^*$, which can be constructed from the primal mesh $\mathscr{T}$. The so-called control volumes $V \in \mathscr{T}^*$ are constructed by connecting the center of gravity of an element $T \in \mathscr{T}$ with the midpoint of the edges $E \in \mathscr{E}_T$; see [3, Fig. 1]. Note that for every vertex $a_i \in \mathscr{N}$ ($i = 1 \ldots \#\mathscr{N}$), we can assign a unique box $V_i \in \mathscr{T}^*$ only containing a_i.

Finally, with $\mathscr{S}^1(\mathscr{T})$ we define the piecewise affine and globally continuous function space on $\mathscr{T}$ and $\mathscr{S}^1_*(\mathscr{E}_\Gamma)$ is $\mathscr{S}^1(\mathscr{E}_\Gamma)$ ($\mathscr{S}^1$ on $\mathscr{E}_\Gamma$) with integral mean zero. We denote by $\mathscr{P}^0(\mathscr{E}_\Gamma)$ and $\mathscr{P}^0(\mathscr{T}^*)$ the $\mathscr{E}_\Gamma$-piecewise and $\mathscr{T}^*$-piecewise constant function spaces. For $v^* \in \mathscr{P}^0(\mathscr{T}^*)$ we may use $v^* := \sum_{a_i\in\mathscr{N}} v_i^* \chi_i^*$, $v_i^* \in \mathbb{R}$, where χ_i^* is the characteristic function of $V_i \in \mathscr{T}^*$.

Non-symmetric FVM-BEM coupling: Now we can introduce the non-symmetric FVM-BEM coupling method which reads: Find $u_h \in \mathscr{S}^1(\mathscr{T})$ and $\phi_h \in \mathscr{P}^0(\mathscr{E}_\Gamma)$ such that

$$\begin{aligned} \mathscr{A}_V(u_h, v^*) - \langle \phi_h, v^*\rangle_\Gamma &= (f, v^*)_\Omega + \langle t_0, v^*\rangle_\Gamma, \\ \langle (1/2 - \mathscr{K})u_h, \psi_h\rangle_\Gamma + \langle \mathscr{V}\phi_h, \psi_h\rangle_\Gamma &= \langle (1/2-\mathscr{K})u_0, \psi_h\rangle_\Gamma \end{aligned} \tag{2}$$

for all $v^* \in \mathscr{P}^0(\mathscr{T}^*)$, $\psi_h \in \mathscr{P}^0(\mathscr{E}_\Gamma)$ with the finite volume bilinear form

$$\begin{aligned} \mathscr{A}_V(u_h, v^*) := \sum_{a_i\in\mathscr{N}} v_i^* \Bigg(&\int_{\partial V_i\backslash\Gamma} (-\mathbf{A}\nabla u_h + \mathbf{b}u_h)\cdot\mathbf{n}\, ds \\ &+ \int_{V_i} c u_h\, dx + \int_{\partial V_i\cap\Gamma^{out}} \mathbf{b}\cdot\mathbf{n}\, u_h\, ds \Bigg), \end{aligned} \tag{3}$$

the single layer operator $(\mathscr{V}\phi_h)(x) := -\frac{1}{2\pi}\int_\Gamma \phi_h(y)\log|x-y|\, ds_y$, and the double layer operator $(\mathscr{K}u_h)(x) := -\frac{1}{2\pi}\int_\Gamma u_h(y)\frac{\partial}{\partial\mathbf{n}_y}\log|x-y|\, ds_y$, $x \in \Gamma$.

The system (2) approximates u by u_h and the conormal ϕ by ϕ_h. However, for convection dominated problems the central approximation of the convection term can lead to strong oscillations in the FVM solution. Since FVM is based on the balance equation we can easily apply a full upwinding stabilization which avoids these oscillations but still preserves local flux conservation: Given $V_i \in \mathscr{T}^*$, we consider the intersections $\tau_{ij} = V_i \cap V_j \neq \emptyset$ with the neighboring boxes $V_j \in \mathscr{T}^*$; see also [3, Fig. 1]. Then we replace $\mathbf{b}u_h$ on interior dual edges $\partial V_i\backslash\Gamma$ in $\mathscr{A}_V$ (3) by an upwind approximation. Instead of u_h on τ_{ij} we use $u_{h,ij} := u_h(a_i)$ if $\frac{1}{|\tau_{ij}|}\int_{\tau_{ij}} \mathbf{b}\cdot\mathbf{n}_i\, ds \ge 0$, otherwise $u_{h,ij} := u_h(a_j)$. Here, $\mathbf{n}_i$ points outwards with respect to V_i.

The stability and convergence analysis (also with the upwind option) [3, Theorem 2 and 3] holds under a minimal eigenvalue condition on $\mathbf{A}$ (constraint from the ellipticity of the non-symmetric variational form [3, Theorem 1]). With the usual regularity assumptions this scheme leads to first order convergence.

Three-field FVM-BEM coupling: The three-field coupling uses a different formulation of the exterior problem (i.e., the full Calderón system) and reads: Find $u_h \in \mathscr{S}^1(\mathscr{T})$, $\xi_h \in \mathscr{S}^1_*(\mathscr{E}_\Gamma)$ and $\phi_h \in \mathscr{P}^0(\mathscr{E}_\Gamma)$ such that

$$\begin{aligned} \mathscr{A}_V(u_h, v^*) - \langle \phi_h, v^* \rangle_\Gamma &= (f, v^*)_\Omega + \langle t_0, v^* \rangle_\Gamma, \\ -\langle u_h, \psi_h \rangle_\Gamma - \langle \mathscr{V} \phi_h, \psi_h \rangle_\Gamma + \langle (1/2 + \mathscr{K}) \xi_h, \psi_h \rangle_\Gamma &= -\langle u_0, \psi_h \rangle_\Gamma, \\ \langle (1/2 + \mathscr{K}^*) \phi_h, \theta_h \rangle_\Gamma + \langle \mathscr{W} \xi_h, \theta_h \rangle_\Gamma &= 0 \end{aligned} \tag{4}$$

for all $v^* \in \mathscr{P}^0(\mathscr{T}^*)$, $\theta_h \in \mathscr{S}^1_*(\mathscr{E}_\Gamma)$, $\psi_h \in \mathscr{P}^0(\mathscr{E}_\Gamma)$. Here, we additionally use the adjoint double layer operator $(\mathscr{K}^* \phi_h)(x) := -\frac{1}{2\pi} \int_\Gamma \phi_h(y) \frac{\partial}{\partial \mathbf{n}_x} \log|x-y| \, ds_y$ and the hypersingular integral operator $(\mathscr{W} \xi_h)(x) := \frac{1}{2\pi} \frac{\partial}{\partial \mathbf{n}_x} \int_\Gamma \xi_h(y) \frac{\partial}{\partial \mathbf{n}_y} \log|x-y| \, ds_y$, $x \in \Gamma$. Note that the system (4) additionally approximates the trace ξ by ξ_h and that the upwind option in $\mathscr{A}_V$ described above applies here as well. An a priori convergence analysis (also with the upwind option but without the eigenvalue restriction) can be found in [1]. With the usual regularity assumptions this scheme leads to first order convergence as well. Although the three-field coupling leads to a larger system of linear equations than the non-symmetric coupling one should apply it if the trace ξ_h is explicitly important or if the right-hand side contribution $\mathscr{K} u_0$ is difficult to evaluate.

3 Residual Based a Posteriori Error Estimator

In order to introduce an element-wise refinement indicator, which is a part of our a posteriori error estimator, we define the residual $R := R(u_h) = f - \operatorname{div}(-\mathbf{A}\nabla u_h + \mathbf{b} u_h) - c u_h$ on $T \in \mathscr{T}$ and an edge-residual or jump $J : L^2(\mathscr{E}) \to \mathbb{R}$ by

$$J|_E := J(u_h)|_E = \begin{cases} \left[(-\mathbf{A}\nabla u_h)|_{E,T} - (-\mathbf{A}\nabla u_h)|_{E,T'} \right] \cdot \mathbf{n} & \text{for all } E \in \mathscr{E}_I, \\ (-\mathbf{A}\nabla u_h + \mathbf{b} u_h) \cdot \mathbf{n} + \phi_h + t_0 & \text{for all } E \in \mathscr{E}_\Gamma^{in}, \\ -\mathbf{A}\nabla u_h \cdot \mathbf{n} + \phi_h + t_0 & \text{for all } E \in \mathscr{E}_\Gamma^{out}. \end{cases}$$

with $E = T \cap T' \in \mathscr{E}_I$, $T, T' \in \mathscr{T}$. Note that $\varphi|_{E,T}$ denotes the trace of $\varphi \in H^1(T)$ on E and the normal vector $\mathbf{n}$ points from T to T'.

To prove a robust upper bound of the energy error we need some further notation. In order to apply a robust interpolant, the diffusion distribution in Ω has to be quasi-monotone; for a definition we refer to [2, 4]. To simplify notation we restrict ourselves here to a piecewise constant diffusion coefficient $\alpha \in \mathscr{P}^0(\mathscr{T})$ with $\mathbf{A} = \alpha \mathbf{I}$. For the $\mathscr{T}$-piecewise constant function $\alpha \in \mathscr{P}^0(\mathscr{T})$ we write $\alpha_T := \alpha|_T$ for

all $T \in \mathscr{T}$. Furthermore, we define $\alpha_E := \max\{\alpha_{T_1}, \alpha_{T_2}\}$ for $E \in \mathscr{E}_I$ with $E \subset T_1 \cap T_2$, $\alpha_E := \alpha_T$ for $E \in \mathscr{E}_\Gamma$ with $E \subset \partial T$. For convection and reaction we define $\beta_T := \min_{x\in T}\{(\operatorname{div}\mathbf{b}(x))/2 + c(x)\}$ for all $T \in \mathscr{T}$, $\beta_E := \min\{\beta_{T_1}, \beta_{T_2}\}$ for $E \in \mathscr{E}_I$ with $E \subset T_1 \cap T_2$ and $\beta_E := \beta_T$ for $E \in \mathscr{E}_\Gamma$ with $E \subset \partial T$. Next, we define $\mu_T := \min\{\beta_T^{-1/2}, h_T\alpha_T^{-1/2}\}$ and $\mu_E := \min\{\beta_E^{-1/2}, h_E\alpha_E^{-1/2}\}$ for all $T \in \mathscr{T}$ and all $E \in \mathscr{E}$, respectively. Note that we take the second argument if $\beta_T = 0$ or $\beta_E = 0$.

Then, the semi-robust refinement indicator for the *non-symmetric coupling* reads for all $T \in \mathscr{T}$

$$\eta_T^2 := \mu_T^2 \|R\|_{L^2(T)}^2 + \frac{1}{2}\sum_{E\in\mathscr{E}_I\cap\mathscr{E}_T} \alpha_E^{-1/2}\mu_E\|J\|_{L^2(E)}^2 + \sum_{E\in\mathscr{E}_\Gamma\cap\mathscr{E}_T} \alpha_E^{-1/2}\mu_E\|J\|_{L^2(E)}^2 + \sum_{E\in\mathscr{E}_\Gamma\cap\mathscr{E}_T} h_E\|\partial/\partial s\big((1/2-\mathscr{K})(u_0-u_h) - \mathscr{V}\phi_h\big)\|_{L^2(E)}^2, \tag{5}$$

where $\partial/\partial s$ denotes the arc length derivative. For the *three-field coupling* the semi-robust refinement indicator differs slightly, since the exterior trace is approximated separately. Hence, for all $T \in \mathscr{T}$ we get

$$\begin{aligned}\eta_T^2 := \mu_T^2 \|R\|_{L^2(T)}^2 &+ \frac{1}{2}\sum_{E\in\mathscr{E}_I\cap\mathscr{E}_T} \alpha_E^{-1/2}\mu_E\|J\|_{L^2(E)}^2 + \sum_{E\in\mathscr{E}_\Gamma\cap\mathscr{E}_T} \alpha_E^{-1/2}\mu_E\|J\|_{L^2(E)}^2 \\ &+ \sum_{E\in\mathscr{E}_\Gamma\cap\mathscr{E}_T} h_E\|\partial u_h/\partial s - \partial/\partial s\big(u_0 - \mathscr{V}\phi_h + (1/2+\mathscr{K})\xi_h\big)\|_{L^2(E)}^2 \\ &+ \sum_{E\in\mathscr{E}_\Gamma\cap\mathscr{E}_T} h_E\|\mathscr{W}\xi_h + (1/2+\mathscr{K}^*)\phi_h\|_{L^2(E)}^2\end{aligned} \tag{6}$$

If we apply the *upwind stabilization* option, an additional refinement quantity is necessary. For both coupling systems this reads for all $T \in \mathscr{T}$

$$\eta_{T,up}^2 := \alpha_T^{-1/2}\mu_T \sum_{\tau_{ij}^T\in\mathscr{D}^T} \|\mathbf{b}\cdot\mathbf{n}_i(u_h - u_{h,ij})\|_{L^2(\tau_{ij}^T)}^2 \tag{7}$$

with $\mathscr{D}^T := \big\{\tau_{ij}^T \big| \tau_{ij}^T = V_i \cap V_j \cap T$ for $V_i, V_j \in \mathscr{T}^*, V_i \neq V_j, V_i \cap T \neq \emptyset, V_j \cap T \neq \emptyset\big\}$ and the upwind value $u_{h,ij}$ from Sect. 2. With the refinement indicators (5) and (6) (plus (7)) we can define an error estimator

$$\eta := \Big(\sum_{T\in\mathscr{T}} \eta_T^2 (+\eta_{T,up}^2)\Big)^{1/2} \tag{8}$$

for the non-symmetric (2) and the three-field (4) FVM-BEM coupling. For both couplings η is reliable and efficient with respect to the error in the energy norm

$$E_h := \begin{cases} |\!|\!| u - u_h |\!|\!|_\Omega + \|\phi - \phi_h\|_{\mathscr{V}} & \text{non-symmetric,} \\ |\!|\!| u - u_h |\!|\!|_\Omega + \|\phi - \phi_h\|_{\mathscr{V}} + \|\xi - \xi_h\|_{\mathscr{W}} & \text{three-field} \end{cases} \tag{9}$$

with $|\!|\!| v |\!|\!|_\Omega := \|\mathbf{A}^{1/2}\nabla v\|^2_{L^2(\Omega)} + \|\,(\operatorname{div}\mathbf{b}/2 + c)^{1/2}\, v\|^2_{L^2(\Omega)}$, $\|\cdot\|^2_{\mathscr{V}} := \langle \mathscr{V}\cdot, \cdot\rangle_\Gamma$ and $\|\cdot\|^2_{\mathscr{W}} := \langle \mathscr{W}\cdot, \cdot\rangle_\Gamma$. For both couplings the upper bound (reliability) is *robust* with respect to the variation of the model data. For the non-symmetric coupling, however, we have a minimal eigenvalue restriction of the diffusion matrix $\mathbf{A}$ again. The analytical proof for the lower bound (efficiency) in both couplings holds only for a quasi-uniform mesh on the boundary Γ. An improved efficiency result in slightly stronger norms (but for a shape regular triangulation also on the boundary) has recently been published in [4]. Additionally, the constant in the lower bound is only *semi-robust*, i.e., it depends on the local Péclet number. For more details and discussions on the bounds we refer to [4] (non-symmetric) and [2] (three-field).

4 Numerical Experiments

With the refinement indicators (5) and (6) (plus (7)) we devise an adaptive algorithm with the well known Dörfler marking strategy, where we consider a sequence $\mathscr{T}^{(k)}$, $k = 0, 1, \ldots$ of triangulations: Throughout, let $\theta = 0.5$, then at refinement step k choose $\mathscr{M}^{(k)} \subset \mathscr{T}^{(k)}$ with minimal cardinality such that

$$\sum_{T\in\mathscr{M}^{(k)}} \left(\eta_T^2 (+\eta_{T,up}^2)\right) \geq \theta \sum_{T\in\mathscr{T}^{(k)}} \left(\eta_T^2 (+\eta_{T,up}^2)\right).$$

Then refine the elements in the set $\mathscr{M}^{(k)}$ with a red-green-blue refinement which ensures the shape regularity of the new mesh $\mathscr{T}^{(k+1)}$.

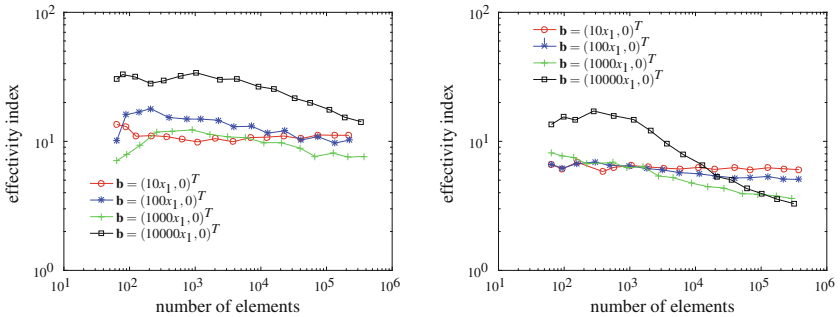
4.1 Convection-Diffusion Problem

For our first problem we choose $\Omega = (0, 1/2) \times (0, 1/2)$ and prescribe the solution in the interior to be $u(x_1, x_2) = 0.5\left(1 - \tanh\left(\frac{0.25-x_1}{0.02}\right)\right)$ for $x = (x_1, x_2) \in \Omega$, and the solution in the exterior domain Ω_e to be $u_e(x_1, x_2) = \log\sqrt{(x_1 - 0.25)^2 + (x_2 - 0.25)^2}$. We choose the jumping diffusion coefficient as $\alpha = 0.42$ for $x_2 < 0.25$ and 10 for $x_2 \geq 0.25$, the convection field $\mathbf{b} = (1000x_1, 0)^T$ and the reaction coefficient $c = 0$. Since this is a convection dominated problem we will use the full upwind stabilization. The right-hand side f and the jumps are calculated by means of the analytical solution.

Table 1 shows the contributions to the error in the energy norm (9) of both adaptive couplings. Note that in the non-symmetric case we compute ξ_h by $u_h|_\Gamma - u_0$ which is motivated by (1d). It can be observed that the error for the three-field coupling

Table 1 Errors for different refinement levels k for both coupling systems of the first example

k	Scheme	$\#\mathcal{T}$	$\|\|u-u_h\|\|_\Omega$	$\|\|\xi-\xi_h\|\|_{\mathscr{W}}$	$\|\|\phi-\phi_h\|\|_{\mathscr{V}}$	$\|\|u-u_h\|\|_{L^2(\Omega)}$
8	Non-symmetric	5834	$4.48e-01$	$6.11e-02$	$4.79e-02$	$6.62e-03$
	Three-field	4542	$4.60e-01$	$5.81e-02$	$4.86e-02$	$5.65e-03$
12	Non-symmetric	66959	$1.80e-01$	$1.95e-02$	$1.68e-02$	$2.32e-03$
	Three-field	52065	$1.76e-01$	$1.12e-02$	$1.07e-02$	$1.81e-03$
16	Non-symmetric	671921	$6.17e-02$	$4.40e-03$	$4.61e-03$	$7.61e-04$
	Three-field	534051	$5.84e-02$	$3.25e-03$	$3.11e-03$	$5.38e-04$



Non-symmetric coupling. Three-field coupling.

Fig. 1 The effectivity index η/E_h for the two coupling methods for the first example

is slightly better (less elements but smaller errors). In Fig. 1 we show the effectivity index η/E_h for $\mathbf{b} = \{(10x_1, 0)^T; (100x_1, 0)^T; (1000x_1, 0)^T; (10000x_1, 0)^T\}$. In both cases we observe the dependency on the local Péclet number, i.e., once we have resolved the shock region, the effectivity index convergences as well.

4.2 A More Practical Example

For the second example we do not know an analytical solution of (1). Additionally, we replace the radiation condition (1c) by $u_e(x) = a_\infty + \mathcal{O}(1/|x|)$ for $|x| \to \infty$. Thus we have to assume the scaling condition $\langle \partial u_e/\partial \mathbf{n}, 1\rangle_\Gamma = 0$; see [2] and have to modify our discretization. The domain will be the classical L-shaped domain $\Omega = (-1/4, 1/4)^2 \setminus [0, 1/4] \times [-1/4, 0]$. We fix the piecewise constant diffusion coefficient α to 1 for $x_1 > 0$, 0.1 for $x_2 \leq 0$ and 0.5 else, $\mathbf{b} = (1500, 1000)^T$, and $c = 0.01$. The right-hand side will be $f(x_1, x_2) = 5$ for $0.2 \leq x_1 \leq -0.1$, $-0.2 \leq x_2 \leq -0.05$ and 0 else and the jumps t_0 and u_0 are set to zero. This problem is again convection dominated, therefore, we use the full upwind stabilization. In Fig. 2 two adaptively generated meshes and contour lines are plotted to show the similarities between the two coupling approaches. Both meshes refine along the steepest parts of

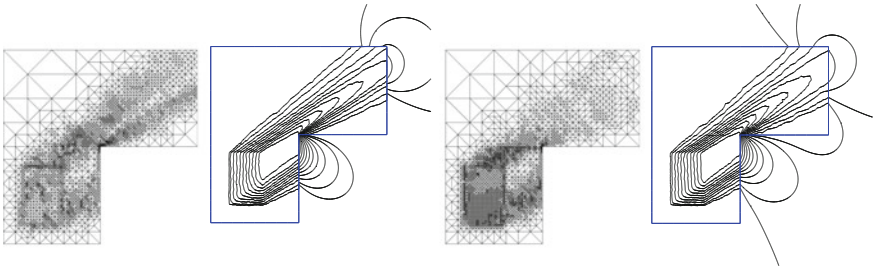


Fig. 2 Adaptively generated mesh and contour lines for the non-symmetric FVM-BEM (*left*) and three-field FVM-BEM (*right*) for the second example

the solution, but they localize at slightly different areas. To generate the contour lines we calculate the values in Ω_e from the Cauchy data ξ_h and ϕ_h and the *representation formula*; see [1, 3]. Therefore, the contour lines show also the flow into the unbounded domain and show the difference in the accuracy of the approximation of the exterior solution.

5 Conclusions

We presented the adaptive non-symmetric and the adaptive three-field FVM-BEM coupling. For both methods we established an error estimator which is reliable and efficient. In contrast to the three-field coupling the upper bound for the non-symmetric coupling imposes a lower bound on the smallest eigenvalue of the diffusion matrix which seems to be only a theoretical constraint. The effectivity index for both methods is semi-robust against variation of the model data. The three-field coupling leads to slightly better results than the non-symmetric coupling with respect to the same number of elements. However, the three-field coupling is computationally more expensive since it approximates the exterior trace directly. On the other hand the input data does not appear in an integral operator.

Acknowledgements The work of the second author is supported by the *Excellence Initiative* of the German Federal and State Governments and the *Graduate School of Computational Engineering* at Technische Universität Darmstadt.

References

1. Erath, C.: Coupling of the finite volume element method and the boundary element method: an a priori convergence result. SIAM J. Numer. Anal. **50**(2), 574–594 (2012). doi:10.1137/110833944
2. Erath, C.: A posteriori error estimates and adaptive mesh refinement for the coupling of the finite volume method and the boundary element method. SIAM J. Numer. Anal. **51**(3), 1777–1804 (2013). doi:10.1137/110854771
3. Erath, C., Of, G., Sayas, F.J.: A non-symmetric coupling of the finite volume method and the boundary element method. Numer. Math. **135**(3), 895–922 (2017). doi:10.1007/s00211-016-0820-3
4. Erath, C., Schorr, R.: An adaptive non-symmetric finite volume and boundary element coupling method for a fluid mechanics interface problem. in press, SIAM J. Sci. Comput. (2017)

On the Conditions for Coupling Free Flow and Porous-Medium Flow in a Finite Volume Framework

Thomas Fetzer, Christoph Grüninger, Bernd Flemisch and Rainer Helmig

Abstract This article presents model concepts for the coupling of one-phase compositional non-isothermal Navier-Stokes flow to two-phase compositional non-isothermal Darcy flow in a finite volume framework. The focus of the presented coupling conditions is on defining appropriate conditions for momentum transfer without introducing additional degrees of freedom at the interface. Four different methods are presented and compared with the help of numerical simulations of flow around an evaporating porous medium. The results show that simply assigning the porous medium gas pressure as the gas pressure at the interface (CM1) leads to high, non-physical velocities in cells at the corner of the porous medium. This effect can be weakened by recalculating the interface gas pressure with the help of the total mass balance and additional assumptions concerning the state at the interface (CM2). Allowing only momentum transfer between the gas phases (CM3) leads to an increase of the resistance against inflow, if the porous medium is filled with water. However, in order to minimize the assumptions made, an additional system of equations can be introduced and solved to recalculate the pressure at the interface (CM4). This method is computationally more expensive but shows the expected physical behavior regarding the velocity profile.

Keywords Free flow · Porous-medium flow · Coupling

1 Introduction

Modeling the exchange processes between a free flow and flow in a porous medium is important in a variety of applications, ranging from salinization of agricultural land [13], flow through oil filters [12], rocket cooling [4], material science [5], to nuclear waste storage [15]. The same variety is found when looking at the used

The original version of the book was revised: Missed out corrections have been updated. The erratum to the book is available at https://doi.org/10.1007/978-3-319-57394-6_58

T. Fetzer (✉) · C. Grüninger · B. Flemisch · R. Helmig
Universität Stuttgart, Pfaffenwaldring 61, 70569 Stuttgart, Germany
e-mail: Thomas.Fetzer@iws.uni-stuttgart.de

C. Cancés and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_37

solution and modeling techniques. Decoupling strategies in space [4, 6, 15] and time [19] are utilized as well as fully-implicit monolithic approaches [1, 17]. Several different finite volume methods (FVM), using collocated grids [1, 17], pure staggered grid / marker-and-cell (MAC) schemes [12], or coupling MAC and cell-centered approaches [15, 19] can be found, some are used in combination with finite element methods [4]. Problems arise if the discretization does not provide all degrees of freedom at the interface between the two subdomains. In this work, the sensitivity of normal momentum exchange is analyzed using the example of evaporation from a porous medium.

2 Free Flow and Porous-Medium Flow

For modeling evaporation from a porous medium, see [7, 18], non-isothermal flow and transport of two phases (α), liquid (l) and gas (g), have to be considered. The phases are composed of two components (κ), air (a) and water (w).

The free flow is modeled using the Navier-Stokes equations under the following assumptions: (i) one-phase gas flow, (ii) Newtonian fluid, and (iii) Fickian diffusion. The equations for the free flow (1)–(4) are summarized in Table 1. The free flow momentum balance (2) is discretized using finite volumes on a MAC scheme, cf. [10]. The other balance equations are discretized using cell-centered FVM. The primary variables gas pressure p_g, water vapor mass fraction X_g^w, and temperature T are located at the finite volume cell center, whereas the velocity components $v_{\mathrm{g},i}$ and their control volumes are shifted to the faces, see Fig. 1.

The flow inside the porous medium is modeled by Darcy's law and is based on the following assumptions: (i) gas and liquid as mobile phases, (ii) Newtonian fluids, (iii) creeping flow with $\mathrm{Re} < 1$, (iv) rigid solid phase, (v) Fickian diffusion, and (vi) local thermodynamic equilibrium. The porous medium balance equations (5), (7), and (8), are discretized using cell-centered FVM. The primary variables p_g, liquid phase saturation S_l, and T are located at the finite volume cell center, as shown in Fig. 1. If a cell has dried out, X_g^w replaces S_l as a primary variable. The velocities, which are secondary variables, are approximated at the face centers by finite differences.

3 Coupling Conditions

The conditions for coupling the different model concepts are based on the assumption of local thermodynamic equilibrium and the continuity of fluxes which are commonly used in literature, e.g., [14, 16, 17]. In the following, the coupling conditions are presented with focus on the exchange of normal momentum. This work is restricted to (i) axis-parallel grids and (ii) porous-medium free-flow interfaces which are composed of faces between cells in both subdomains. The discretization and implementation of the coupling conditions involves some common steps that

Table 1 Governing equations for the free flow, porous-medium flow, and coupling. Diffusive fluxes for each subdomain $\omega \in \{\text{ff},\text{pm}\}$ are defined as: $\mathbf{j}_\alpha^{\kappa,\omega} = -D_\alpha^\omega \rho_{\text{mol,g}} M^\kappa \nabla x_\alpha^\kappa$, with D_α^{pm} including dispersion caused by tortuous pathways. Note that (10) is zero for all momentum balances tangentially to the interface

	Type	Equation	
Free flow	Mass	$\frac{\partial \rho_\text{g}}{\partial t} + \nabla \cdot \left(\rho_\text{g} \mathbf{v}_\text{g}\right) = 0$	(1)
	Momentum	$\frac{\partial \left(\rho_\text{g} \mathbf{v}_\text{g}\right)}{\partial t} + \nabla \cdot \left(\rho_\text{g} \mathbf{v}_\text{g} \mathbf{v}_\text{g}^\intercal\right) - \nabla \cdot \left(\rho_\text{g} \nu_\text{g} \left(\nabla \mathbf{v}_\text{g} + \nabla \mathbf{v}_\text{g}^\intercal\right)\right) + \nabla \cdot \left(p_\text{g} \mathbf{I}\right) - \rho_\text{g} \mathbf{g} = 0$	(2)
	Water mass	$\frac{\partial \left(\rho_\text{g} X_\text{g}^\text{w}\right)}{\partial t} + \nabla \cdot \left(\rho_\text{g} X_\text{g}^\text{w} \mathbf{v}_\text{g}\right) + \nabla \cdot \mathbf{j}_\text{g}^{\text{w,ff}} = 0$	(3)
	Energy	$\frac{\partial \left(\rho_\text{g} u_\text{g}^\text{e}\right)}{\partial t} + \nabla \cdot \left(\rho_\text{g} h_\text{g} \mathbf{v}_\text{g}\right) + \sum_\kappa \nabla \cdot \left(h_\text{g}^\kappa \mathbf{j}_\text{g}^{\kappa,\text{ff}}\right) - \nabla \cdot \left(\lambda_\text{g}^\text{ff} \nabla T\right) = 0$	(4)
Porous medium	Mass	$\sum_\alpha \left(\phi \frac{\partial (\rho_\alpha S_\alpha)}{\partial t} + \nabla \cdot (\rho_\alpha \mathbf{v}_\alpha)\right) = 0$	(5)
	Momentum	$\mathbf{v}_\alpha = -\frac{k_{\text{r},\alpha} K}{\nu_\alpha \rho_\alpha} \left(\nabla p_\alpha - \rho_\alpha \mathbf{g}\right)$	(6)
	Water mass	$\sum_\alpha \left(\phi \frac{\partial (\rho_\alpha S_\alpha X_\alpha^\text{w})}{\partial t} + \nabla \cdot \left(\rho_\alpha X_\alpha^\text{w} \mathbf{v}_\alpha\right) + \nabla \cdot \mathbf{j}_\alpha^{\text{w,pm}}\right) = 0$	(7)
	Energy	$\sum_\alpha \left(\phi \frac{\partial (\rho_\alpha S_\alpha u_\alpha^\text{e})}{\partial} + \nabla \cdot (\rho_\alpha h_\alpha \mathbf{v}_\alpha)\right) + (1-\phi) \frac{\partial (\rho_\text{s} c_\text{s} T)}{\partial t} - \nabla \cdot \left(\lambda^{\text{pm}} \nabla T\right) = 0$	(8)
Coupling	Mass	$\left[\left(\rho_\text{g} \mathbf{v}_\text{g}\right) \cdot \mathbf{n}\right]^\text{ff} = -\left[\left(\rho_\text{g} \mathbf{v}_\text{g} + \rho_\text{l} \mathbf{v}_\text{l}\right) \cdot \mathbf{n}\right]^\text{pm}$	(9)
	Norm. mom.	$\left[\left(\rho_\text{g} \mathbf{v}_\text{g} \mathbf{v}_\text{g}^\intercal - \rho_\text{g} \nu_\text{g} \left(\nabla \mathbf{v}_\text{g} + \nabla \mathbf{v}_\text{g}^\intercal\right) + p_\text{g} \mathbf{I}\right) \mathbf{n}\right]^\text{ff} = -\left[\left(p_\text{g} \mathbf{I}\, \mathbf{n}\right)\right]^\text{pm}$	(10)
	Tang. mom.	$\left[\left(-\frac{\sqrt{K}}{\alpha_\text{BJ}} \left(\nabla \mathbf{v}_\text{g}\right) \mathbf{n}\right) \cdot \mathbf{t}_i\right]^\text{ff} = \left[\mathbf{v}_\text{g} \cdot \mathbf{t}_i\right]^\text{ff}$	(11)
	Water mass	$\left[\left(\rho_\text{g} X_\text{g}^\text{w} \mathbf{v}_\text{g} + \mathbf{j}_\text{g}^{\text{w,ff}}\right) \cdot \mathbf{n}\right]^\text{ff} = -\left[\sum_\alpha \left(\rho_\alpha X_\alpha^\text{w} \mathbf{v}_\alpha + \mathbf{j}_\alpha^{\text{w,pm}}\right) \cdot \mathbf{n}\right]^\text{pm}$	(12)
	Energy	$\left[\left(\rho_\text{g} h_\text{g} \mathbf{v}_\text{g} + \sum_\kappa h_\text{g}^\kappa \mathbf{j}_\text{g}^{\kappa,\text{ff}} - \lambda_\text{g}^\text{ff} \nabla T\right) \cdot \mathbf{n}\right]^\text{ff} = -\left[\left(\sum_\alpha \rho_\alpha h_\alpha \mathbf{v}_\alpha - \lambda^{\text{pm}} \nabla T\right) \cdot \mathbf{n}\right]^\text{pm}$	(13)

may later be loosened when different coupling methods are discussed. For more details on the implementation in the numerical framework, please refer to [9].

Finite volume schemes require continuity of fluxes at the interface. Here, the emphasis is on modeling the interface fluxes with the available information. This means that only one side of (9)–(13) has to be specified. Except for (10), this is the free-flow side, because the interface-normal free-flow velocity is located at the interface. Further, liquid fluxes inside the porous medium cannot be directly calculated. In addition, based on the assumption of local thermodynamic equilibrium at the interface, corresponding primary variables are assumed to be continuous across the interface: $X_\text{g}^{\text{w,ff,if}} = X_\text{g}^{\text{w,pm,if}}$ and $T^{\text{ff,if}} = T^{\text{pm,if}}$. Their gradients can be built and used to derive diffusive and conductive fluxes, see Fig. 1. Other quantities needed for

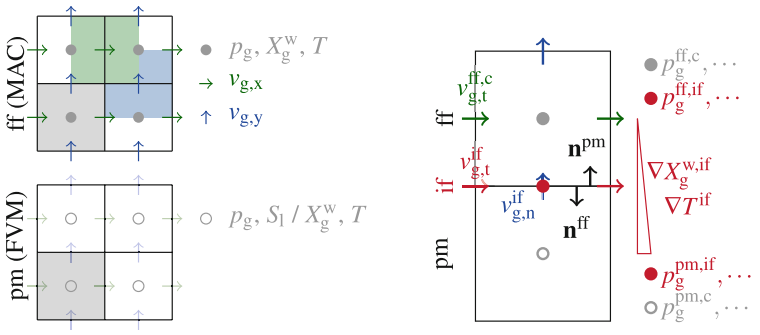


Fig. 1 Discretization schemes for the free flow (ff) and the porous medium (pm), *left*. Location of unknowns and pseudo-unknowns at the coupling interface (if), *right*

calculating the diffusive and conductive fluxes are taken from the cell center (c) of the respective subdomain, as their physical meaning is not identical in the two subdomains. For the advective transport an upwinding scheme is used, if the transported quantity occurs in both subdomains. Otherwise, the quantity at the cell center of the respective subdomain is used.

For each tangential momentum condition (11), cf. [2, 20], $v_{g,t_i} := \mathbf{v}_g \cdot \mathbf{t}_i$ can be separated, solved, and set as a Dirichlet condition. When coupling normal momentum (10), both pressures at the interface, $p_g^{ff,if}$ and $p_g^{pm,if}$, are unknown. They differ due to the different model concepts and the assumption of continuity of normal stresses. Determining $p_g^{pm,if}$ is necessary to couple the free-flow momentum balance equation with flow processes inside the porous medium. $p_g^{pm,if}$ will not directly be used to calculate fluxes inside the porous medium. In the next paragraphs, four different methods to determine $p_g^{pm,if}$ are presented.

Simple Momentum Coupling (***CM1***): This method assumes: $p_g^{pm,if} = p_g^{pm,c}$. Once the discretization is fine enough, this should be a good approximation of the gas pressure at the interface. For coarse grids or high fluxes, the pressure and thus the flow resistance, is over- or underestimated at the interface. Further, around corners of the porous medium no pressure drop can be predicted.

Total Momentum Coupling (***CM2***): The gas phase pressure at the interface is recalculated using (6) and (9):

$$\left[\rho_g \mathbf{v}_g^{if} \cdot \mathbf{n}\right]^{ff} = \left[\sum_\alpha \rho_\alpha \frac{k_{r,\alpha} K}{\rho_\alpha \nu_\alpha} \left(\frac{p_\alpha^{if} - p_\alpha^c}{\left(\mathbf{x}^{if} - \mathbf{x}^c\right) \cdot \mathbf{n}} - \rho_\alpha \mathbf{g} \cdot \mathbf{n} \right)\right]^{pm}. \tag{14}$$

In (14), the liquid saturation $S_l^{pm,if}$, and thus the liquid pressure at the interface $p_l^{pm,if}$, are unknown. Therefore, no difference in the liquid phase saturation is assumed: $S_l^{pm,if} = S_l^{pm,c}$. This simplification results in the same pressure difference for both phases and thus the same gradients contributing to the liquid and the gas phase flux. With this simplification, the interface gas pressure can be expressed

by:

$$p_{\mathrm{g}}^{\mathrm{pm,if}} = \frac{\left[\rho_{\mathrm{g}}\mathbf{v}_{\mathrm{g}}^{\mathrm{if}} \cdot \mathbf{n}\right]^{\mathrm{ff}} + \left[\sum_{\alpha} \frac{k_{\mathrm{r},\alpha} K}{\nu_{\alpha}} \rho_{\alpha} \mathbf{g} \cdot \mathbf{n}\right]^{\mathrm{pm}}}{\sum_{\alpha} \frac{k_{\mathrm{r},\alpha} K}{\nu_{\alpha}}} \left(\mathbf{x}^{\mathrm{pm,if}} - \mathbf{x}^{\mathrm{pm,c}}\right) \cdot \mathbf{n}^{\mathrm{pm}} + p_{\mathrm{g}}^{\mathrm{pm,c}} . \quad (15)$$

If gravitational forces are neglected, the porous medium and its discretization is given; then the lowest gas pressure and thus the highest gas velocities along the interface occur for small flow resistances $\left(\sum_{\alpha} k_{\mathrm{r},\alpha}/\nu_{\alpha}\right)^{-1}$ which are found at $S_{\mathrm{l}} = 1$, see Fig. 2.

Gas Momentum Coupling ***(CM3)***: In this approach, the liquid phase is not allowed to take up momentum from the free gas flow. This means the liquid terms in (14) and (15) are dropped and the gas pressure at the interface becomes:

$$p_{\mathrm{g}}^{\mathrm{pm,if}} = \frac{\left[\rho_{\mathrm{g}}\mathbf{v}_{\mathrm{g}}^{\mathrm{if}} \cdot \mathbf{n}\right]^{\mathrm{ff}} + \left[\frac{k_{\mathrm{r,g}} K}{\nu_{\mathrm{g}}} \rho_{\mathrm{g}} \mathbf{g} \cdot \mathbf{n}\right]^{\mathrm{pm}}}{\frac{k_{\mathrm{r,g}} K}{\nu_{\mathrm{g}}}} \left(\mathbf{x}^{\mathrm{pm,if}} - \mathbf{x}^{\mathrm{pm,c}}\right) \cdot \mathbf{n}^{\mathrm{pm}} + p_{\mathrm{g}}^{\mathrm{pm}} . \quad (16)$$

A fully liquid-saturated system acts, in contrast with CM2, as an impermeable barrier. No additional assumptions about the liquid state are required. Now, the lowest resistance $(k_{\mathrm{r,g}}/\nu_{\mathrm{g}})^{-1}$ and thus the highest velocities are found for $S_{\mathrm{l}} = 0$, see Fig. 2.

Coupling via an Interface Solver ***(CM4)***: The aim of this method is to make as few assumptions concerning interface conditions as possible. For each primary variable, a pseudo-unknown exists at the interface. These pseudo-unknowns are $p_{\mathrm{g}}^{\mathrm{pm,if}}$, $X_{\mathrm{g}}^{\mathrm{w,ff,if}}$, $S_{\mathrm{l}}^{\mathrm{pm,if}}$ or $X_{\mathrm{g}}^{\mathrm{w,pm,if}}$, $T^{\mathrm{ff,if}}$, and $T^{\mathrm{pm,if}}$. With the above-mentioned assumption of local thermodynamic equilibrium, two of them can be eliminated. The three necessary equations are (14), (12), and (13). In contrast to the assumptions required in CM2, assuming $S_{\mathrm{l}}^{\mathrm{pm,if}} = S_{\mathrm{l}}^{\mathrm{pm,c}}$ is no longer necessary. This means the pressure gradients of the liquid and of the gas phase might differ. Further, interface quantities are used in upwinding decisions and to construct the gradients with quantities from only one subdomain. The diffusive and conductive fluxes in (12) and (13) can be expressed by:

$$\mathbf{j}_{\alpha}^{\kappa,\omega} = -D_{\mathrm{g}}^{\omega} \rho_{\mathrm{mol,g}} M^{\kappa} \frac{x_{\alpha}^{\mathrm{w},\omega,\mathrm{if}} - x_{\alpha}^{\mathrm{w},\omega,\mathrm{c}}}{\left(\mathbf{x}^{\omega,\mathrm{if}} - \mathbf{x}^{\omega,\mathrm{c}}\right) \cdot \mathbf{n}^{\omega}} , \quad -\lambda_{\mathrm{g}}^{\omega} \nabla T = -\lambda_{\mathrm{g}}^{\omega} \frac{T^{\omega,\mathrm{if}} - T^{\omega,\mathrm{c}}}{\left(\mathbf{x}^{\omega,\mathrm{if}} - \mathbf{x}^{\omega,\mathrm{c}}\right) \cdot \mathbf{n}^{\omega}} .$$

In each step, to calculate the fluxes across the interface, this system of equations has to be solved. If it has converged, the pseudo-unknown $p_{\mathrm{g}}^{\mathrm{pm,if}}$ is used to calculate all fluxes. In comparison to all previous coupling methods, this method is able to consider co-current flow of liquid and gas for estimating the normal momentum flux and thus the interface gas pressure. To avoid numerical problems with unphysical solutions, a slope-limiter for the solution of the pseudo-unknowns can be applied. This is necessary at the beginning of a simulation, when the processes and state variables inside the porous medium and the free flow are far from their equilibrium state. The

slope-limiter can also be necessary on coarse grids when, e.g., the temperature which should evolve at the interface is much lower than the temperature within the two cells. However, the converged solution should not be affected by the slope-limiter.

4 Numerical Results and Discussion

In this section, the advantages and drawbacks of the different coupling methods CM1-CM4 are discussed based on numerical results. The model is implemented in the open-source simulator DuMu^{x} [8, 11], using DUNE [3], DUNE-PDELab, and DUNE-MultiDomain. All code is open-source and available via: https://git.iws.uni-stuttgart.de/dumux-pub/Fetzer2017a. More details on the implementation and software can be found in [9]. An implicit Euler scheme is used for time-stepping, the monolithic matrix is solved using the direct solver UMFPack.

The setup for numerical analysis features a free gas flow, from top to bottom, with a no-slip condition at the left wall and symmetry conditions on the right, see Fig. 3 and Table 2. The maximum velocity is chosen to obtain a Reynolds number of 1000. The porous medium is filled with gas and liquid. Its properties, which corresponds to a well sorted coarse sand, are given in Table 3. Gravitational forces are neglected in both subdomains. The equidistant base grid has five cells in each direction. A grid convergence study with five refinement steps is performed. The simulation is started with $\Delta t = 0.1$ s and ends after $t_{\text{end}} = 43\,200$ s.

The resulting velocity profiles at the end of the simulation for all coupling methods on the finest grid are shown in Fig. 4. They indicate some differences between CM1 and the other coupling methods. For CM1, the gas velocity on the porous-medium side of the vertical interface is even higher than the velocity on the free-flow side from the other methods. The fact that only the closest cell to the vertical interface is affected, prevails throughout the grid refinement. In addition, the refinement even leads to an increment of velocity in that cell. For a parallel free flow adjacent to the porous medium or in cases that the free flow completely enters the porous medium, CM1 does not show such singular behavior, see [9]. Because of the evaporation process, the gas velocity at the interface is directed from the porous medium into the free flow, see Fig. 4. As shown in the pressure plots in Fig. 5, except for CM3, the gas flow inside the porous medium is directed towards its center. The water flow has the reversed direction, to compensate for the mass lost by evaporation. The

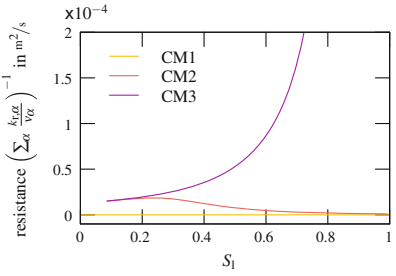


Fig. 2 Generic flow resistance for entering the porous medium

Table 2 Initial and boundary conditions

Parameter	Value
$v_{g,x}^{ff}$ [m/s]	0
$v_{g,y}^{ff}$ [m/s]	$-0.0161\,x^2$
p_g^{ff} [Pa]	1E5
$X_g^{w,ff}$ [−]	0.0005
T^{ff} [K]	303.15
p_g^{pm} [Pa]	1E5
S_l^{pm} [−]	0.85
T^{pm} [K]	293.15

Table 3 Soil properties (coarse sand)

Parameter	Value
$K\,[\mathrm{m}^2]$	3.87E − 10
ϕ [−]	0.32
$S_{r,l}$ [−]	0.0875
$S_{r,g}$ [−]	0.02
α_{vg} [1/Pa]	8.77E − 4
n_{vg} [−]	12.7
λ_s [W/(m K)]	5.26
$c_s\,[\mathrm{K/(kg\,K)}]$	790
$\rho_s\,[\mathrm{kg/m}^3]$	2700
α_{BJ} [−]	1.0

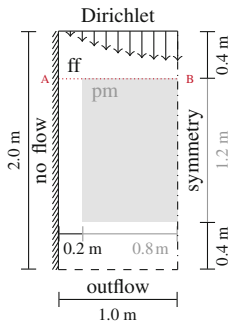


Fig. 3 Setup for numerical simulations

increased interface-normal momentum flux caused by this water mass flux and the different flow direction of the two phases can only be considered with CM4 which consequently reveals higher velocities above the porous medium, Fig. 4.

First, the grid convergence for each coupling method against its solution on the finest grid shows a similar trend for the velocity profile above the narrow free-flow

Table 4 Rel. L^2-errors of vertical velocities for different refinement levels (RL) compared to a reference solution (ref) on finest grid of each coupling method. One simulation did not converge (dnc)

RL	Above ff [-]				Above pm [-]				pm (ref CM4) [-]			CPU times [s]	
	CM1	CM2	CM3	CM4	CM1	CM2	CM3	CM4	CM1	CM2	CM3	CM2	CM4
0	0.568	0.178	0.177	dnc	4.768	0.989	0.906	dnc	1379.0	1.005	0.926	0.8	dnc
1	0.353	0.258	0.257	0.257	3.714	0.942	0.812	0.712	1092.0	0.963	0.853	4.1	6.1
2	0.215	0.203	0.203	0.202	2.751	0.854	0.642	0.489	837.6	0.890	0.722	20	30
3	0.112	0.126	0.125	0.125	1.905	0.722	0.386	0.236	625.3	0.811	0.527	112	161
4	0.055	0.065	0.065	0.065	1.092	0.545	0.143	0.119	438.7	0.778	0.346	787	1328
5	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	285.0	0.890	0.234	8532	10031

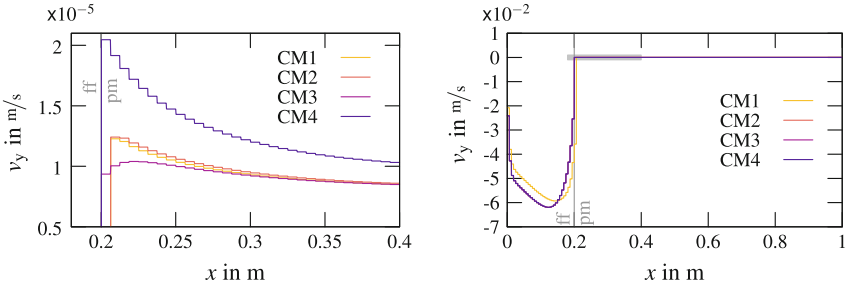


Fig. 4 Gas velocity profiles on finest grid at the end of the simulation. *Left*, for the entire cross-section at the top of the porous medium (y = 1.6 m, A-B in Fig. 3). *Right*, zoom on the edge of the porous medium

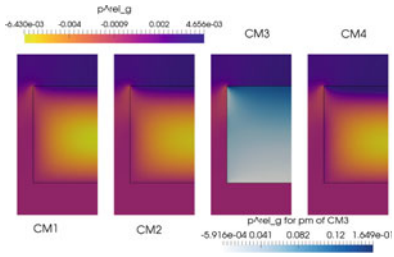


Fig. 5 Relative gas pressure $p_g^{rel} = p_g - 1E5$ Pa on finest grid at the end of the simulation

channel, see Table 4. For the section above the porous medium, the convergence rate is improved from CM1 to CM4. Note that, compared to CM2, the convergence rate for CM3 improves with increasing refinement. If the finest grid of CM4 is considered as a reference solution, CM1 and CM2 do not converge against the reference solution. Above the porous medium, CM1 reveals large errors, the error for CM2 is almost constant, but for CM3 the error is reduced with finer grids. For the free-flow section in contrast, the convergence rates of CM2 and CM3 are insensitive with respect to the chosen reference solution (results are not shown).

The CPU times in Table 4 indicate about 50% higher CPU times for CM4, which are caused by the additional interface system of equations required to be solved. The CPU times for CM1, CM2, and CM3 are very similar to each other.

References

1. Baber, K., Mosthaf, K., Flemisch, B., Helmig, R., Müthing, S., Wohlmuth, B.: Numerical scheme for coupling two-phase compositional porous-media flow and one-phase compositional free flow. IMA J. Appl. Math. **77**(6) (2012). doi:10.1093/imamat/hxs048
2. Beavers, G.S., Joseph, D.D.: Boundary conditions at a naturally permeable wall. J. Fluid Mech. **30**(1) (1967). doi:10.1017/S0022112067001375
3. Blatt, M., Burchardt, A., Dedner, A., Engwer, C., Fahlke, J., Flemisch, B., Gersbacher, C., Gräser, C., Gruber, F., Grüninger, C., Kempf, D., Klöfkorn, R., Malkmus, T., Müthing, S., Nolte, M., Piatkowski, M., Sander, O.: The distributed and unified numerics environment, version 2.4. Arch. Numer. Softw. **4**(100) (2016). doi:10.11588/ans.2016.100.26526
4. Dahmen, W., Gotzen, T., Müller, S., Rom, M.: Numerical simulation of transpiration cooling through porous material. Int. J. Numer. Methods Fluids **76**(6) (2014). doi:10.1002/fld.3935

5. Defraeye, T.: Advanced computational modelling for drying processes. Rev. Appl. Energy **131** (2014). doi:10.1016/j.apenergy.2014.06.027
6. Discacciati, M., Gervasio, P., Giacomini, A., Quarteroni, A.: The interface control domain decomposition method for Stokes–Darcy coupling. SIAM J. Numer. Anal. **54**(2) (2016). doi:10.1137/15M101854X
7. Fetzer, T., Smits, K.M., Helmig, R.: Effect of turbulence and roughness on coupled porous-medium/free-flow exchange processes. Trans. Porous Med. **114**(2) (2016). doi:10.1007/s11242-016-0654-6
8. Flemisch, B., Darcis, M., Erbertseder, K., Faigle, B., Lauser, A., Mosthaf, K., Müthing, S., Nuske, P., Tatomir, A., Wolff, M., Helmig, R.: DuMuX: DUNE for multi-phase, component, scale, physics, ... flow and transport in Porous Media. Adv. Water Resour. **34**(9) (2011). doi:10.1016/j.advwatres.2011.03.007
9. Grüninger, C., Fetzer, T., Flemisch, B., Helmig, R.: Coupling DuMuX and DUNE-PDELab to investigate evaporation at the interface between Darcy and Navier-Stokes flow. Arch. Numer. Softw. (2016). (submitted)
10. Harlow, F.H., Welch, J.E.: Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface. Phys. Fluids **8**(12) (1965). doi:10.1063/1.1761178
11. Hommel, J., Ackermann, S., Beck, M., Becker, B., Class, H., Fetzer, T., Flemisch, B., Gläser, D., Grüninger, C., Heck, K., Kissinger, A., Koch, T., Schneider, M., Seitz, G., Weishaupt, K.: DuMuX 2.10.0 (2016). doi:10.5281/zenodo.159007
12. Iliev, O., Laptev, V.: On numerical simulation of flow through oil filters. Comput. Vis. Sci. **6**(2) (2004). doi:10.1007/s00791-003-0118-8
13. Jambhekar, V.A., Helmig, R., Schröder, N., Shokri, N.: Free-flow–Porous-Media coupling for evaporation-driven transport and precipitation of salt in soil. Trans. Porous Med. **110**(2) (2015). doi:10.1007/s11242-015-0516-7
14. Layton, W.J., Schieweck, F., Yotov, I.: Coupling fluid flow with porous media flow. SIAM J. Numer. Anal. **40**(6) (2002). doi:10.1137/S0036142901392766
15. Masson, R., Trenty, L., Zhang, Y.: Coupling compositional liquid gas Darcy and free gas flows at porous and free-flow domains interface. J. Comput. Phys. **321** (2016). doi:10.1016/j.jcp.2016.06.003
16. Moeckel, G.P.: Thermodynamics of an interface. Arch. Ration. Mech. Anal. **57**(3) (1975). doi:10.1007/BF00280158
17. Mosthaf, K., Baber, K., Flemisch, B., Helmig, R., Leijnse, A., Rybak, I., Wohlmuth, B.: A coupling concept for two-phase compositional porous-medium and single-phase compositional free flow. Water Resour. Res. **47**(10) (2011). doi:10.1029/2011WR010685
18. Mosthaf, K., Helmig, R., Or, D.: Modeling and analysis of evaporation processes from porous media on the REV scale. Water Resour. Res. **50**(2) (2014). doi:10.1002/2013WR014442
19. Rybak, I., Magiera, J., Helmig, R., Rohde, C.: Multirate time integration for coupled saturated/unsaturated porous medium and free flow systems. Comput. Geosci. **19**(2) (2015). doi:10.1007/s10596-015-9469-8
20. Saffman, P.: On the boundary condition at the surface of a porous medium. Stud. Appl. Math. **50**(2) (1971). doi:10.1002/sapm197150293

Non-conforming Localized Model Reduction with Online Enrichment: Towards Optimal Complexity in PDE Constrained Optimization

Mario Ohlberger and Felix Schindler

Abstract We propose a new non-conforming localized model reduction paradigm for efficient solution of large scale or multiscale PDE constrained optimization problems. The new conceptual approach goes beyond the classical offline/online splitting of traditional projection based model order reduction approaches for the underlying state equation, such as the reduced basis method. Instead of first constructing a surrogate model that has globally good approximation quality with respect to the whole parameter range, we propose an iterative enrichment procedure that refines and locally adapts the surrogate model specifically for the parameters that are depicted during the outer optimization loop.

Keywords Model reduction · Reduced basis method · LRBMS · Optimization · Control · Online enrichment · Discontinuous Galerkin

MSC (2010): 35Q93 · 65K10 · 65N30

1 Introduction

We are concerned with model reduction for parameter optimization of general elliptic multiscale problems, where the optimization functional is defined on a macro scale and the material design parameters are considered to have influence on the micro scale. Such optimization problems naturally arise, e.g., in optimal design of composed materials or in the design of technical devices that rely on multiscale

M. Ohlberger (✉)
Applied Mathematics, Center for Nonlinear Sciences and Center for Multiscale Theory and Computation, University of Muenster, Einsteinstr. 62, 48149 Münster, Germany
e-mail: mario.ohlberger@uni-muenster.de

F. Schindler
Applied Mathematics and Center for Nonlinear Sciences, University of Muenster, Einsteinstr. 62, 48149 Münster, Germany
e-mail: felix.schindler@uni-muenster.de

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_38

processes, such as fuel cells or batteries. In previous works [13, 14] we considered such problems under the assumption of scale separation, or even local periodicity, which allowed us to suggest a model reduction approach based on the two scale limit equation of the homogenized problem. However, in many real applications such a structural assumption is too restrictive and hence more general approaches need to be developed in general heterogeneous situations. In recent contributions, we introduced and analyzed the localized reduced basis multiscale method (LRBMS) [12, 15, 16] which is particularly designed to efficiently cope with general heterogeneous parameterized multiscale problems. In particular we developed an efficient localized a posteriori error estimator against the underlying true solution of the parameterized problem and demonstrated how this error estimator can be used to overcome the classical offline/online splitting of reduced basis (RB) methods, by means of the newly developed concept of online enrichment. The proposed method merely requires a very cheap preparation step and then iteratively enriches localized snapshot spaces using the localized a posteriori error information. In this contribution, we combine our development from the previous two approaches to suggest, for the first time, an efficient solution algorithm for parameter optimization of elliptic multiscale or large scale problems on the basis of our non-conforming localized model reduction approach with online enrichment.

In detail we look at the following multiscale or large scale optimization setting:

$$\left.\begin{aligned} \text{Find} \quad & \mu^* = \arg\min J\left(u^\varepsilon(\mu), \mu\right) \\ \text{subject to} \quad & C_j\left(u^\varepsilon(\mu), \mu\right) \leq 0 \quad \forall j = 1, \dots, m, \\ & \mu \in \mathscr{P} \end{aligned}\right\} \tag{1}$$

with a compact *parameter set* $\mathscr{P} \subset \mathbb{R}^P$ for $P \in \mathbb{N}$. In (1), the *state variable* $u^\varepsilon(\mu)$ is given as the solution of the following (parametrized) multiscale problem:

$$\left.\begin{aligned} -\nabla \cdot \left(A^\varepsilon(\mu) \nabla u^\varepsilon(\mu)\right) &= f(\mu) \quad (\text{in } \Omega) \\ u^\varepsilon(\mu) &= g_D \quad (\text{on } \partial\Omega) \end{aligned}\right\} \tag{2}$$

In (2), $\Omega \subset \mathbb{R}^d$ for $d = 1, 2, 3$ is a bounded domain and A^ε denotes a general multiscale diffusion tensor (the multiscale nature of which is denoted by $\varepsilon > 0$) without any further structural assumptions. We make use of the short notation $u(\mu) := u^\varepsilon(\mu) := u^\varepsilon(\cdot; \mu)$ and will use analogue expressions for all functions that depend on both spatial variables and parameters.

There is a large variety of numerical algorithms for general optimization problem such as (1), see for example [20]. Typically, these algorithms are based on necessary and/or sufficient optimality conditions for local optima which involve higher order derivatives of the participating functions. In order to obtain a reduced approximation of (1) it is thus not sufficient to provide fast approximations solely of the state $u(\mu)$, but also for derivatives with respect to the parameter μ.

As the solution of such optimization problems usually requires repeated evaluations of the underlying multiscale partial differential equation (PDE) for different sets of parameters, model reduction is applied to increase computational efficiency. In this contribution we are concerned with a localized generalization of the RB approach [8]. The application of the RB approach to parameter optimization of elliptic PDEs was first presented in [17], and generalized to multiscale problems in [13]. A posteriori error estimates for reduced approximation of linear-quadratic optimization problems and parametrized optimal control problems with control constraints were studied, e.g., in [4, 7, 19, 21]. In the context of optimal control and mesh adaptivity of the underlying finite element discretizations we refer to [2, 9, 11, 18] and the references therein.

The rest of this paper is organized as follows: In Sect. 2 we introduce a weak formulation of the underlying parameterized multiscale problem in broken Sobolev spaces employing a non-conforming discontinuous Galerkin (DG) variational formulation. Based on this non-conforming setting, we then introduce in Sect. 3 localized reduced spaces and related localized RB surrogate models, both for the state equation and for the equations that characterize the derivatives of the state with respect to the parameters. Finally, in Sect. 4, a new iterative solution concept based on successively enhanced surrogate models is proposed and discussed.

2 Non-conforming Weak Formulation of the Parameterized Multiscale Problem

In order to derive a suitable non-conforming weak formulation for our model reduction approach, we first assume that a non-overlapping decomposition of the underlying domain Ω is given by a coarse triangulation $\mathscr{T}_H$ with cells $T_j \in \mathscr{T}_H$, $j = 1, \dots N_H$. Furthermore, each macro cell T_j is further decomposed by a local fine resolution triangulation τ_h^j, that resolves all fine scale features of the multiscale problem. We then define the global fine scale partition τ_h as the union of all its local contributions, i.e., $\tau_h = \bigcup_{j=1}^{N_H} \tau_h^j$. Hence, τ_h is a nested refinement of $\mathscr{T}_H$ as schematically depicted in Fig. 1. Let $H^1(\tau_h) := \{v \in L^2(\Omega) \mid v|_t \in H^1(t) \ \ \forall t \in \tau_h\}$ denote the broken Sobolev space on τ_h, which naturally inherits the decomposition $H^1(\tau_h) = \bigoplus_{j=1}^{N_H} H^1(\tau_h^j)$.

Definition 1 (*Weak solution of the multiscale problem in broken spaces*) We call $u(\mu) \in H^1(\tau_h)$ weak solution of (2), if

$$a_{\mathrm{DG}}\big(u(\mu), v; \mu\big) = L_{\mathrm{DG}}(v; \mu) \qquad \text{for all } v \in H^1(\tau_h). \tag{3}$$

Here, the DG bilinear form a_{DG} and the right hand side L_{DG} are given as

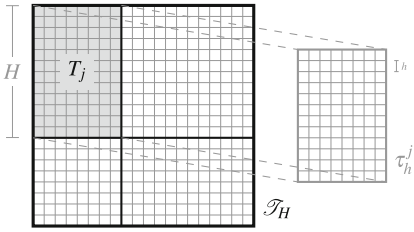


Fig. 1 Sketch of domain decomposition into macro elements $T_j \in \mathscr{T}_H$ for the definition of the LRBMS and the underlying fine grids $\tau_h^j \subset \mathscr{T}_H$ that are used for the construction of the local approximation spaces

$$
\begin{aligned}
a_{\mathrm{DG}}(v, w; \mu) &:= \sum_{t \in \tau_h} \int_t A^\varepsilon(\mu) \nabla v \cdot \nabla w - \sum_{e \in \mathscr{F}_h^I} \int_e \{A^\varepsilon(\mu) \nabla v \cdot \mathbf{n_e}\}[w] \\
&\quad - \sum_{e \in \mathscr{F}_h^I} \int_e \{A^\varepsilon(\mu) \nabla w \cdot \mathbf{n_e}\}[v] + \sum_{e \in \mathscr{F}_h^I} \frac{\sigma_e(\boldsymbol{\mu})}{|e|^\beta} \int_e [v][w], \\
L_{\mathrm{DG}}(v; \mu) &:= \sum_{t \in \tau_h} \int_t f v + \sum_{e \in \mathscr{F}_h^D} \int_e \left(\frac{\sigma_e(\boldsymbol{\mu})}{|e|^\beta} v - A^\varepsilon(\mu) \nabla v \cdot \mathbf{n} \right) g_D,
\end{aligned}
$$

where the parametric positive penalty function $\sigma_e(\boldsymbol{\mu}) : \mathscr{P} \to \mathbb{R}$ and the averages and jumps $\{\cdot\}$ and $[\cdot]$ across inner and boundary interfaces $e \in \mathscr{F}_h^I \cup \mathscr{F}_h^D$ are chosen similar to SWIPDG [6, 16].

3 Localized Model Reduction for PDE Constrained Optimization

Our non-conforming localized model reduction approach is based on the construction of appropriate low dimensional local approximation spaces $U_N^j \subset H^1(\tau_h^j)$ of local dimensions N_j that form the global reduced solution space via

$$
U_H^N = \bigoplus_{j=1}^{N_H} U_N^j, \qquad N := \dim(U_H^N) = \sum_{j=1}^{N_H} N_j. \tag{4}
$$

Once such a reduced approximation space is constructed, the LRBMS approximation is defined as follows.

Definition 2 (*The localized reduced basis multiscale method*) We call $u_N(\mu) \in U_H^N$ a localized reduced basis multiscale approximation of (3) if it satisfies

$$
a_{\mathrm{DG}}(u_N(\mu), v_N; \mu) = L_{\mathrm{DG}}(\mu; v_N) \qquad \text{for all } v_N \in U_H^N. \tag{5}
$$

Note that (5) is a globally coupled reduced problem, where all arising quantities can nevertheless be locally computed w.r.t the local reduced spaces U_N^j.

To simplify the presentation, we assume in the sequel that the bilinear form a_{DG} and the right hand side L_{DG} are affinely decomposable (see [5] and the references therein for the treatment of general nonlinear operators), which allows for an efficient offline/online splitting of the resulting reduced problem:

$$a_{\mathrm{DG}}(u, v; \mu) = \sum_{q=1}^{Q_A} \theta_q^A(\mu) a_{\mathrm{DG}}^q(u, v), \qquad L_{\mathrm{DG}}(\mu; v) = \sum_{q=1}^{Q_L} \theta_q^L(\mu) L_{\mathrm{DG}}^q(v). \tag{6}$$

In order to solve the overall optimization problem, we will also need to compute parameter derivatives of the state $u_N(\mu)$. Equations for these quantities are established by differentiating the defining equation (5) of $u_N(\mu)$ with respect to μ_i. Due to linearity and the chain rule, we obtain the following weak formulation for $\partial_{\mu_i} u_N(\mu) \in U_H^N$, for all $v_N \in U_H^N$:

$$a_{\mathrm{DG}}(\partial_{\mu_i} u_N(\mu), v_N; \mu) = -\partial_{\mu_i} a_h(u_N(\mu), v_N; \mu) + \partial_{\mu_i} L_{\mathrm{DG}}(\mu; v). \tag{7}$$

Since a_{DG} and L_{DG} are affinely decomposable, we have

$$\partial_{\mu_i} a_h(u, v; \mu) = \sum_{q=1}^{Q_A} \partial_{\mu_i} \theta_q^A(\mu) a_{\mathrm{DG}}^q(u, v), \quad \partial_{\mu_i} L_{\mathrm{DG}}(\mu; v) = \sum_{q=1}^{Q_L} \partial_{\mu_i} \theta_q^L(\mu) L_{\mathrm{DG}}^q(v).$$

Hence, we can reuse both the same reduced spaces and precomputed reduced system matrices as for the approximation of $u_N(\mu)$. Thus, the computational overhead to compute the parameter derivatives is of the same order as the cost to compute a reduced state equation. Higher order derivatives can be computed analogously by further differentiation of (7).

In recent contributions [1, 3, 10, 16] we discussed several possibilities to construct local reduced spaces U_N^j from global or localized snapshot computations. Thereby, in the concept presented above, it is possible to use finite volume, DG or conforming finite element approximations on the underlying fine partition τ_h or restrictions thereof to a local neighborhood of the macro elements $T_j \in \mathscr{T}_H$. In what follows, we consider the iterative construction of reduced approximation spaces and related surrogate models based on localized a posteriori error control and local enrichment as recently introduced in [16]. In these circumstances we obtain the following estimate on the error w.r.t. the unknown weak solution of (2).

Theorem 1 (Localizable a posteriori error estimate) *With the assumptions and the notation of [16, Cor. 4.5], the following estimate on the full approximation error in the energy norm* $|||v|||_{\overline{\mu}} := \sum_{t \in \tau_h} \int (A^\varepsilon(\overline{\mu}) \nabla v) \cdot \nabla v$ *holds for arbitrary* $\overline{\mu}, \hat{\mu} \in \mathscr{P}$,

$$|||u(\mu) - u_N(\mu)|||_{\overline{\mu}} \leq \eta\big(u_N(\mu)\big) := C(\mu, \overline{\mu}, \hat{\mu})\Bigg[\sum_{j=1}^{N_H} \Big(\eta_j^{nc}(u_N(\mu))^2\Big)^{1/2} + \sum_{j=1}^{N_H} \Big(\eta_j^{r}(u_N(\mu))^2\Big)^{1/2} + \sum_{j=1}^{N_H} \Big(\eta_j^{df}(u_N(\mu))^2\Big)^{1/2}\Bigg]$$

with a computable constant $C(\mu, \overline{\mu}, \hat{\mu}) > 0$ and fully computable local indicators η_j^{nc}, η_j^{r} and η_j^{df} corresponding to the local non-conformity errors, residual errors, and diffusive flux reconstruction errors, respectively.

We refer to [16] for a more detail presentation and derivation of this result.

4 A New Iterative Solution Concept Based on Successively Enhanced Surrogate Models

In classical model reduction approaches for PDE constrained optimization problems, a surrogate model to approximate the parameterized state equation is constructed in a so called offline phase and then successively used for fast evaluations of the state equation for parameters that are selected during the outer optimization loop (see, e.g., [17]). As the selected parameter values during the optimization loop are not known a priori, the surrogate models in such approaches need to be prepared to uniformly approximate the state equation with respect to the whole parameter regime. This might lead to a quite expensive offline construction phase that involves a large number of usually global snapshot computations for suitably selected parameter values. In an optimization loop, however, typically only parameters along a path towards the optimal parameter are depicted, as sketched in Fig. 2. Based on the concept of local enrichment from [16] we thus suggest a new iterative procedure to successively built up or enhance the surrogate model (5), (7) by using only localized snapshot computations for the parameters that are selected during the optimization loop. The resulting approach is thus tailored towards the specific optimization problem in an a posteriori manner.

In more detail, in a first step we initialize the local reduced spaces U_N^j with a classical polynomial coarse scale DG basis of prescribed order, thus ensuring that any reduced solution of the state equation is at least as good as a DG solution on the coarse triangulation $\mathscr{T}_H$. We then optionally employ a discrete weak Greedy algorithm with only very few (typically one) global snapshot computations and enrich

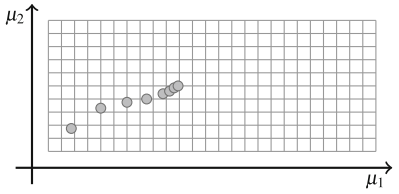


Fig. 2 Sketch of parameter selection during an optimization loop. For each selected parameter point $\mu = (\mu_1, \mu_2)$ an approximation of $u(\mu)$ and its derivatives with respect to μ need to be computed efficiently

U_N^j accordingly. Finally, $U_H^N = \bigoplus_{j=1}^{N_H} U_N^j$ and a related initial surrogate model is constructed as sketched in Sect. 3 above.

During the following optimization loop, given any $\boldsymbol{\mu} \in \mathscr{P}$ from the optimization algorithm, we compute a reduced solution $u_N(\boldsymbol{\mu}) \in U_H^N$ and efficiently assess its quality using the localized a posteriori error estimator $\eta(\boldsymbol{\mu}) := (\sum_{j=1}^{N_H} \eta_j(\boldsymbol{\mu})^2)^{1/2}$ derived in [15, 16]. If the estimated error is above a prescribed tolerance, $\Delta > 0$, we start an intermediate local enrichment phase to enhance the surrogate model in the SEMR (solve → estimate → mark → refine) spirit of adaptive mesh refinement. We refer to Algorithm 4.1 and [16] for a detailed description and evaluation of this enrichment procedure that only involves local snapshot computations for the given parameter $\boldsymbol{\mu}$ on some local neighborhoods $\omega(T_j)$, $T_j \in \mathscr{T}_H$ with Dirichlet boundary values obtained from the insufficient previous reduced surrogate. The algorithm calls a routine `OPT` that performs one optimization step with a descent method based on the old parameter value, the corresponding state and its derivatives with respect to the parameters. It returns the new parameter value and `success=true`, if the optimization criteria has been met.

Algorithm 4.1 Parameter optimization with adaptive enrichment.

Require: $\boldsymbol{\mu}^{(0)} \in \mathscr{P}$, initial local bases ${\Phi_j}^{(0)}$, $\Delta_{\text{model}}, \Delta_{\text{opt}} > 0$, a marking strategy `MARK` and an orthonormalization procedure `ONB` (see [16, Sec. 5]), an optimization routine `OPT` (returning a new parameter and status of convergence).

$n \leftarrow 0$, $U_H^{N\,(0)} \leftarrow \bigoplus_{j=1}^{N_H} \operatorname{span}\big({\Phi_j}^{(0)}\big)$

repeat

Solve (5), (7) *for* $u_N(\boldsymbol{\mu}^{(n)}), \partial_{\mu_i} u_N(\mu^{(n)}) \in U_H^{N\,(n)}$, $i = 1, \ldots, P$.

$m \leftarrow n$

while $\eta(\boldsymbol{\mu}^{(n)}) > \Delta_{\text{model}}$ **do**

for all $j = 1, \ldots, N_H$ **do**

Compute local error indicators $\eta_j(\boldsymbol{\mu}^{(n)})$ *according to [16, Cor. 4.5].*

end for

$\tilde{\mathscr{T}}_H \leftarrow \texttt{MARK}\big(\mathscr{T}_H, \{\eta_j(\boldsymbol{\mu}^{(n)})\}_{j=1}^{N_H}\big)$

for all $T_j \in \tilde{\mathscr{T}}_H$ **do**

Solve locally on $\omega(T_j)$ *for enhanced local snapshot* $u_h^j(\boldsymbol{\mu}^{(n)}) \in H^1(\tau_h^j)$.

${\Phi_j}^{(m+1)} \leftarrow \texttt{ONB}\big(\{{\Phi_j}^{(m)}, u_h^j(\boldsymbol{\mu}^{(n)})\}\big)$

end for

$U_H^{N\,(m+1)} \leftarrow \bigoplus_{T_j \in \tilde{\mathscr{T}}_H} \operatorname{span}\big({\Phi_j}^{(m+1)}\big) \oplus \bigoplus_{T_j \in \mathscr{T}_H \setminus \tilde{\mathscr{T}}_H} \operatorname{span}\big({\Phi_j}^{(m)}\big)$

Solve (5), (7) *for* $u^N(\boldsymbol{\mu}^{(n)}), \partial_{\mu_i} u^N(\mu^{(n)}) \in U_H^{N\,(m+1)}$, $i = 1, \ldots, P$.

$m \leftarrow m + 1$

end while

$(\mu^{(n+1)}, \texttt{success}) \leftarrow \texttt{OPT}\big(\boldsymbol{\mu}^{(n)}, \Delta_{\text{opt}}, u^N(\boldsymbol{\mu}_n), \{\partial_{\mu_i} u^N(\mu_n)\}_{i=1}^{P}\big)$

$n \leftarrow n + 1$

until `success`

return optimal parameter $\boldsymbol{\mu}^{(n)}$ and state $u^N(\boldsymbol{\mu}^{(n)})$

5 Conclusion and Outlook

In this contribution we proposed how to efficiently use the localized reduced basis multiscale method with local enrichment in the context of surrogate modeling for the solution of large scale or multiscale PDE constrained optimization problems. The resulting approach iteratively constructs enhanced surrogate models specifically tailored to the solution of the optimization problem in an a posteriori manner and thereby overcomes offline/online splitting of traditional projection based model reduction approaches. A deeper numerical analysis of the presented approach as well as its thorough analysis in numerical experiments are subject to ongoing research.

References

1. Albrecht, F., Haasdonk, B., Kaulmann, S., Ohlberger, M.: The localized reduced basis multiscale method. In: Proceedings of Algoritmy 2012. Conference on Scientific Computing, Vysoke Tatry, Podbanske, 9–14 Sept. 2012, pp. 393–403. Slovak University of Technology in Bratislava, Publishing House of STU (2012)
2. Benedix, O., Vexler, B.: A posteriori error estimation and adaptivity for elliptic optimal control problems with state constraints. Comput. Optim. Appl. **44**(1), 3–25 (2009)
3. Buhr, A., Engwer, C., Ohlberger, M., Rave, S.: ArbiLoMod, a simulation technique designed for arbitrary local modifications (2015). http://arxiv.org/abs/1512.07840
4. Dedè, L.: Reduced basis method and error estimation for parametrized optimal control problems with control constraints. J. Sci. Comput. **50**(2), 287–305 (2012)
5. Drohmann, M., Haasdonk, B., Ohlberger, M.: Reduced basis approximation for nonlinear parametrized evolution equations based on empirical operator interpolation. SIAM J. Sci. Comput. **34**, A937–A969 (2012). doi:10.1137/10081157X
6. Ern, A., Stephansen, A.F., Zunino, P.: A discontinuous galerkin method with weighted averages for advection-diffusion equations with locally small and anisotropic diffusivity. IMA J. Numer. Anal. **29**(2), 235–256 (2009)
7. Grepl, M.A., Kärcher, M.: Reduced basis a posteriori error bounds for parametrized linear-quadratic elliptic optimal control problems. C. R. Math. Acad. Sci. Paris **349**(15–16), 873–877 (2011)
8. Haasdonk, B.: Reduced basis methods for parametrized PDEs–A tutorial introduction for stationary and instationary problems. In: Model Reduction and Approximation: Theory and Algorithms. Benner, P., Cohen, A., Ohlberger, M., and Willcox, K. (eds.). SIAM, Philadelphia, PA (2017)
9. Hintermüller, M., Hinze, M., Hoppe, R.H.W.: Weak-duality based adaptive finite element methods for PDE-constrained optimization with pointwise gradient state-constraints. J. Comput. Math. **30**(2), 101–123 (2012)
10. Kaulmann, S., Ohlberger, M., Haasdonk, B.: A new local reduced basis discontinuous Galerkin approach for heterogeneous multiscale problems. C. R. Math. Acad. Sci. Paris **349**(23–24), 1233–1238 (2011)
11. Liu, W., Yan, N.: A posteriori error estimates for distributed convex optimal control problems. Adv. Comput. Math. **15**(1–4), 285–309 (2001)
12. Ohlberger, M., Rave, S., Schindler, F.: Model reduction for multiscale lithium-ion battery simulation. In: Karasözen, B., et al. (ed.) Numerical Mathematics and Advanced Applications ENUMATH 2015. Lecture Notes in Computational Science and Engineering, vol. 112, pp. 317–331. Springer (2016)

13. Ohlberger, M., Schaefer, M.: A reduced basis method for parameter optimization of multiscale problems. In: Proceedings of Algoritmy 2012, Conference on Scientific Computing, Vysoke Tatry, Podbanske, 9-14 Sept. pp. 272–281 (2012)
14. Ohlberger, M., Schaefer, M.: Error control based model reduction for parameter optimization of elliptic homogenization problems. In: Le Gorrec, Y. (ed.) 1st IFAC Workshop on Control of Systems Governed by Partial Differential Equations, CPDE 2013; Paris; France; 25 September 2013 through 27 Sept. 2013; Code 103235, vol. 1, pp. 251–256. International Federation of Automatic Control (IFAC) (2013)
15. Ohlberger, M., Schindler, F.: A-posteriori error estimates for the localized reduced basis multi-scale method. In: Fuhrmann, J., Ohlberger, M., Rohde, C. (eds.) Finite Volumes for Complex Applications VII-Methods and Theoretical Aspects, Springer Proceedings in Mathematics & Statistics, vol. 77, pp. 421–429. Springer International Publishing (2014)
16. Ohlberger, M., Schindler, F.: Error control for the localized reduced basis multi-scale method with adaptive on-line enrichment. SIAM J. Sci. Comput. **37**(6), A2865–A2895 (2015)
17. Oliveira, I.B., Patera, A.T.: Reduced-basis techniques for rapid reliable optimization of systems described by affinely parametrized coercive elliptic partial differential equations. Optim. Eng. **8**(1), 43–65 (2007)
18. Rösch, A., Wachsmuth, D.: A-posteriori error estimates for optimal control problems with state and control constraints. Numer. Math. **120**(4), 733–762 (2012)
19. Tröltzsch, F., Volkwein, S.: POD a-posteriori error estimates for linear-quadratic optimal control problems. Comput. Optim. Appl. **44**(1), 83–115 (2009)
20. Vanderbei, R.J., Shanno, D.F.: An interior-point algorithm for nonconvex nonlinear programming. Comput. Optim. Appl. **13**(1–3), 231–252 (1999). (Computational optimization—a tribute to Olvi Mangasarian, Part II)
21. Vossen, G., Volkwein, S.: Model reduction techniques with a-posteriori error analysis for linear-quadratic optimal control problems. Numer. Algebr. Control Optim. **2**(3), 465–485 (2012)

Combining the Hybrid Mimetic Mixed Method and the Eulerian Lagrangian Localised Adjoint Method for Approximating Miscible Flows in Porous Media

Hanz Martin Cheng and Jérôme Droniou

Abstract We design a numerical scheme for a miscible displacement in porous media. This scheme is based on the Hybrid Mimetic Mixed method, which is applicable on generic meshes, and uses a characteristic method for dealing with the advection.

Keywords Hybrid Mimetic Mixed (HMM) schemes · ELLAM miscible displacement · Porous media

1 Introduction

One of the tertiary oil recovery processes consists in injecting, in an underground oil reservoir, a solvent that mixes with the residing oil and reduces its viscosity, thus enabling its displacement towards a production well. Let Ω be a bounded domain in $\mathbb{R}^d$ and $[0, T]$ be a time interval. Denote by $\mathbf{K}(\mathbf{x})$ and $\phi(\mathbf{x})$ the permeability tensor and the porosity of the medium, respectively. Neglecting gravity, the mathematical model is [8]:

$$\begin{aligned} \nabla \cdot \mathbf{u} &= q^+ - q^- := q \qquad \text{on } \Omega \times [0, T] \\ \mathbf{u} &= -\frac{\mathbf{K}}{\mu(c)} \nabla p \qquad \text{on } \Omega \times [0, T] \end{aligned} \tag{1a}$$

$$\phi \frac{\partial c}{\partial t} + \nabla \cdot (\mathbf{u}c - \mathbf{D}(\mathbf{x}, \mathbf{u}) \nabla c) = q^+ - cq^- := q_c \qquad \text{on } \Omega \times [0, T] \tag{1b}$$

This coupled system of PDEs has unknowns $p(\mathbf{x}, t)$ the pressure of the mixture, $\mathbf{u}(\mathbf{x}, t)$ the Darcy velocity, and $c(\mathbf{x}, t)$ the concentration of the injected solvent. The

H.M. Cheng (✉) · J. Droniou
School of Mathematical Sciences, Monash University, Victoria 3800, Australia
e-mail: hanz.cheng@monash.edu

J. Droniou
e-mail: jerome.droniou@monash.edu

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_39

functions q^+ and q^- represent the injection and production wells respectively, and $\mathbf{D}(\mathbf{x}, \mathbf{u})$ denotes the diffusion tensor

$$\mathbf{D}(\mathbf{x}, \mathbf{u}) = \phi(\mathbf{x}) \left[d_m \mathbf{I} + d_l |\mathbf{u}| \mathscr{P}(\mathbf{u}) + d_t |\mathbf{u}| (\mathbf{I} - \mathscr{P}(\mathbf{u})) \right] \text{ with } \mathscr{P}(\mathbf{u}) = \left(\frac{u_i u_j}{|\mathbf{u}|^2} \right)_{i,j}.$$

Here, d_m is the molecular diffusion coefficient, d_l and d_t are the longitudinal and transverse dispersion coefficients respectively, and $\mathscr{P}(\mathbf{u})$ is the projection matrix along the direction of $\mathbf{u}$. Also, $\mu(c) = \mu(0)[(1-c) + M^{1/4}c]^{-4}$ is the viscosity of the fluid mixture, where $M = \mu(0)/\mu(1)$ is the mobility ratio of the two fluids. We consider no-flow boundary conditions and, as usual for this process, zero initial conditions for the concentration:

$$\mathbf{u} \cdot \mathbf{n} = (\mathbf{D} \nabla c) \cdot \mathbf{n} = 0 \text{ on } \partial\Omega \times [0, T], \qquad c(\cdot, 0) = 0 \text{ in } \Omega. \tag{1c}$$

The pressure is fixed by imposing a zero average: for all $t \in [0, T]$, $\int_\Omega p(\mathbf{x}, t) \mathrm{d}\mathbf{x} = 0$

A number of numerical schemes have been considered for this model, some of which are the Mixed Finite Element–Eulerian Lagrangian Localised Adjoint Methods (MFEM–ELLAM) [10] and a Mixed Finite Volume (MFV) scheme with upwinding [3]. Due to the use of finite element methods, the MFEM–ELLAM is only limited to certain types of meshes. Moreover, a large number of quadrature points is required to produce acceptable results [9]. The MFV (part of the Hybrid Mimetic Mixed (HMM) schemes, which contain in particular mixed-hybrid Mimetic Finite Differences [5]) is adapted to more generic meshes, but the upwinding tends to introduce excess diffusion in the solution. The purpose of this paper is to discretise (1) using the HMM method, thus allowing for generic meshes, and using the ELLAM for the advective term, to avoid the pitfalls of upwinding.

2 The HMM–ELLAM

We consider polytopal meshes as defined in [4], in dimension $d = 2$. Thus, $\mathscr{T} = (\mathscr{M}, \mathscr{E}, \mathscr{P})$ are the set of cells, edges, and points of our mesh, respectively. $\mathscr{E}_K \subset \mathscr{E}$ denotes the set of edges of the cell $K \in \mathscr{M}$. A usual way to approximate (1) is to use a two-step process. Starting from a known value $c^{(n)}$ of c at time level n (for $n = 0$, $c^{(0)} = 0$), a numerical solution $p^{(n+1)}$ for p at time level $n+1$ is computed by approximating (1a) with $c = c^{(n)}$. This computation also provides an approximation $\mathbf{u}^{(n+1)}$ of the Darcy velocity at time level $n+1$, and possibly of secondary quantities (e.g., fluxes). The concentration $c^{(n+1)}$ at time level $n+1$ is then computed by approximating (1b) by using $\mathbf{u} = \mathbf{u}^{(n+1)}$ and the aforementioned secondary quantities.

2.1 Numerical Scheme for the Pressure Equation

Incorporating $\int_\Omega p = 0$, the variational formulation for (1a) for each cell $K \in \mathscr{M}$ is then given by

$$\int_K \frac{\mathbf{K}}{\mu}\nabla p \cdot \nabla v - \int_{\partial K} \frac{\mathbf{K}}{\mu}\nabla p \cdot \boldsymbol{n}_{K,\sigma} v + \int_\Omega p \int_K v = \int_K qv\,, \quad \forall v \in H^1(\Omega). \quad (2)$$

We present the HMM method in its "finite volume" form, see e.g., [5]. The space of degrees of freedom is $X_{\mathscr{T}} := \{w = ((w_K)_{K\in\mathscr{M}}, (w_\sigma)_{\sigma\in\mathscr{E}_K})\}$. For $\sigma \in \mathscr{E}_K$, denote by $T_{K,\sigma}$ the triangle with vertex $\mathbf{x}_K$ and base σ (see Fig. 1), and define

$$\forall w \in X_{\mathscr{T}}\,,\ \nabla_H w(\mathbf{x}) = \overline{\nabla}_K w + \frac{\sqrt{2}}{d_{K,\sigma}}[w_\sigma - w_K - \overline{\nabla}_K w \cdot (\overline{\boldsymbol{x}}_\sigma - \boldsymbol{x}_K)]\boldsymbol{n}_{K,\sigma}\,, \quad (3)$$

where $\overline{\nabla}_K w = |K|^{-1}\sum_{\sigma\in\mathscr{E}_K} |\sigma| w_\sigma \boldsymbol{n}_{K,\tilde{\sigma}}$ is a linearly exact discretization of the gradient (it is exact if $(w_\sigma)_{\sigma\in\mathscr{E}_K}$ interpolate an affine function at the edge midpoints) and $d_{K,\sigma}$ is the orthogonal distance between $\boldsymbol{x}_K$ and σ.

The concentration at time n is also approximated in $X_{\mathscr{T}}$, and so cell values $(c_K^n)_{K\in\mathscr{M}}$ are accessible. We use them to define the pressure fluxes by:

$$\forall K \in \mathscr{M}\,,\ \forall v \in X_{\mathscr{T}}\,,\ \sum_{\sigma\in\mathscr{E}_K} F_{K,\sigma}(v_K - v_\sigma) = \int_K \frac{\mathbf{K}(x)}{\mu(c_K^n)}\nabla_H p(\mathbf{x}) \cdot \nabla_H v(\mathbf{x}) d\mathbf{x}.$$

The discrete form of (2) then follows easily. Taking test functions so that $v_K = 1$ for cell K and 0 for all other cells gives the balance of fluxes, whilst choosing $v_\sigma = 1$ for an edge σ gives either the flux conservativity (internal edges) or the no-flow boundary conditions (boundary edges). The final scheme for the pressure, which provides $p^{(n+1)} \in X_{\mathscr{T}}$, as well as fluxes $(F_{K,\sigma})_{K\in\mathscr{M},\,\sigma\in\mathscr{E}_K}$, is therefore

$$\sum_{\sigma\in\mathscr{E}_K} F_{K,\sigma} + |K| \sum_{M\in\mathscr{M}} |M| p_M = \int_K q. \quad (4)$$

$$\begin{aligned} F_{K,\sigma} + F_{L,\sigma} &= 0 \quad \text{for all edge } \sigma \text{ between two different cells } K \text{ and } L, \\ F_{K,\sigma} &= 0 \quad \text{for all edge } \sigma \text{ of } K \text{ lying on } \partial\Omega. \end{aligned} \quad (5)$$

2.2 Reconstruction of a Darcy Velocity

The ELLAM requires to compute the characteristics of the advective component of (1b), that is the solution, for each $\mathbf{x} \in \Omega$, to the ODEs

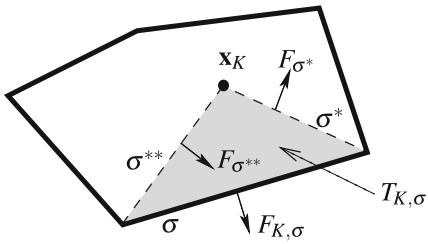


Fig. 1 Triangulation of a generic cell. Here, $s^{\sigma}_{\sigma^*} = +1$ and $s^{\sigma}_{\sigma^{**}} = -1$

$$\frac{d\hat{\mathbf{x}}}{dt}(t) = \frac{\mathbf{u}^{(n+1)}(\hat{\mathbf{x}}(t), t)}{\phi(\hat{\mathbf{x}}(t))}, \qquad \hat{\mathbf{x}}(t^{(n+1)}) = \mathbf{x}. \tag{6}$$

This obviously requires to reconstruct a Darcy velocity $\mathbf{u}^{(n+1)}$ everywhere. Two important features of this velocity need to be accounted for: the no-flow boundary conditions $\mathbf{u}^{(n+1)} \cdot \mathbf{n} = 0$ on $\partial\Omega$, which ensures that the solutions to (6) do not exit the computational domain, and the preservation of the divergence in (1a), to avoid creating regions with artificial wells or sinks (which lead to non-physical flows).

Preserving these features is done by using the technique of [7]. Each cell $K \in \mathscr{M}$ is split into triangles (see Fig. 1), and an oriented interior flux F_{σ^*} is computed on each internal edge created by this subdivision. Then, $\mathbf{u}^{(n+1)}$ is the $\mathbb{RT}_0$ function reconstructed from these fluxes on the triangular subdivision. This function belongs to $H_{\div}(\Omega)$ and, to ensure that its divergence is as dictated by (1a), the internal fluxes F_{σ^*} are constructed so that their balance (along with the fluxes $F_{K,\sigma}$ at the boundary of K) in each triangle corresponds to the balance over the cell K:

$$\forall \sigma \in \mathscr{E}_K\,, \quad \frac{1}{|T_{K,\sigma}|}\left(\sum_{\sigma^* \in \mathscr{E}^{*,\mathrm{int}}_{K,\sigma}} s^{\sigma}_{\sigma^*} F_{\sigma^*} + F_{K,\sigma}\right) = \frac{1}{|K|}\sum_{\sigma' \in \mathscr{E}_K} F_{K,\sigma'},$$

where $s^{\sigma}_{\sigma^*} = 1$ if F_{σ^*} is oriented outside $T_{K,\sigma}$ and -1 otherwise. This local system of equations is underdetermined. The chosen solution is that of minimal l^2 norm.

2.3 Numerical Scheme for the Concentration Equation

As with the pressure equation, we multiply the concentration equation (1b) by a test function v and perform integration by parts to obtain

$$\int_{t^n}^{t^{n+1}} \int_{\Omega} \left(\phi \frac{\partial(cv)}{\partial t} + D\nabla c \cdot \nabla v\right) - \int_{t^n}^{t^{n+1}} \int_{\Omega} c(\phi v_t + u \cdot \nabla v) = \int_{t^n}^{t^{n+1}} \int_{\Omega} q_c v. \tag{7}$$

The test functions are then selected to eliminate the advective term, and to match the piecewise constant functions at the core of the HMM method. We therefore take v

such that $\phi v_t + \mathbf{u}^{n+1} \cdot \nabla v = 0$ in the sense of distributions, and $v(t^{(n+1)}, \cdot) = \mathbf{1}_K = 1$ on K and 0 outside K. With characteristics computed through (6), this leads to $v(t^{(n)}, \mathbf{x}) = \mathbf{1}_K(\hat{\mathbf{x}}) = \mathbf{1}_{\hat{K}}(\mathbf{x})$, where $\hat{K}$ is K traced back from $t^{(n+1)}$ to $t^{(n)}$ through (6).

The diffusion term is discretised separately from the advective term, by using an implicit scheme. Fluxes $D_{K,\sigma}$ are defined as for the pressure equation (1a), using a piecewise Darcy velocity $\mathbf{u}^{(n+1)}$ given by the reconstructed pressure gradient (3) and the viscosity at $c^{(n)}$. For the source term, we also treat c implicitly. Without discretising the source term, this leads to the following scheme for (7):

$$\phi \int_K c^{n+1} - \phi \int_{\hat{K}} c^n + \Delta t \sum_{\sigma \in \mathscr{E}_K} D_{K,\sigma} = \int_{t^n}^{t^{n+1}} \int_\Omega q_{c^{n+1}} v.$$

For each cell K, the traceback region $\hat{K}$ is approximated in the following manner: for each of the vertices and edge midpoints of K, we solve (6) starting from $\mathbf{x} =$ that vertex or midpoint, we then compute $\hat{\mathbf{x}}(t^{(n)})$ and we approximate $\hat{K}$ by the polygon defined by these points $\hat{\mathbf{x}}(t^{(n)})$. The integrals are then computed by writing $\phi \int_K c^{n+1} = \phi |K| c_K^{n+1}$ and $\int_{\hat{K}} c^n = \sum_{M \in \mathscr{M}} |\hat{K} \cap M| c_M^n$, which leads to the following discretised form of the concentration equation

$$\phi |K| c_K^{n+1} + \Delta t \sum_{\sigma \in \mathscr{E}_K} D_{K,\sigma} = \phi \sum_{M \in \mathscr{M}} |\hat{K} \cap M| c_M^n + \int_{t^n}^{t^{n+1}} \int q_{c^{n+1}}.$$

2.4 The Integral of the Source Term $q_{c^{n+1}}$

For the integral involving the source term, we use a weighted trapezoid rule in time $\int_{t^n}^{t^{n+1}} \int q_{c^{n+1}} = w \int_{\hat{K}} q_{c^{n+1}} + (1 - w) \int_K q_{c^{n+1}}$. The left and right rules correspond to $w = 1$ and $w = 0$, respectively. Let E be an injection cell. A proper weight that will yield mass conservation has been derived for Cartesian meshes on [1]. The weight $w = (1 - e^{-\alpha})^{-1} - \alpha^{-1}$, where $\alpha = \Delta t \int_E q^{n+1} / \int_E \phi$, can easily be generalized for arbitrary meshes. A separate treatment will be made for cells that trace back into the injection well. These integrals will be computed using a forward tracing algorithm as described in [2].

3 Numerical Results

We take: $\Omega = (0, 1000) \times (0, 1000)$ ft^2; timestep of $\Delta t = 36$ days; injection well at (1000, 1000) and production well at (0, 0), both with flow rate of 30ft^2/day; constant

porosity ϕ=0.1; constant permeability tensor $\mathbf{K} = 80\mathbf{I}$; oil viscosity $\mu(0) = 1.0$ cp; mobility ratio $M = 41$; $\phi d_m = 0.0\text{ft}^2/\text{day}$, $\phi d_l = 5.0\text{ft}$, and $\phi d_t = 0.5\text{ft}$.

Figures 2 and 3 show the numerical solution for the concentration at $t = 3$ years on a Cartesian mesh using the left and the right rule, respectively. As can be seen here, the left rule provides us with an underestimate of the concentration at the injection well, and an overestimate somewhere along the neighborhood of the injection well. The right rule, implemented in [10], is also a bad choice since it provides us with an overshoot of the concentration at the injection well, as already proved for the MFEM-ELLAM in [9]. This is due to the fact that all of the source has been dumped into the injection well in one time step.

Figures 4 and 5 show the numerical solution for the concentration using the proper weight for the trapezoidal rule, as described in the previous section. These results present a significant improvement from those obtained through the right and left rule. Numerical results using Hexahedral meshes are presented in Figs. 6 and 7. The concentration spikes up along the boundary at $t = 10$ years. To mitigate this, the approximation of the traceback region is improved by using 3 points per edge (instead of only the edge midpoints); Figs. 8 and 9 show the significant improvement this enables. To understand more generally how many points to choose to obtain acceptable approximate traceback regions, we introduce the regularity parameter $m_{\text{reg}} = \max_{K\in\mathcal{M}}(\text{diam}(K)^2/|K|)$ of the mesh. It is then observed a reasonable numerical solution (overshoot $\leq 10\%$) is obtained by taking $\lceil \log_2(m_{\text{reg}})\rceil$ points along each edge, see Table 1. Further increasing the number of points per edge does not provide any significant improvement to our numerical solution. Finally, we show

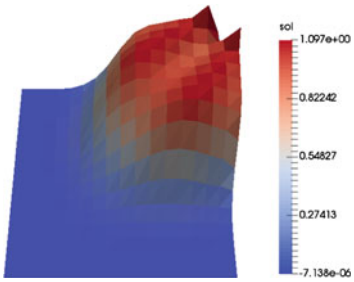


Fig. 2 Cartesian mesh, $t = 3$ years, *left rule* for source terms

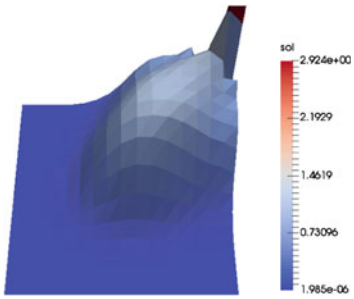


Fig. 3 Cartesian mesh, $t = 3$ years, *right rule* for source terms

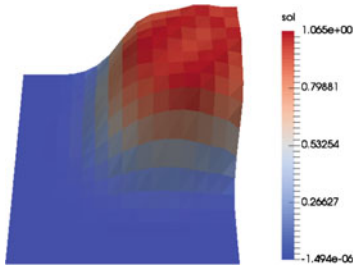


Fig. 4 Cartesian mesh, $t = 3$ years, weighted trapezoid rule for source terms

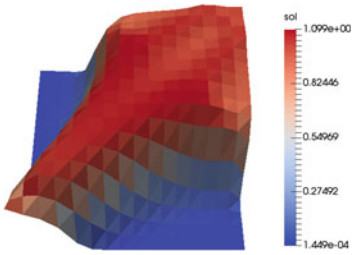


Fig. 5 Cartesian mesh, $t = 10$ years, weighted trapezoid rule for source terms

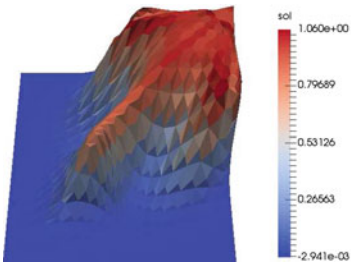


Fig. 6 Hexahedral mesh, $t = 3$ years, weighted trapezoid rule for source terms

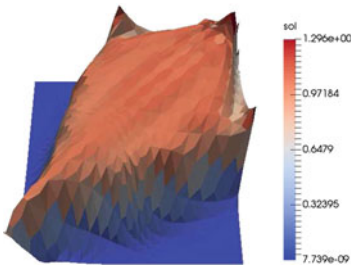


Fig. 7 Hexahedral mesh, $t = 10$ years, weighted trapezoid rule for source terms

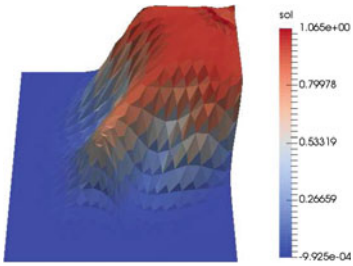


Fig. 8 Hexahedral mesh, $t = 3$ years, 3 points per edge, weighted trapezoid rule for source terms

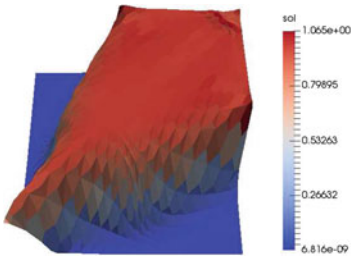


Fig. 9 Hexahedral mesh, $t = 10$ years, 3 points per edge, weighted trapezoid rule for source terms

Table 1 Regularity parameter of the meshes and nb of points to approximate the trace-back cells

Mesh	m_{reg}	$\log_2(m_{\text{reg}})$	Points per edge
Cartesian	2	1	1
Hexahedral	5.4772	2.4534	3
Kershaw	32.0274	5.0012	6

the numerical solutions for "Kershaw" meshes [6] on Figs. 10 and 11. We used here a proper quadrature rule for the source term, and an appropriate number of points per edge (see Table 1). The solutions on both Cartesian and hexahedral meshes are very similar, showing a certain robustness of the method with respect to the choice of mesh. The solution on the Kershaw mesh is noticeably different, due to the mesh being very distorted; this leads to a skewed approximation of the Darcy velocity, and thus a skewed advection of the fluid.

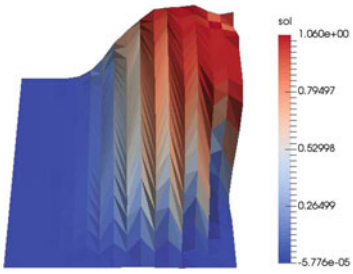


Fig. 10 Kershaw mesh, $t = 3$ years, 6 points per edge, weighted trapezoid rule for source terms

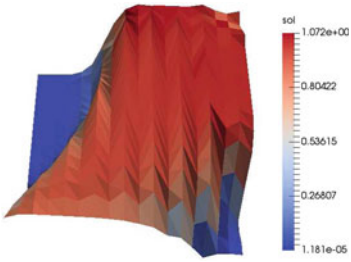


Fig. 11 Kershaw mesh, $t = 10$ years, 6 points per edge, weighted trapezoid rule for source terms

4 Summary

In previous work, the pressure equation (1a) and the concentration equation (1b) are often treated separately. This work presents a complete scheme for both equations, which is usable on generic meshes as encountered in real-world applications – with the usual caveats on distorted meshes. Our analysis demonstrates the importance of choosing a proper quadrature rule for integrating the source terms, and of selecting a correct number of approximation points – depending on the regularity of the mesh – to trace the cells. Further research will focus on reducing the grid effects on skewed meshes, and on finding better quadrature rules to deal with larger time steps.

Acknowledgements This work was supported by the ARC DP scheme (project DP170100605).

References

1. Arbogast, T., Huang, C.: A fully mass and volume conserving implementation of a characteristic method for transport problems. SIAM J. Sci. Comput. **28**(6), 20012022 (2006)
2. Arbogast, T., Wang, W.H.: Stability, monotonicity, maximum and minimum principles, and implementation of the volume corrected characteristic method. SIAM J. Sci. Comput. **33**(4), 15491573 (2011)
3. Chainais-Hillairet, C., Droniou, J.: Convergence analysis of a mixed finite volume scheme for an elliptic-parabolic system modeling miscible fluid flows in porous media. SIAM J. Numer. Anal. **45**(5), 2228–2258 (electronic) (2007)
4. Droniou, J., Eymard, R., Gallouët, T., Guichard, C., Herbin, R.: The gradient discretisation method (2016). https://hal.archives-ouvertes.fr/hal-01382358. Submitted
5. Droniou, J., Eymard, R., Gallouët, T., Herbin, R.: A unified approach to mimetic finite difference, hybrid finite volume and mixed finite volume methods. Math. Models Methods Appl. Sci. **20**(2), 265–295 (2010)
6. Herbin, R., Hubert, F.: Benchmark on discretization schemes for anisotropic diffusion problems on general grids. In: Finite Volumes for Complex Applications V, pp. 659–692. ISTE, London (2008)
7. Kuznetsov, Y., Repin, S.: New mixed finite element method on polygonal and polyhedral meshes. Russ. J. Numer. Anal. Math. Model. **18**(3) (2003)
8. Peaceman, D.W., Rachford Jr., H.H.: Numerical calculation of multidimensional miscible displacement. Soc. Pet. Eng. J. **2**(4), 327–339 (1962)

9. Sweeney, J.: Numerical methods for an oil recovery model (2015). Honours thesis, Monash University
10. Wang, H., Liang, D., Ewing, R.E., Lyons, S.L., Qin, G.: An approximation to miscible fluid flows in porous media with point sources and sinks by an Eulerian-Lagrangian localized adjoint method and mixed finite element methods. SIAM J. Sci. Comput. **22**(2), 561–581 (electronic) (2000)

Mixed Finite Volume Methods for Linear Elasticity

I. Ambartsumyan, E. Khattatov and I. Yotov

Abstract We present a new mixed finite element method for linear elasticity with weakly enforced stress symmetry on simplicial grids. Motivated by the multipoint flux mixed finite element method for Darcy flow, we consider a special quadrature rule that allows for elimination of the stress and rotation variables and leads to a cell-centered system for the displacements. Theoretical and numerical results indicate first-order convergence for all variables in the natural norms.

Keywords Mixed finite element · Finite volume · Multipoint stress · Elasticity

MSC (2010): 65N08 · 65N30 · 65N15 · 74S05 · 74S10

1 Problem Set up

We consider the static linear elasticity problem in a domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$. Given the body force vector field f on Ω, the stress σ and the displacement u satisfy the constitutive and equilibrium equations

$$A\sigma = \epsilon(u), \quad \text{div } \sigma = f \quad \text{in } \Omega. \tag{1}$$

I. Ambartsumyan · E. Khattatov · I. Yotov (✉)
Department of Mathematics, University of Pittsburgh, Pittsburgh 15260, USA
e-mail: yotov@math.pitt.edu

I. Ambartsumyan
e-mail: ILA6@pitt.edu

E. Khattatov
e-mail: ELK58@pitt.edu

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_40

Here $A(x)$ is the symmetric and positive definite compliance tensor, which in the case of an isotropic body is

$$A\sigma = \frac{1}{2\mu}\left(\sigma - \frac{\lambda}{2\mu + d\lambda}\,\mathrm{tr}(\sigma) I\right),$$

where I is the $d \times d$ identity matrix and $\mu(x) > 0$ and $\lambda(x) \geq 0$ are the Lamé coefficients, and $\epsilon(u) = \frac{1}{2}\left(\nabla u + (\nabla u)^T\right)$. The boundary conditions are $u = g \quad \text{on } \Gamma_D$, $\sigma n = 0 \quad \text{on } \Gamma_N$, where $\Gamma_D \cup \Gamma_N = \partial\Omega$, n is the outward unit normal vector on $\partial\Omega$, and we assume for simplicity that $\Gamma_D \neq \emptyset$. Throughout the paper div is the usual divergence for vector fields, and it produces a matrix field when applied to a matrix field by taking the divergence of each row.

Let $\mathbb{M}$ and $\mathbb{N}$ be the spaces of real $d \times d$ matrices and skew-symmetric matrices, respectively. Introducing the Lagrange multiplier γ, a skew-symmetric matrix representing the rotation, to penalize the asymmetry of the stress tensor, we arrive at the weak formulation for (1): find $(\sigma, u, \gamma) \in \mathbb{X} \times V \times \mathbb{W}$ such that

$$(A\sigma,\ \tau) + (u,\ \mathrm{div}\ \tau) + (\gamma,\ \tau) = \langle g,\ \tau n\rangle_{\Gamma_D}, \quad \forall \tau \in \mathbb{X}, \tag{2}$$

$$(\mathrm{div}\ \sigma,\ v) = (f,\ v), \quad \forall v \in V, \tag{3}$$

$$(\sigma,\ \xi) = 0, \quad \forall \xi \in \mathbb{W}, \tag{4}$$

where $\mathbb{X} = \left\{\tau \in H(\mathrm{div};\, \Omega, \mathbb{M}) : \tau n = 0 \text{ on } \Gamma_N\right\}$, $V = L^2(\Omega, \mathbb{R}^d)$, and $\mathbb{W} = L^2(\Omega, \mathbb{N})$. The reason for considering a mixed formulation with weak stress symmetry is that its lowest order mixed finite element approximation developed in [3] is suitable for a local stress elimination via a quadrature rule, resulting in a cell-centered system for displacement (and the rotation). This approach is motivated by the multipoint flux mixed finite element method [11], as well as the multipoint stress approximation [8, 9].

2 Numerical Approximation

Consider a polygonal domain $\Omega \in \mathbb{R}^d$ and let $\mathcal{T}_h$ be a finite element partition of Ω consisting of triangles in two dimensions and tetrahedra in three dimensions. For any element $E \in \mathcal{T}_h$ there exists a bijection mapping $F_E : \hat{E} \to E$, where $\hat{E}$ is a reference element. Denote the Jacobian matrix by DF_E and let $J_E = \det(DF_E)$. Let $\mathbb{X}_h \times V_h \times \mathbb{W}_h^l = (\mathrm{BDM}_1)^d \times (\mathrm{P}_0)^d \times (\mathrm{P}_l)^{d\times d, skew}$, $l = 0, 1 \subset \mathbb{X} \times V \times \mathbb{W}$, where BDM_1 is the lowest order Brezzi-Douglas-Marini space [5]. On the reference triangle these spaces are defined as

$$\hat{\mathbb{X}}_h(\hat{E}) = \mathrm{P}_1(\hat{E})^2 \times \mathrm{P}_1(\hat{E})^2 = \begin{pmatrix} \alpha_1\hat{x} + \beta_1\hat{y} + \gamma_1 & \alpha_2\hat{x} + \beta_2\hat{y} + \gamma_2 \\ \alpha_3\hat{x} + \beta_3\hat{y} + \gamma_3 & \alpha_4\hat{x} + \beta_4\hat{y} + \gamma_4 \end{pmatrix},$$

$$\hat{V}_h(\hat{E}) = \mathrm{P}_0(\hat{E}) \times \mathrm{P}_0(\hat{E}), \quad \hat{\mathbb{W}}_h^l(\hat{E}) = \begin{pmatrix} 0 & p \\ -p & 0 \end{pmatrix}, \ p \in \mathrm{P}_l(\hat{E}) \text{ for } l = 0, 1.$$

The definition of the spaces on tetrahedra is obtained naturally from the one above. The corresponding spaces on any element $E \in \mathscr{T}_h$ are defined via the transformations, for $\chi \in \mathbb{X}_h$, $v \in V_h$, and $w \in \mathbb{W}_h^l$,

$$\chi \leftrightarrow \hat{\chi} : \chi = \frac{1}{J_E} DF_E \hat{\chi} \circ F_E^{-1}, \quad v \leftrightarrow \hat{v} : v = \hat{v} \circ F_E^{-1}, \quad w \leftrightarrow \hat{w} : w = \hat{w} \circ F_E^{-1}.$$

The mixed finite element approximation of (2)–(4) is shown to be stable and first order accurate for all of variables in their natural norms in [3] ($l = 0$) and [7] ($l = 1$). The drawback is that the resulting algebraic system is a two-level saddle point system with three coupled variables, and thus expensive to solve. We next propose a quadrature rule that allows for local elimination of the stresses and rotations which leads to a cell-centered displacement-rotation, or further, displacement-only system.

A quadrature rule. We employ a trapezoidal-type quadrature rule for the stress bilinear form. For $\chi,\ \tau \in \mathbb{X}_h$, define

$$(A\chi, \tau)_Q = \sum_{E \in \mathscr{T}_h} (A\chi, \tau)_{Q,E}, \quad (A\chi, \tau)_{Q,E} = \frac{|E|}{s} \sum_{i=1}^{s} A(\mathbf{r}_i)\chi(\mathbf{r}_i) : \tau(\mathbf{r}_i),$$

where $s = 3$ on triangles and $s = 4$ on tetrahedra. In the case of linear rotations, a similar quadrature rule is employed for the stress-rotation bilinear forms.

Two multipoint stress mixed finite element methods. We seek $\sigma_h \in \mathbb{X}_h$, $u_h \in V_h$ and $\gamma_h \in \mathbb{W}_h^l$, $l = 0, 1$, such that

$$(A\sigma_h, \tau)_Q + (u_h, \operatorname{div} \tau) + (\gamma_h, \tau)_Q = \langle g, \tau n \rangle_{\Gamma_D}, \qquad \tau \in \mathbb{X}_h, \tag{5}$$

$$(\operatorname{div} \sigma_h, v) = (f, v), \qquad v \in V_h, \tag{6}$$

$$(\sigma_h, \xi)_Q = 0, \qquad \xi \in \mathbb{W}_h^l. \tag{7}$$

We note that the quadrature rule in $(\gamma_h, \tau)_Q$ and $(\sigma_h, \xi)_Q$ is applied only for $l = 1$. It is in fact exact for $l = 0$. We refer to the methods with $l = 0$ and $l = 1$ as the MSMFE-0 and the MSMFE-1 method, respectively. The well-posedness of (5)–(7) can be established using the classical Babuŝka-Brezzi conditions [6], which in our case are as follows:

(S1) There exists a constant $c > 0$ such that

$$c\|\tau\|_{\mathrm{div}}^2 \le (A\tau, \tau)_Q \text{ for } \tau \in \mathbb{X}_h \text{ s.t. } (\operatorname{div} \tau, v) + (\tau, \xi)_Q = 0, \ \forall (v, \xi) \in V_h \times \mathbb{W}_h^l,$$

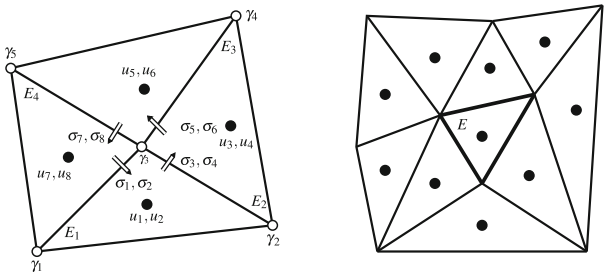


Fig. 1 Finite elements and stress degrees of freedom sharing a vertex (*left*) and cell-centered stencil (*right*)

(S2) There exists $\beta > 0$ such that

$$\inf_{0\neq(v,\xi)\in V_h\times\mathbb{W}_h^l}\ \sup_{0\neq\tau\in\mathbb{X}_h}\ \frac{(\operatorname{div}\tau,\ v)+(\tau,\ \xi)_Q}{\|\tau\|_{\operatorname{div}}\left(\|v\|+\|\xi\|\right)}\geq\beta.$$

Here $\|\cdot\|_{\operatorname{div}}$ and $\|\cdot\|$ denote the $H(\operatorname{div};\Omega)$ and the $L^2(\Omega)$ norms, respectively. Condition **(S1)** can be easily established by showing that $(A\tau,\ \tau)_Q$ is equivalent to $\|\tau\|^2$, see, e.g., [11]. Condition **(S2)** has been shown in [3, 4] for $l=0$ and in [2] for $l=1$. The latter case is challenging, due to the presence of the quadrature rule. It requires inf-sup stability for the bilinear form $(\operatorname{div} q,\ w)_Q$ for the Taylor-Hood $\mathrm{P}_2-\mathrm{P}_1$ spaces. This is shown in [2] using a macro-element argument motivated by [10].

Reduction to a cell-centered scheme. The algebraic system that arises from the (5)–(7) is of the form

$$\begin{pmatrix} A_{\sigma\sigma} & A_{\sigma u}^T & A_{\sigma\gamma}^T \\ A_{\sigma u} & 0 & 0 \\ A_{\sigma\gamma} & 0 & 0 \end{pmatrix}\begin{pmatrix}\sigma\\ u\\ \gamma\end{pmatrix}=\begin{pmatrix} g\\ f\\ 0\end{pmatrix},$$

where $(A_{\sigma\sigma})_{ij}=(A\tau_i,\tau_j)_Q$, $(A_{\sigma u})_{ij}=(\operatorname{div}\tau_i,v_j)$ and $(A_{\sigma\gamma})_{ij}=(\tau_i,\xi_j)_Q$. It is known [5, 6] that the degrees of freedom for BDM_1 can be chosen to be the values of the normal fluxes at any d points on each edge (face) $\hat{e}$ of the reference element $\hat{E}$. This naturally extends to normal stresses in our case of $\mathrm{BDM}_1\times\mathrm{BDM}_1$. We choose these points to be the vertices of $\hat{e}$. Let us consider any interior vertex $\mathbf{r}$ and suppose that it is shared by l elements $E_1,\ldots,E_m$ as shown in Fig. 1 with $m=4$. Let $e_1,\ldots,e_k$ be the edges (faces) that share the vertex $\mathbf{r}$ and let $\tau_1,\ldots,\tau_{dk}$, be the stress basis functions on these edges (faces) associated with the vertex $\mathbf{r}$. Denote the corresponding values of the normal components of σ_h by $\sigma_1,\ldots,\sigma_{dk}$. Note that for for the sake of clarity the normal stresses are drawn at a distance from the vertex. The quadrature rule $(A\tau_i,\tau_j)_Q$ decouples $\sigma_1,\ldots,\sigma_{dk}$ from the rest of the stress degrees of freedom. As a result, the matrix $A_{\sigma\sigma}$ is block-diagonal with $dk\times dk$ blocks associated with

the mesh vertices. Due to **(S1)**, these local blocks are symmetric and positive definite blocks. Hence, eliminating σ leads to a cell-centered displacement-rotation system

$$\begin{pmatrix} A_{\sigma u}A_{\sigma\sigma}^{-1}A_{\sigma u}^T & A_{\sigma u}A_{\sigma\sigma}^{-1}A_{\sigma\gamma}^T \\ A_{\sigma\gamma}A_{\sigma\sigma}^{-1}A_{\sigma u}^T & A_{\sigma\gamma}A_{\sigma\sigma}^{-1}A_{\sigma\gamma}^T \end{pmatrix} \begin{pmatrix} u \\ \gamma \end{pmatrix} = \begin{pmatrix} \tilde{f} \\ \tilde{h} \end{pmatrix}. \tag{8}$$

Further reduction is possible in the case of MSMFE-1, $l = 1$, where the quadrature rule $(\tau_i, \xi_j)_Q$ results in a block-diagonal rotation matrix $A_{\sigma\gamma}A_{\sigma\sigma}^{-1}A_{\sigma\gamma}^T$ with $d(d-1)/2 \times d(d-1)/2$ blocks associated with mesh vertices. It is symmetric and positive definite, due to **(S2)**. Hence, local elimination of the rotation in (8) results in a displacement-only cell-centered system

$$(A_{\sigma u}A_{\sigma\sigma}^{-1}A_{\sigma u}^T - A_{\sigma u}A_{\sigma\sigma}^{-1}A_{\sigma\gamma}^T(A_{\sigma\gamma}A_{\sigma\sigma}^{-1}A_{\sigma\gamma}^T)^{-1}A_{\sigma\gamma}A_{\sigma\sigma}^{-1}A_{\sigma u}^T)u = \hat{f}. \tag{9}$$

The cell-centered stencil in (8) and (9) is shown in Fig. 1 (right). The displacements (and rotations) in each element E are coupled to the displacements (and rotations) in all elements that share a vertex with E.

Error estimates. As shown above, the MSMFE-0 and MSMFE-1 methods allow to eliminate locally the stress (and rotation) variables, thus significantly reducing the size of the global problem. The following result from [2] addresses the accuracy of the methods. First-order convergence is obtained for all variables in their natural norms. Moreover, the computed displacement is h^2-close to the true displacement when measured at the center of mass of each element.

Theorem 1 *If $A \in W^{1,\infty}(\Omega)$, then*

$$\|\sigma - \sigma_h\| + \|u - u_h\| + \|\gamma - \gamma_h\| \le Ch(\|\sigma\|_1 + \|u\|_1 + \|\gamma\|_1). \tag{10}$$

Moreover, if $A \in W^{2,\infty}(\Omega)$, then

$$\|Q_h u - u_h\| \le Ch^2 \left(\|\sigma\|_1 + \|\operatorname{div} \sigma\|_1 + \|\gamma\|_1\right), \tag{11}$$

where Q_h denotes the L^2-projection onto the space V_h.

3 Numerical Results

We present several numerical tests confirming the theoretical convergence rates and illustrating the behavior of the method. The first two examples test convergence on the unit hypercube in 2 and 3 dimensions, respectively, while the last example models a pulley under centripetal load. For the first two examples, the boundary conditions are Dirichlet on the entire boundary and the analytical solution is given. All tests have been performed using the FEniCS finite element package [1].

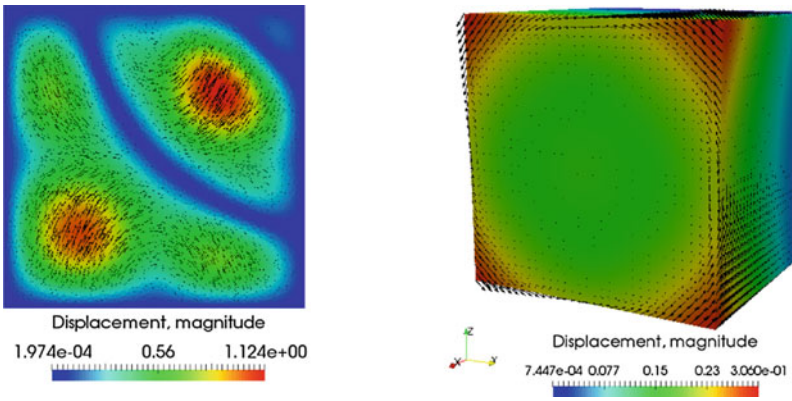


Fig. 2 Computed displacement on the finest mesh with the MSMFE-1 method in Example 1 *left* and Example 2 *right*

Table 1 Relative errors computed with the MSMFE-0 method for Example 1

h	$\|\sigma - \sigma_h\|_{\text{div}}$	Rate	$\|u - u_h\|$	Rate	$\|Q_h u - u_h\|$	Rate	$\|\gamma - \gamma_h\|$	Rate
1/2	2.26E+00	–	2.18E+00	–	1.72E+00	–	1.68E+00	–
1/4	1.57E+00	0.5	7.45E-01	1.6	7.02E-01	1.7	9.32E-01	0.9
1/8	6.70E-01	1.2	2.67E-01	1.5	1.43E-01	2.4	5.42E-01	0.8
1/16	2.99E-01	1.2	1.20E-01	1.2	2.94E-02	2.3	2.97E-01	0.9
1/32	1.48E-01	1.0	6.03E-02	1.0	7.80E-03	1.9	1.54E-01	1.0
1/64	7.37E-02	1.0	3.03E-02	1.0	2.03E-03	1.9	7.80E-02	1.0

Example 1 We solve problem (1) with a given displacement solution

$$u = \begin{pmatrix} \cos(2\pi xy)\sin(\pi x)\sin(\pi y) \\ \cos(2\pi xy)\sin(2\pi x)\sin(2\pi y) \end{pmatrix}$$

on the unit square. The Lamé parameters are set to be $\lambda = 123.0$ and $\mu = 79.3$. The computational grid is obtained by a Delauney triangulation of the given domain on several levels of refinement. The computed displacement with the MSMFE-1 method on the finest level $h = 1/64$ is shown in Fig. 2 (left). We present the relative errors and convergence rates for the MSMFE-0 and MSMFE-1 methods in Tables 1 and 2, respectively. As expected from the theory, both methods exhibit at least linear convergence for all variables in their natural norms, with a quadratic convergence for the displacement error computed at the cell-centers. We also note that the rotation error converges with order $O(h^{1.5})$ for the MSMFE-1 method, which is due to the use of a P_1 finite element space for this variable.

Table 2 Relative errors computed with MSMFE-1 method for Example 1

h	$\|\sigma - \sigma_h\|_{\text{div}}$	Rate	$\|u - u_h\|$	Rate	$\|Q_h u - u_h\|$	Rate	$\|\gamma - \gamma_h\|$	Rate
1/2	2.26E+00	–	2.49E+00	–	1.92E+00	–	1.28E+00	–
1/4	1.57E+00	0.5	7.50E-01	1.7	7.10E-01	1.8	7.77E-01	0.7
1/8	6.77E-01	1.2	2.68E-01	1.5	1.48E-01	2.3	3.39E-01	1.2
1/16	3.01E-01	1.2	1.20E-01	1.2	3.12E-02	2.2	1.11E-01	1.6
1/32	1.48E-01	1.0	6.04E-02	1.0	8.64E-03	1.9	3.95E-02	1.5
1/64	7.40E-02	1.0	3.03E-02	1.0	2.27E-03	1.9	1.50E-02	1.4

Table 3 Relative errors computed with MSMFE-0 method for Example 2

h	$\|\sigma - \sigma_h\|_{\text{div}}$	Rate	$\|u - u_h\|$	Rate	$\|Q_h u - u_h\|$	Rate	$\|\gamma - \gamma_h\|$	Rate
1/2	1.42E+00	–	4.88E-01	0.0	3.09E-01	–	4.53E-01	0.0
1/4	7.06E-01	1.0	2.22E-01	1.1	7.23E-02	2.1	2.14E-01	1.1
1/8	3.47E-01	1.0	1.08E-01	1.0	1.78E-02	2.0	1.03E-01	1.1
1/16	1.72E-01	1.0	5.37E-02	1.0	4.43E-03	2.0	5.07E-02	1.0
1/32	8.59E-02	1.0	2.68E-02	1.0	1.11E-03	2.0	2.52E-02	1.0

Table 4 Relative errors computed with MSMFE-1 method for Example 2

h	$\|\sigma - \sigma_h\|_{\text{div}}$	Rate	$\|u - u_h\|$	Rate	$\|Q_h u - u_h\|$	Rate	$\|\gamma - \gamma_h\|$	Rate
1/2	1.47E+00	–	4.82E-01	0.0	2.98E-01	–	3.53E-01	0.0
1/4	7.32E-01	1.0	2.22E-01	1.1	7.37E-02	2.0	1.49E-01	1.2
1/8	3.61E-01	1.0	1.08E-01	1.0	1.90E-02	2.0	5.68E-02	1.4
1/16	1.78E-01	1.0	5.37E-02	1.0	4.86E-03	2.0	2.04E-02	1.5
1/32	8.83E-02	1.0	2.68E-02	1.0	1.23E-03	2.0	7.21E-03	1.5

Example 2 For the second test, we model a simultaneous twisting and compression of the unit cube, with a given displacement solution:

$$\begin{pmatrix} -0.1(e^x - 1)\sin(\pi x)\sin(\pi y) \\ -(e^x - 1)(y - \cos(\frac{\pi}{12})(y - 0.5) + \sin(\frac{\pi}{12})(z - 0.5) - 0.5) \\ -(e^x - 1)(z - \sin(\frac{\pi}{12})(y - 0.5) - \cos(\frac{\pi}{12})(z - 0.5) - 0.5) \end{pmatrix}.$$

We allow the Young's modulus to change over the domain as $E = e^{4x}$ and take the Poisson ratio $\nu = 0.2$. The Lamé parameters are obtained from the relationships $\lambda = \frac{E\nu}{(1+\nu)(1-2\nu)}$ and $\mu = \frac{E}{2(1+\nu)}$. The computed displacement with the MSMFE-1 method on the finest level $h = 1/32$ is shown in Fig. 2 (right). The relative errors and convergence rates obtained by the MSMFE-0 and MSMFE-1 methods, as shown in Tables 3 and 4, respectively, behave similarly to the two-dimensional Example 1 (Tables 1 and 2).

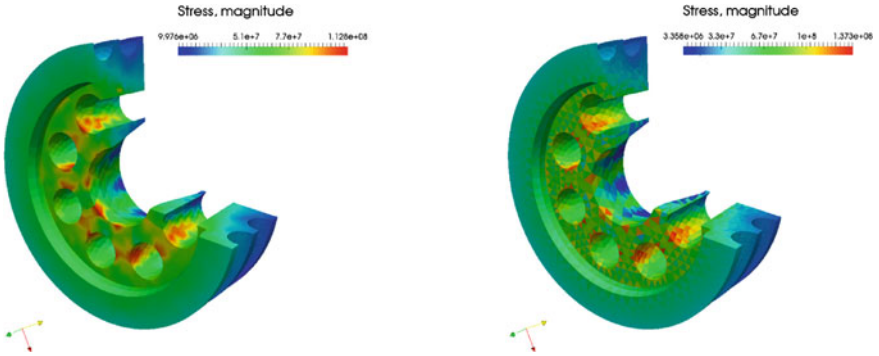


Fig. 3 Stress solution magnitude of the MSMFE-1 method *left* and a non-mixed method *right* for Example 3

Example 3 The last example is from the FEniCS repository and it models a pulley under centripetal load given by

$$f = \left(\rho\omega^2 x,\ \rho\omega^2 y,\ 0\right)',$$

with rotation rate $\omega = 10\ (rad/s)$ and mass density $\rho = 300\ (kg/m^3)$. The displacement is set to be zero at the shaft of the pulley, while zero traction boundary conditions are enforced on the rest of the boundary. We compare the computed stress of the MSMFE-1 method to the one obtained by solving the classical displacement formulation for linear elasticity, with a post-processing step to recover the stress, see Fig. 3). For the sake of space, we omit the displacement solutions as they are visually of equal quality. However, we observe smoother approximation of the stress variable provided by the MSMFE-1 method, as the method computes a locally conservative H-div approximation and does not require a numerical differentiation, hence avoiding extra loss of accuracy.

References

1. Alnæs, M.S., Blechta, J., Hake, J., Johansson, A., Kehlet, B., Logg, A., Richardson, C., Ring, J., Rognes, M.E., Wells, G.N.: The FEniCS project version 1.5. Archive of Numerical Software **3**(100) (2015). doi:10.11588/ans.2015.100.20553
2. Ambartsumyan, I., Khattatov, E., Nordbotten, J., Yotov, I.: A multipoint stress mixed finite element method for elasticity I: Simplicial grids (2016). Preprint
3. Arnold, D., Falk, R., Winther, R.: Mixed finite element methods for linear elasticity with weakly imposed symmetry. Math. Comput. **76**(260), 1699–1723 (2007)
4. Brezzi, F., Boffi, D., Demkowicz, L., Durán, R., Falk, R., Fortin, M.: Mixed Finite Elements, Compatibility Conditions, and Applications. Springer (2008)
5. Brezzi, F., Douglas Jr., J., Marini, L.D.: Two families of mixed finite elements for second order elliptic problems. Numer. Math. **47**(2), 217–235 (1985)

6. Brezzi, F., Fortin, M.: Mixed Hybrid Finite Element Methods. Springer Series in Computational Mathematics, vol. 15. Springer, Berlin (1991)
7. Cockburn, B., Gopalakrishnan, J., Guzmán, J.: A new elasticity element made for enforcing weak stress symmetry. Math. Comput. **79**(271), 1331–1349 (2010)
8. Nordbotten, J.M.: Cell-centered finite volume discretizations for deformable porous media. Int. J. Numer. Methods Eng. **100**(6), 399–418 (2014). doi:10.1002/nme.4734. http://dx.doi.org/10.1002/nme.4734
9. Nordbotten, J.M.: Convergence of a cell-centered finite volume discretization for linear elasticity. SIAM J. Numer. Anal. **53**(6), 2605–2625 (2015)
10. Stenberg, R.: Analysis of mixed finite elements methods for the Stokes problem: a unified approach. Math. Comput. **42**(165), 9–23 (1984)
11. Wheeler, M.F., Yotov, I.: A multipoint flux mixed finite element method. SIAM J. Numer. Anal. **44**(5), 2082–2106 (2006)

A Nonlinear Domain Decomposition Method to Couple Compositional Gas Liquid Darcy and Free Gas Flows

Nabil Birgle, Roland Masson and Laurent Trenty

Abstract A domain decomposition algorithm is proposed to couple at the interface a gas liquid compositional Darcy flow and a compositional free gas flow. At each time step, our algorithm solves iteratively the nonlinear system coupling the compositional Darcy flow in the porous medium, the RANS gas flow in the free flow domain, and the convection diffusion of the species in the free flow domain. In order to speed up the convergence of the algorithm, the transmission conditions at the interface are replaced by Robin boundary conditions. Each Robin coefficient is obtained from a diagonal approximation of the Dirichlet to Neumann operator related to a scalar simplified model in the neighbouring subdomain. The efficiency of our domain decomposition algorithm is assessed in the case of the modelling of the mass exchanges at the interface between the geological formation and the ventilation galleries of geological radioactive waste disposal.

Keywords Drying model · Coupling algorithm · Nonlinear domain decomposition method · Compositional gas liquid darcy flow · Free gas flow

1 Formulation of the Coupled Model

Let us denote by $\Omega_{\rm pm}$ the porous medium domain, by $\Omega_{\rm ff}$ the free flow domain and by $\Gamma = \partial\Omega_{\rm pm} \cap \partial\Omega_{\rm ff}$ the interface. Let $\mathscr{P} = \{g, \ell\}$ denote the set of gas and liquid phases assumed to be both defined by a mixture of components $i \in \mathscr{C}$ among which

N. Birgle (✉) · R. Masson
Laboratoire J.A. Dieudonné, Team Coffee, Université d'Azur, Inria, CNRS,
Parc Valrose, 06108 Nice Cedex 02, France
e-mail: nabil.birgle@unice.fr

R. Masson
e-mail: roland.masson@unice.fr

L. Trenty
Andra, 1-7 Rue Jean Monnet, 92290 Chatenay-malabry, France
e-mail: laurent.trenty@andra.fr

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_41

the water component denoted by w which can vaporize in the gas phase, and a set of gaseous components $j \in \mathscr{C} \setminus \{w\}$ which can dissolve in the liquid phase. The model is assumed to be isothermal with a fixed temperature T.

Compositional Darcy flow in the porous medium Ω_{pm}: following [5], the liquid gas Darcy flow formulation uses the gas pressure p^g, the liquid pressure p^ℓ, and the component fugacities $f = (f_i)_{i\in\mathscr{C}}$ as primary unknowns. In this formulation, following [4], the component molar fractions $c^\alpha = (c_i^\alpha)_{i\in\mathscr{C}}$ of each phase $\alpha \in \mathscr{P}$ are the functions $c^\alpha(p^\alpha, f)$ defined by inversion of the equations $f_i^\alpha(c^\alpha, p^\alpha) = f_i$, $i \in \mathscr{C}$, where f_i^α is the fugacity of the component i in the phase α. In addition, for $\alpha \in \mathscr{P}$, the phase pressure p^α is extended in the absence of the phase in such a way that the closure law $\sum_{i\in\mathscr{C}} c_i^\alpha(p^\alpha, f) = 1$ is always imposed (see [5]). The phase molar and mass densities, as well as the phase viscosities functions are denoted in the following by respectively $\zeta^\alpha(p^\alpha, c^\alpha)$, $\rho^\alpha(p^\alpha, c^\alpha)$, $\mu^\alpha(p^\alpha, c^\alpha)$ for $\alpha \in \mathscr{P}$. Finally, we define the liquid saturation as the function $s^\ell(\mathbf{x}, p^g - p^\ell)$ defined by the inverse of the monotone graph extension of the capillary pressure function, and we set $s^g(\mathbf{x}, .) = 1 - s^\ell(\mathbf{x}, .)$.

Let us define the two-phase Darcy velocities $\mathbf{u}^\alpha = \frac{-k_r^\alpha}{\mu^\alpha}\mathbf{K}(\nabla p^\alpha - \rho^\alpha \mathbf{g})$ where $k_r^\alpha(\mathbf{x}, s^\alpha)$ is the phase relative permeability, $\mathbf{K}(\mathbf{x})$ the porous medium permeability tensor, and $\mathbf{g}$ the gravitational acceleration vector. Let us also introduce, for each component $i \in \mathscr{C}$, the total number of mole per unit pore volume $\mathrm{n}_i = \sum_{\alpha\in\mathscr{P}} s^\alpha \zeta^\alpha c_i^\alpha$ and the component molar flow rate $\mathbf{v}_i = \sum_{\alpha\in\mathscr{P}} \zeta^\alpha c_i^\alpha \mathbf{u}^\alpha$. The model in Ω_{pm} accounts for the mole conservation of each component $i \in \mathscr{C}$ coupled with the sum to 1 of the molar fractions for each phase $\alpha \in \mathscr{P}$:

$$\begin{aligned} \phi\partial_t \mathrm{n}_i + \nabla\cdot\mathbf{v}_i = 0,\ i \in \mathscr{C}, \quad &\text{on } \Omega_{\text{pm}} \times (0, t_{\text{f}}), \\ \sum_i c_i^\alpha = 1,\ \alpha \in \mathscr{P}, \ &\text{on } \Omega_{\text{pm}} \times (0, t_{\text{f}}), \end{aligned} \tag{1}$$

where $\phi(\mathbf{x})$ is the porous medium porosity and $(0, t_{\text{f}})$ the simulation time interval.

Flow and transport model in the free flow domain Ω_{ff}: the primary unknowns in the free flow domain are defined by the gas pressure p, the gas molar fractions $c = (c_i)_{i\in\mathscr{C}}$, and the gas velocity $\mathbf{u}$. The flow is described by a Reynolds Averaged Navier–Stokes (RANS) model and assumed to be quasi-stationary at the time scale of the porous medium. Let us first define the uncoupled mean turbulent flow as the solution $(\mathbf{u}^0, p^0)$ of the following RANS model given the initial gas molar fractions c^0:

$$\begin{aligned} \nabla\cdot(\rho^g(p^0, c^0)\mathbf{u}^0 \otimes \mathbf{u}^0 - \mu_{\text{t}}^0(\nabla\mathbf{u}^0 + \nabla^t\mathbf{u}^0)) + \nabla p^0 = \rho^g(p^0, c^0)\mathbf{g}, \quad &\text{in } \Omega_{\text{ff}}, \\ \nabla\cdot(\zeta^g(p^0, c^0)\mathbf{u}^0) = 0, \quad &\text{in } \Omega_{\text{ff}}, \\ \mathbf{u}^0 = 0, \quad &\text{on } \Gamma, \end{aligned} \tag{2}$$

where the turbulent viscosity $\mu_{\text{t}}^0(\mathbf{x})$ is obtained using an algebraic turbulent model or a more advanced $k - \varepsilon$ model [1]. This turbulent flow is responsible for a turbulent diffusivity denoted by $d_{\text{t}}^0(\mathbf{x})$ typically depending on the turbulent viscosity, on the gas

Fickian diffusion and on the Schmidt number. Assuming that the velocity perturbation induced by the coupling is small compared to the flow velocity, the turbulent viscosity μ_t^0 and diffusivity d_t^0 are used for the coupled model. Thus, the primary unknowns $\mathbf{u}$, p, c in the free flow domain satisfy the following system of equations

$$\begin{aligned} \nabla \cdot (\rho^g(p,c)\mathbf{u} \otimes \mathbf{u} - \mu_t^0(\nabla \mathbf{u} + \nabla^t \mathbf{u})) + \nabla p &= \rho^g(p,c)\mathbf{g}, \quad \text{on } \Omega_{\mathrm{ff}} \times (0, t_{\mathrm{f}}), \\ \nabla \cdot (\zeta^g(p,c)\mathbf{u}) &= 0, \quad \text{on } \Omega_{\mathrm{ff}} \times (0, t_{\mathrm{f}}), \\ \nabla \cdot \mathbf{w}_i &= 0,\ i \in \mathscr{C}, \quad \text{on } \Omega_{\mathrm{ff}} \times (0, t_{\mathrm{f}}), \end{aligned} \tag{3}$$

with the component molar flow rate $\mathbf{w}_i = \zeta^g(p,c)(c_i\mathbf{u} - d_t^0 \nabla c_i)$.

Transmission conditions at the interface Γ: at the interface Γ between the free-flow domain and the porous medium, the coupling conditions are those stated in [1, 6, 7] where we have replaced the Beaver Joseph condition by the simpler no slip condition due to the low permeability of the porous medium in our application. They state the gas molar fraction and molar normal flux continuity, the gas liquid thermodynamical equilibrium, the no slip condition, and the normal component of the normal stress continuity as follows:

$$\begin{aligned} c_i^g &= c_i, \quad \text{on } \Gamma \times (0, t_{\mathrm{f}}), \\ \mathbf{v}_i \cdot \mathbf{n}_{\mathrm{pm}} + \mathbf{w}_i \cdot \mathbf{n}_{\mathrm{ff}} &= 0, \quad \text{on } \Gamma \times (0, t_{\mathrm{f}}), \\ \sum_i c_i^\alpha &= 1, \quad \text{on } \Gamma \times (0, t_{\mathrm{f}}), \\ \left(\rho^g \mathbf{u} \otimes \mathbf{u} - \mu_t^0(\nabla \mathbf{u} + \nabla^t \mathbf{u})\right) \mathbf{n}_{\mathrm{ff}} \cdot \mathbf{n}_{\mathrm{ff}} + p &= p^g, \quad \text{on } \Gamma \times (0, t_{\mathrm{f}}), \\ \mathbf{u} \cdot \boldsymbol{\tau} &= 0, \quad \text{on } \Gamma \times (0, t_{\mathrm{f}}), \end{aligned} \tag{4}$$

with $i \in \mathscr{C}, \alpha \in \mathscr{P}$, and where $\mathbf{n}_{\mathrm{pm}}$ and $\mathbf{n}_{\mathrm{ff}}$ are the unit normal vectors at the interface Γ oriented outward from the porous medium domain and the free flow domain respectively.

2 Domain Decomposition Algorithm

The coupled model (1), (3) and (4) is integrated in time using an Euler implicit scheme which leads to solve at each time step a fully coupled nonlinear system. This nonlinear system is solved using a domain decomposition algorithm detailed below. This approach has two advantages. Firstly it allows to use different codes for the porous medium and the free flow problems. Secondly, it reduces the complexity of the nonlinear and linear systems which results in a better efficiency compared with a monolithic Newton algorithm solving the fully coupled system [1, 7].

In the following, the time step count n is omitted for the sake of clarity and the component total number of mole in the porous medium at the previous time step is denoted by n_i^{n-1}. The domain decomposition count is denoted by the superscript k. As usual, the algorithm is initialized by the previous time step solution. The algorithm

solves iteratively, until convergence to the fully coupled solution, the compositional gas liquid flow in the porous medium with Robin type transmission conditions, the RANS model in the free flow domain and the convection diffusion equations in the free flow domain with Robin type transmission conditions.

Porous medium flow with Robin boundary conditions on Γ: compute the phase pressures $p^{\alpha,k}$, $\alpha \in \mathscr{P}$, the fugacity vector f^k in the porous medium Ω_{pm} and a normal velocity correction denoted by $\delta_{\mathbf{u}}^k$ at the interface Γ and oriented outward to the free flow domain, such that

$$\begin{aligned}
\tfrac{\phi}{\Delta t^n}(\mathrm{n}_i^k - \mathrm{n}_i^{n-1}) + \nabla \cdot \mathbf{v}_i^k &= 0, \quad \text{in } \Omega_{\mathrm{pm}},\\
\sum_i c_i^{\alpha,k} &= 1, \quad \text{in } \Omega_{\mathrm{pm}},\\
\beta_{\mathrm{pm}} c_i^{g,k} - \mathbf{v}_i^k \cdot \mathbf{n}_{\mathrm{pm}} - c_i^{g,k}\zeta_{\mathrm{pm}}^{g,k}\delta_{\mathbf{u}}^k &= \beta_{\mathrm{pm}} c_i^{k-1} - \mathbf{w}_i^{k-1} \cdot \mathbf{n}_{\mathrm{pm}}, \quad \text{on } \Gamma,\\
p^{g,k} &= p^{k-1} + (\rho_{\mathrm{ff}}^g \mathbf{u} \otimes \mathbf{u} - \mu_{\mathrm{t}}^0(\nabla \mathbf{u} + \nabla^t \mathbf{u}))^{k-1}\mathbf{n}_{\mathrm{ff}} \cdot \mathbf{n}_{\mathrm{ff}}, \quad \text{on } \Gamma,\\
\sum_i c_i^{\alpha,k} &= 1, \quad \text{on } \Gamma.
\end{aligned} \tag{5}$$

with $i \in \mathscr{C}$, $\alpha \in \mathscr{P}$ and $\zeta_{\mathrm{pm}}^{g,k} = \zeta^g(p^{g,k}, c^{g,k})$. Note that the additional unknown $\delta_{\mathbf{u}}^k$ accounts for the correction of the normal gas velocity $\mathbf{u}^{k-1} \cdot \mathbf{n}_{\mathrm{ff}}$ at the interface induced by the coupling with the porous medium.

RANS flow with Dirichlet boundary condition on Γ: compute the pressure p^k and the gas velocity $\mathbf{u}^k$ such that

$$\begin{aligned}
\nabla \cdot (\rho_{\mathrm{ff}}^{g,k}\mathbf{u}^k \otimes \mathbf{u}^k - \mu_{\mathrm{t}}^0(\nabla \mathbf{u}^k + \nabla^t \mathbf{u}^k)) + \nabla p^k &= \rho_{\mathrm{ff}}^{g,k}\mathbf{g}, \quad \text{in } \Omega_{\mathrm{ff}},\\
\nabla \cdot (\zeta_{\mathrm{ff}}^{g,k}\mathbf{u}^k) &= 0, \quad \text{in } \Omega_{\mathrm{ff}},\\
\zeta_{\mathrm{ff}}^{g,k}\mathbf{u}^k &= \zeta_{\mathrm{ff}}^{g,k-1}\mathbf{u}^{k-1} + \zeta_{\mathrm{pm}}^{g,k}\delta_{\mathbf{u}}^k\,\mathbf{n}_{\mathrm{ff}}, \quad \text{on } \Gamma,
\end{aligned} \tag{6}$$

with $\zeta_{\mathrm{ff}}^{g,k} = \zeta^g(p^k, c^{k-1})$ and $\rho_{\mathrm{ff}}^{g,k} = \rho^g(p^k, c^{k-1})$.

Convection diffusion equations with Robin boundary conditions on Γ: compute c^k such that for all $i \in \mathscr{C}$

$$\begin{aligned}
\nabla \cdot \mathbf{w}_i^k &= 0, \quad \text{in } \Omega_{\mathrm{ff}},\\
\beta_{\mathrm{ff}} c_i^k - \mathbf{w}_i^k \cdot \mathbf{n}_{\mathrm{ff}} &= \beta_{\mathrm{ff}} c_i^{g,k} - \mathbf{v}_i^k \cdot \mathbf{n}_{\mathrm{ff}}, \quad \text{on } \Gamma,
\end{aligned} \tag{7}$$

with $\mathbf{w}_i^k = \zeta_{\mathrm{ff}}^{g,k}(c_i^k \mathbf{u}^k - d_{\mathrm{t}}^0 \nabla c_i^k)$.

The domain decomposition algorithm is iterated until the following stopping criterion at the interface Γ is satisfied for a given tolerance ε:

$$\frac{\sum_{i\in\mathscr{C}}\|c_i^{g,k}-c_i^k\|}{\sum_{i\in\mathscr{C}}\|c_i^k\|}+\frac{\sum_{i\in\mathscr{C}}\|(\mathbf{v}_i^k-\mathbf{w}_i^k)\cdot\mathbf{n}_{\mathrm{ff}}\|}{\sum_{i\in\mathscr{C}}\|\mathbf{w}_i^k\cdot\mathbf{n}_{\mathrm{ff}}\|}+\frac{\|\delta_{\mathbf{u}}^k\|}{\|\mathbf{u}^k\cdot\mathbf{n}_{\mathrm{ff}}\|}<\varepsilon. \tag{8}$$

2.1 Computation of the Robin Coefficients β_{pm} and β_{ff}

To speedup the convergence of the domain decomposition method, the Robin coefficients β_{pm} and β_{ff} of each subdomain must approximate the Dirichlet to Neumann (DtN) operator of the neighbouring subdomain problem [3]. For this purpose, a simplified scalar model is defined in each subdomain and a low frequency diagonal approximation of its DtN operator is built.

To compute β_{pm}, the convection diffusion equation in (3) is approximated by using the uncoupled velocity $\mathbf{u}^0$ from (2) and by neglecting the variations of the gas molar density. We end up with a linear operator $\mathscr{L}_{\mathrm{ff}}c=\zeta^g(p^0,c^0)\nabla\cdot(c\mathbf{u}^0-d_{\mathrm{t}}^0\nabla c)$ independent on $i\in\mathscr{C}$. Thus for a molar fraction c_Γ on Γ, we define $\mathrm{DtN}_{\mathrm{ff}}(c_\Gamma)=\mathbf{w}\cdot\mathbf{n}_{\mathrm{ff}}$ where $\mathbf{w}=\zeta^g(p^0,c^0)(c\mathbf{u}^0-d_{\mathrm{t}}^0\nabla c)$ and c is solution of the convection diffusion equation $\nabla\cdot\mathbf{w}=0$ in Ω_{ff} with Dirichlet condition $c=c_\Gamma$ on Γ. To account efficiently for the convection diffusion boundary layer and the tangential convection, the following low frequency diagonal approximation of the $\mathrm{DtN}_{\mathrm{ff}}$ operator

$$\beta_{\mathrm{pm}}=\mathrm{DtN}_{\mathrm{ff}}(1_\Gamma)-\mathrm{DtN}_{\mathrm{ff}}(0_\Gamma)$$

is used rather than a classical order 0 Taylor approximation.

To compute β_{ff}, the gas liquid porous medium flow is approximated by the Richards equation. Let us define $\bar{c}_i^\ell=1$ for $i=w$ and $\bar{c}_i^\ell=0$ for $i\in\mathscr{C}\setminus\{w\}$, and let $\bar{p}^g$ be a constant reference pressure in the free flow domain typically corresponding to the outflow pressure. From these, the state laws are approximated by $\bar{\zeta}^\ell(p^\ell)=\zeta^\ell(p^\ell,\bar{c}^\ell)$, $\bar{\mu}^\ell(p^\ell)=\mu^\ell(p^\ell,\bar{c}^\ell)$ and $\bar{\rho}^\ell(p^\ell)=\rho^\ell(p^\ell,\bar{c}^\ell)$ and the water molar fraction in the gas is given by $\bar{c}_w^g(p^\ell)=c_w^g(\bar{p}^g,f^\ell(\bar{c}^\ell,p^\ell))$. Let us set $\mathrm{n}^\ell(\mathbf{x},p^\ell)=\bar{\zeta}^\ell(p^\ell)s^\ell(\mathbf{x},\bar{p}^g-p^\ell)$ and $M^\ell(\mathbf{x},p^\ell)=\frac{\bar{\zeta}^\ell(p^\ell)}{\bar{\mu}^\ell(p^\ell)}k_r^\ell(\mathbf{x},s^\ell(\mathbf{x},\bar{p}^g-p^\ell))$. The Richards model with prescribed water molar fraction c_w at Γ is defined as follows after time integration using the implicit Euler scheme:

$$\begin{aligned}\tfrac{\phi}{\Delta t^n}(\mathrm{n}^\ell-\mathrm{n}^{\ell,n-1})+\nabla\cdot\mathbf{v}^\ell=0,&\quad\text{in }\Omega_{\mathrm{pm}},\\ \bar{c}_w^g(p^\ell)=c_w,&\quad\text{on }\Gamma,\end{aligned} \tag{9}$$

where $\mathrm{n}^{\ell,n-1}(\mathbf{x})=\mathrm{n}^\ell(\mathbf{x},p^{\ell,n-1})$ and $\mathbf{v}^\ell=-M^\ell\mathbf{K}(\nabla p^\ell-\bar{\rho}^\ell\mathbf{g})$. At each point of the interface Γ the Eq. (9) is linearized with respect to p^ℓ and its coefficients are freezed leading to $\mathscr{L}_{\mathrm{pm}}\delta p^\ell=\eta\delta p^\ell+\nabla\cdot(-\kappa\nabla\delta p^\ell+\psi\delta p^\ell)$ with the Dirich-

let boundary condition $\delta p^\ell = \frac{\delta c_w}{\partial_{p^\ell} \bar{c}_w^g}$ on Γ. The freezed coefficients are defined by $\eta = \frac{\phi}{\Delta t^n} \partial_{p^\ell} \mathrm{n}^\ell, \kappa = M^\ell \mathbf{K}$ and $\psi = -\partial_{p^\ell} M^\ell \mathbf{K} \nabla p^\ell + \partial_{p^\ell} (M^\ell \bar{\rho}^\ell) \mathbf{K} \mathbf{g}$. The Robin coefficient is obtained using the following DtN order 0 Taylor approximation [3]:

$$\beta_{\mathrm{ff}} = \frac{1}{2 \partial_{p^\ell} \bar{c}_w^g} \left(\psi \cdot \mathbf{n}_{\mathrm{ff}} + \sqrt{(\psi \cdot \mathbf{n}_{\mathrm{ff}})^2 + 4 \eta \kappa \mathbf{n}_{\mathrm{ff}} \cdot \mathbf{n}_{\mathrm{ff}}} \right). \tag{10}$$

3 Numerical Experiment

This test case is a simplified two dimensional setting defined with Andra [6] to simulate the mass exchanges occurring within deep geological radioactive waste disposal at the interface between a geological formation and a ventilation excavated gallery.

The computational domain shown in Fig. 1 is a rectangle of length $l = 100\,\mathrm{m}$ and height $h_{\mathrm{pm}} = 15\,\mathrm{m}$, split horizontally into the free flow domain $\Omega_{\mathrm{ff}} = (0, l) \times (0, h_{\mathrm{ff}})$ of height $h_{\mathrm{ff}} = 5\,\mathrm{m}$ and the porous medium $\Omega_{\mathrm{pm}} = (0, l) \times (h_{\mathrm{ff}}, h_{\mathrm{pm}})$. The temperature is fixed to $T = 303\,\mathrm{K}$ both in the porous medium and the free flow domains. The gas and liquid phases are a mixture of air (a) and water (w) components. The liquid and gas properties are defined by $\zeta^\ell = 55555\,\mathrm{mol.m^{-3}}$, $\mu^\ell = 10^{-3}\,\mathrm{Pa.s}$, $\mu^g = 1.851 \cdot 10^{-5}\,\mathrm{Pa.s}$, $\zeta^g = p^g (RT)^{-1}$, $\rho^\alpha = \zeta^\alpha \sum_{i \in \mathscr{C}} c_i^\alpha m^i, \alpha \in \mathscr{P}$ with the molar masses $m^w = 18 \cdot 10^{-3}\,\mathrm{kg.mol^{-1}}$ and $m^a = 29 \cdot 10^{-3}\,\mathrm{kg.mol^{-1}}$. The fugacities are defined in the liquid phase by the Raoult–Kelvin's law for the water component and the Henry's law for the air which leads by inversion to $c_w^\ell(p^\ell, f) = \frac{f_w}{p_{\mathrm{sat}}} e^{\frac{p_{\mathrm{sat}} - p^\ell}{\zeta^\ell RT}}$, $c_a^\ell(p^\ell, f) = \frac{f_a}{H^a}$, where $p_{\mathrm{sat}} = 4138\,\mathrm{Pa}$ at $T = 303\,\mathrm{K}$ and $H^a = 3.33 \cdot 10^9\,\mathrm{Pa}$. The gas fugacities are defined by the Dalton's law for an ideal mixture of perfect gas leading to $c_i^g(p^g, f) = \frac{f_i}{p^g}$.

The porous medium contains two rocktypes: a concrete layer located in $\Omega_{\mathrm{cc}} = (\frac{l}{2}, l) \times (h_{\mathrm{ff}}, h_{\mathrm{cc}})$ with $h_{\mathrm{cc}} = 6\,\mathrm{m}$ and the Callovo Oxfordian clay elsewhere in $\Omega_{\mathrm{cox}} = \Omega_{\mathrm{pm}} \setminus \Omega_{\mathrm{cc}}$. The liquid saturation and the relative permeabilities are given by the Van Genuchten laws as in [6] with parameters set to $n_{\mathrm{r}} = 1.54$, $m_{\mathrm{r}} = 1 - \frac{1}{n_{\mathrm{r}}}$,

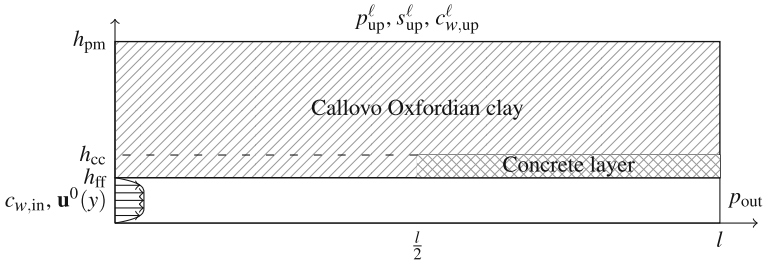


Fig. 1 Computational domain

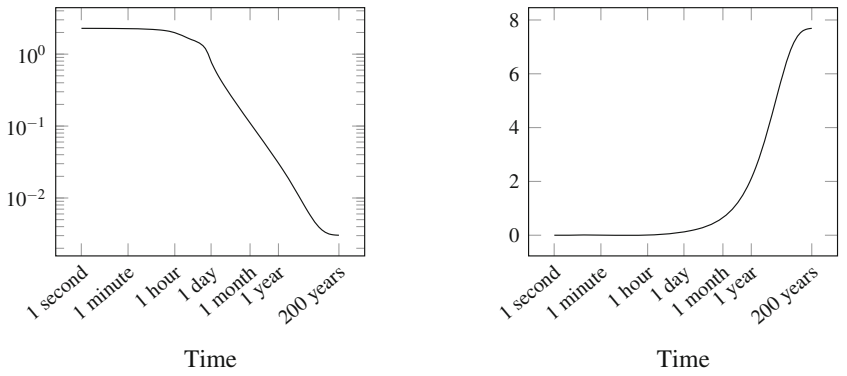


Fig. 2 Evaporation rate at the interface in $\mathrm{l.day}^{-1}.\mathrm{m}^{-1}$ (*left*) and gas volume in the porous medium in m^3 (*right*)

$p_\mathrm{r} = 2 \cdot 10^6\,\mathrm{Pa}$, $s_\mathrm{r}^\ell = 0.01$, $s_\mathrm{r}^g = 0$ for the concrete rocktype and to $n_\mathrm{r} = 1.49$, $m_\mathrm{r} = 1 - \frac{1}{n_\mathrm{r}}$, $p_\mathrm{r} = 15 \cdot 10^6\,\mathrm{Pa}$, $s_\mathrm{r}^\ell = 0.4$, $s_\mathrm{r}^g = 0$ for the Cox rocktype. The porosity is set to $\phi = 0.3$ (resp. $\phi = 0.15$) and the permeability is isotropic and set to $\mathbf{K} = 10^{-18}\,\mathrm{m}^2$ (resp. $\mathbf{K} = 5 \cdot 10^{-20}\,\mathrm{m}^2$) in the concrete rocktype (resp. Cox rocktype).

The liquid pressure, the liquid saturation and the water mole fraction are set both at the initial time and at the top of the porous medium $\Gamma_\mathrm{up} = (0, l) \times \{h_\mathrm{pm}\}$ to $p^{\ell,0} = p_\mathrm{up}^\ell = 4 \cdot 10^6\,\mathrm{Pa}$, $s^{\ell,0} = s_\mathrm{up}^\ell = 1$ and $c_w^{\ell,0} = c_{w,\mathrm{up}}^\ell = 1$ respectively. In the free flow domain, the mean uncoupled turbulent velocity profile $\mathbf{u}^0(y)$ is obtained using the Prandtl algebraic turbulent model [6] which defines the turbulent viscosity μ_t^0. The turbulent diffusion $d_\mathrm{t}^0 = d^g + \frac{\mu_\mathrm{t}^0 - \mu^g}{\rho^g S_\mathrm{c}}$ is deduced using the gas Fickian diffusion $d^g = 2 \cdot 10^{-5}\,\mathrm{m.s}^{-1}$ and the Schmidt number $S_\mathrm{c} = 1$. At the output interface $\Gamma_\mathrm{out} = \{l\} \times (0, h_\mathrm{ff})$, the pressure $p_\mathrm{out} = 10^5\,\mathrm{Pa}$ is the atmospheric pressure which also corresponds to the pressure $\bar{p}^g$ used to compute the Robin coefficient β_ff^n. The velocity at the input boundary $\Gamma_\mathrm{in} = \{0\} \times (0, h_\mathrm{ff})$ is defined by the velocity profile $\mathbf{u}^0(y)$ and is such that $u_\mathrm{in} = -\mid \Gamma_\mathrm{in} \mid^{-1} \int_{\Gamma_\mathrm{in}} \mathbf{u}^0(y) \cdot \mathbf{n}_\mathrm{ff} = 0.5\,\mathrm{m.s}^{-1}$. The input water molar fraction c_w, in corresponds to a relative humidity $H_r = \frac{p_\mathrm{out} c_\mathrm{w,in}}{p_\mathrm{sat}} = 0.5$. Homogeneous Neumann boundary conditions are used at the other boundaries of the domain.

Following [6], a Cartesian mesh of size 100×242 refined at interface Γ is used. The Darcy problem in (5) and the convection diffusion equation in (7) are solved using a two point flux approximation method given in [8] with additional face unknowns at the interface Γ. The RANS problem in (6) is solved using a staggered Marker And Cell scheme given in [2]. An implicit Euler scheme is used in time with a time step $\Delta t^n = (1.2)^{n-1}\,\mathrm{s}$ which varies exponentially up to reach the final simulation time $t_\mathrm{f} = 200$ years, which corresponds roughly to the time of ventilation of the storage and is large enough to reach the stationary state.

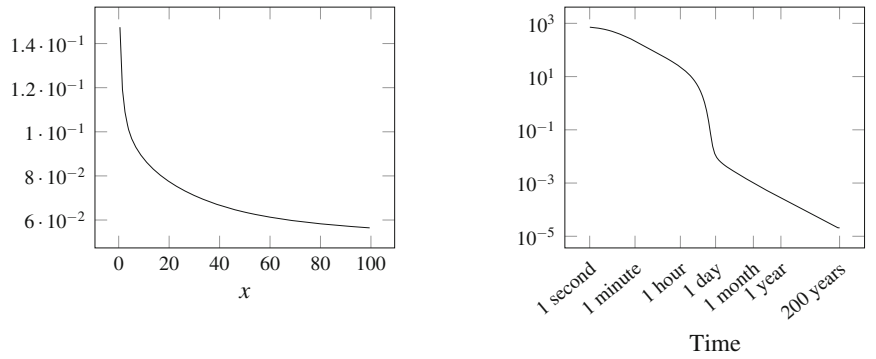


Fig. 3 Value of the Robin coefficient β_{pm} along the interface Γ (*left*) and mean value $\bar{\beta}^n_{\text{ff}} = \mid \Gamma \mid^{-1} \int_\Gamma \beta_{\text{ff}}(\mathbf{x}, t^n)$ as a function of time (*right*)

Figure 2 shows the evaporation rate in the gallery and the gas volume in the porous medium as a function of time. A drop of the evaporation rate occurs at $t \simeq 1$ day when the interface is not longer saturated with water. Figure 3 exhibits the Robin coefficients obtained for this test case. The dependence on time of β_{ff} is crucial to obtain the convergence of the algorithm and corresponds roughly to a Dirichlet condition before the drop of the evaporation rate and to a Neumann boundary condition after the drop of the evaporation rate. It has been checked that for a stopping criterion $\varepsilon = 10^{-6}$ in (8), the domain decomposition algorithm converges in an average of 3.9 iterations per time step. In practice, one or two iterations are enough to obtain the same solution as the fully coupled algorithm.

Acknowledgements This work was supported by ANDRA (the french national agency of the radioactive waste management).

References

1. Fetzer, T., Smits, K., Helmig, R.: Effect of turbulence and roughness on coupled porous-medium/free-flow exchange processes. Transp. Por. Media **114**(2), 395–424 (2016)
2. Harlow, F.H., Welch, J.E.: Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface. Phys. Fluids **8**(12), 2182–2189 (1965)
3. Japhet, C., Nataf, F., Rogier, F.: The optimized order 2 method: application to convection-diffusion problems. Futur. Gener. Comput. Syst. **18**(1), 17–30 (2001)
4. Lauser, A., Hager, C., Helmig, R., Wohlmuth, B.: A new approach for phase transitions in miscible multi-phase flow in porous media. Adv. Water Res. **34**(8), 957–966 (2011)
5. Masson, R., Trenty, L., Zhang, Y.: Formulations of two phase liquid gas compositional Darcy flows with phase transitions. Int. J. Fin. Vol. **11**, 34 (2014)

6. Masson, R., Trenty, L., Zhang, Y.: Coupling compositional liquid gas darcy and free gas flows at porous and free-flow domains interface. J. Comput. Phys. **321**, 708–728 (2016)
7. Mosthaf, K., Baber, K., Flemisch, B., Helmig, R., Leijnse, A., Rybak, I., Wohlmuth, B.: A coupling concept for two-phase compositional porous-medium and single-phase compositional free flow. Water Res. Res. (2011)
8. Peaceman, D.W.: Fundamentals of Numerical Reservoir Simulation. Developments in Petroleum Science. Elsevier (1977)

Hybrid Finite-Volume/Finite-Element Schemes for $p(x)$-Laplace Thermistor Models

Jürgen Fuhrmann, Annegret Glitzky and Matthias Liero

Abstract We introduce an empirical PDE model for the electrothermal description of organic semiconductor devices by means of current and heat flow. The current flow equation is of $p(x)$-Laplace type, where the piecewise constant exponent $p(x)$ takes the non-Ohmic behavior of the organic layers into account. Moreover, the electrical conductivity contains an Arrhenius-type temperature law. We present a hybrid finite-volume/finite-element discretization scheme for the coupled system, discuss a favorite discretization of the $p(x)$-Laplacian at hetero interfaces, and explain how path following methods are applied to simulate S-shaped current-voltage relations resulting from the interplay of self-heating and heat flow.

Keywords Finite volume scheme · $p(x)$-Laplace thermistor model · Path following

MSC (2010): 65M08 · 35J92 · 35G60 · 35Q79 · 80M12 · 80A20

1 Introduction

Presently, carbon-based semiconductors are used in smartphone displays and increasingly in TV screens. Due to the fascinating properties of organic light-emitting diodes (OLEDs), e.g. large-area surface emission, semi-transparency, flexibility, also lighting applications are of great interest. However, lighting requires a much higher brightness than displays and hence higher currents are necessary. These cause substantial Joule self-heating accompanied by unpleasant brightness inhomogeneities of the panels. An appropriate simulation tool for the electrothermal description of OLEDs

J. Fuhrmann (✉) · A. Glitzky · M. Liero
Weierstrass Institute, Mohrenstraße 39, 10117 Berlin, Germany
e-mail: fuhrmann@wias-berlin.de

A. Glitzky
e-mail: glitzky@wias-berlin.de

M. Liero
e-mail: liero@wias-berlin.de

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_42

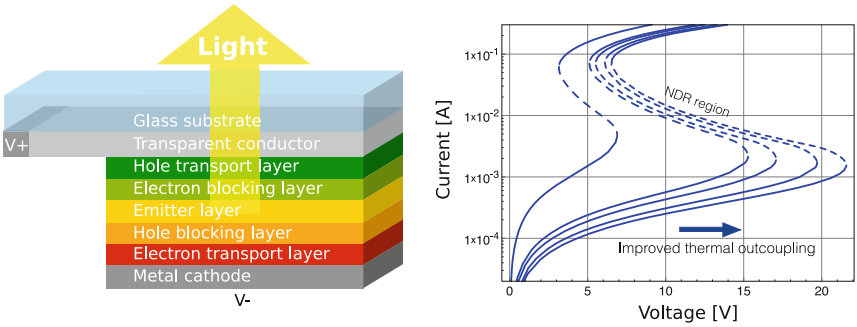


Fig. 1 Schematic view of an OLED stack (*left*) and simulated current-voltage characteristics for different thermal outcoupling regimes, regions of negative differential resistance are dashed (*right*)

can help to validate cost-efficient device concepts by accounting for nonlinear self-heating effects.

Applying a voltage to an organic semiconductor device induces a current flow which leads to a power dissipation by Joule heating and hence also a temperature rise. As higher temperatures improve the electrical conductivity in organic materials, higher currents occur. Thus, a positive feedback loop develops that either leads to the destruction of the device by thermal runaway if the generated heat cannot be dispersed into the environment or results in S-shaped current-voltage characteristics. In particular, in the latter case regions of negative differential resistance (S-NDR) appear, where currents increase despite of decreasing voltages, see Fig. 1 (right). Devices that show such an electrothermal interplay are called thermistors.

S-NDR has been verified for the organic material C_{60} in [7] and for organic materials used in OLEDs in [6], where the temperature dependence of the conductivity is modeled by an exponential law of Arrhenius type, which features an activation energy that is linked to the energetic disorder in the organic material. Due to the huge aspect ratios of OLED panels, such devices cannot be regarded as a single spatially homogeneous thermistor device, but rather as an array of thermally and electrically coupled thermistor devices. In particular, the self-heating, and hence also the local differential resistance, is now a collective property of neighboring thermistors. One attempt is followed in [6] where the electrothermal behavior of OLEDs is investigated by means of electrically and thermally coupled thermistor networks and SPICE simulations.

In this paper, we present a mathematical model for the current and heat flow in organic semiconductor devices consisting of a coupled PDE system for the electrostatic potential and the temperature, see Sect. 2. This PDE modeling approach gives much more flexibility concerning variations in geometry and material composition than network models. An essential feature of our PDE model is that the current flow equation is of $p(x)$-Laplace type, where the exponent $p(x)$ takes the non-Ohmic behavior of the organic layers into account. The exponent is in general

discontinuous but piecewise constant as the different functional layers exhibit different power laws. In Sect. 3 we introduce a numerical scheme for the simulation of current and heat flow in OLEDs. One of the major challenges is the derivation of a stable scheme for the $p(x)$-Laplacian, which we address by using a hybrid finite-volume/finite-element approach which is discussed in Sect. 4. Whereas the presented scheme preserves lower bounds for the temperature, a convergence proof is still under discussion. Challenges arise from bad analytical properties of the Joule heat term. Finally, in Sect. 5 we describe a path following method which enables us to simulate the electrothermal behavior of organic semiconductor devices also in the S-NDR regime.

2 PDE Modeling of Current and Heat Flow

To describe the interplay of current and heat flow in OLEDs the following empirical PDE model was developed in [9]. It consists of the current flow equation for the electrostatic potential φ and the heat equation for the temperature T

$$\begin{aligned} -\nabla \cdot S(x, T, \nabla\varphi) &= 0, \\ -\nabla \cdot \big(\lambda(x)\nabla T\big) &= H(x, T, \nabla\varphi) \end{aligned} \qquad \text{on } \Omega \subset \mathbb{R}^d \tag{1}$$

with electrical current density S, heat conductivity λ, and Joule heat term H. The special features of the model are the Arrhenius-like temperature law as well as the non-Ohmic current-voltage relations incorporated by a power law in the function S,

$$S(x, T, \nabla\varphi) = \kappa_0(x)F(x, T)|\nabla\varphi|^{p(x)-2}\nabla\varphi, \quad F(x, T) = \exp\Big[-\frac{E_{\text{act}}(x)}{k_{\text{B}}}\Big(\frac{1}{T} - \frac{1}{T_a}\Big)\Big].$$

Here, $T_a > 0$ and k_{B} denote the fixed ambient temperature and Boltzmann's constant. The quantitites κ_0, p, and E_{act} are material dependent effective conductivity, power law exponent, and activation energy, respectively, which have to be extracted from measurements. The first equation in (1) becomes of $p(x)$-Laplacian type with discontinuous but piecewise constant exponent p. In particular, we have $p(x) \equiv 2$ in Ohmic materials such as electrodes and different values $p(x) > 2$ in organic layers. The Joule heat term in the second equation of (1) takes the form

$$H(x, T, \nabla\varphi) = \eta(x, T, \nabla\varphi)\kappa_0(x)F(x, T)|\nabla\varphi|^{p(x)},$$

where $\eta(x, T, \nabla\varphi) \in [0, 1]$ represents the light-outcoupling factor. The system is complemented by Dirichlet and no-flux boundary conditions for the potential φ at the contacts Γ_D and the insulating parts Γ_N of the boundary, and Robin boundary conditions for the heat flow to describe the coupling to the environment

$$\begin{aligned} &\varphi = \varphi^D \text{ on } \Gamma_D, \qquad S(x, T, \nabla\varphi) \cdot \nu = 0 \text{ on } \Gamma_N, \\ &-\lambda(x)\nabla T \cdot \nu = \gamma(x)(T - T_a) \text{ on } \Gamma = \partial\Omega. \end{aligned} \tag{2}$$

Since the Joule heat term H is a priori only in L^1, the mathematical treatment of the system is not straightforward. For analytical results concerning the existence, boundedness and regularity of solutions to Problem (1), (2) we refer to [3, 4, 8].

3 Numerical Scheme

Since we have to deal with piecewise constant functions $p(x)$, we subdivide the computational domain $\overline{\Omega} = \bigcup_{r\in\mathscr{R}} \overline{\Omega}_r$ into disjoint subdomains coinciding with the regions of continuity of the coefficients. We call the surface between two neighboring regions *hetero interface*. Due to its ability to preserve the maximum principle of the current flow equation and the positivity of the temperature, we prefer a two-point flux finite-volume method for the dicretization of (1), (2) over methods defined on more general meshes (e.g. [5]). Our control volumes are Voronoi cells based on a grid with the boundary conforming Delaunay property with respect to boundaries and hetero interfaces [11]. Let $\mathscr{V}$ denote the set of Voronoi boxes and $m = \#\mathscr{V}$ be the number of cells. We assume that each control volume $K \in \mathscr{V}$ contains a collocation point $x_K \in \overline{\Omega}$.

Let $K \in \mathscr{V}$ be an internal Voronoi box meaning that $\mathrm{mes}_{d-1}(\overline{K} \cap \overline{\partial\Omega}) = 0$. We apply Gauss's theorem to the integral of the flux divergence to obtain for the current flow equation in (1) the flux balance with further subdivision into contributions from adjacent subdomains ($\kappa_{0,r}$ and F_r indicate the corresponding values in region Ω_r):

$$0 = \int_K \nabla \cdot S(x, T, \nabla\varphi)\,\mathrm{d}x = \sum_{r\in\mathscr{R}} \sum_{L\sim K} \int_{\overline{K}\cap\overline{L}\cap\Omega_r} \kappa_{0,r}\, F_r(T) |\nabla\varphi|^{p_r-2} \nabla\varphi \cdot \nu_{KL}\,\mathrm{da}, \tag{3}$$

where $L \sim K$ indicates that L is adjacent to K and ν_{KL} is the unit normal vector pointing from K into L. Note that the normal flux over a surface $\overline{K} \cap \overline{L} \cap \Omega_r$ does not only depend on the normal components of $\nabla\varphi$ but on the modulus of the full gradient. To take this into account we compute the approximation of $|\nabla\varphi|^2$ on $\overline{K} \cap \overline{L} \cap \Omega_r$ as the average squared norms of the P1 finite element gradients $\nabla_\tau\varphi$ over the set $\mathscr{T}_{K,L,r}$ of all simplices τ (triangles in 2D) in the underlying Delaunay triangulation adjacent to the edge $\overline{x_K x_L}$ and belonging to Ω_r:

$$\left|\nabla\varphi\right|^2|_{\overline{K}\cap\overline{L}\cap\Omega_r} \approx G^2_{K,L,r}(\varphi) := \frac{\sum_{\tau\in\mathscr{T}_{K,L,r}} |\tau||\nabla_\tau\varphi|^2}{\sum_{\tau\in\mathscr{T}_{K,L,r}} |\tau|}. \tag{4}$$

By this approach we find an approximation of the right-hand side of (3) consisting in replacing the surface integral by a simple quadrature, and the gradient projection by a finite difference expression

$$0 = \sum_{r\in\mathscr{R}} \sum_{L\sim K} \frac{|\overline{K} \cap \overline{L} \cap \Omega_r|}{|x_K - x_L|} \kappa_{0,r} F_r(T_{KL}) G_{K,L,r}(\varphi)^{p_r-2} (\varphi_L - \varphi_K). \tag{5}$$

The same method for calculating the conductivity in the Joule heat term is combined with the technique proposed in [2] allowing to evaluate the Joule heating approximation by edge contributions: Gauss's theorem in the heat equation yields

$$0 = \sum_{L\sim K} \int_{\overline{K}\cap\overline{L}} \lambda(x)\nabla T \cdot \nu_{KL}\,\mathrm{da} + \int_K \eta(x)\kappa_0(x)F(x,T)|\nabla\varphi|^{p(x)}\,\mathrm{d}x, \tag{6}$$

and the suggested approach yields the approximation of the heat flow equation on K

$$\begin{aligned} 0 = \sum_{r\in\mathscr{R}}\sum_{L\sim K}\Bigg(&\frac{|\overline{K}\cap\overline{L}\cap\Omega_r|}{|x_K - x_L|}\lambda_r(T_L - T_K) \\ &+ \frac{1}{2}\frac{|\overline{K}\cap\overline{L}\cap\Omega_r|}{|x_K - x_L|}\eta_r\kappa_{0,r}F_r(T_{KL})G_{K,L,r}(\varphi)^{p_r-2}(\varphi_L - \varphi_K)^2\Bigg), \end{aligned} \tag{7}$$

where $T_{KL} = (T_K + T_L)/2$.

For Voronoi boxes $K \in \mathscr{V}$ with $\mathrm{mes}_{d-1}(\overline{K}\cap\overline{\partial\Omega}) > 0$ we additionally have to implement Dirichlet, no-flux or Robin boundary conditions, respectively,

$$w = w^D \text{ or } \quad -\nu\cdot(b\nabla w) = 0 \text{ or } \quad -\nu\cdot(b\nabla w) = e(w - w^D) \text{ on } \overline{K}\cap\overline{\Omega}_r\cap\overline{\partial\Omega}.$$

We write the flux over an outer face $\overline{K}\cap\overline{\Omega}_r\cap\overline{\partial\Omega}$ as $e_r(w_K - w_r^D)|\overline{K}\cap\overline{\Omega}_r\cap\overline{\partial\Omega}|$, where w_r^D corresponds to a mean of w^D on $\overline{K}\cap\overline{\Omega}_r\cap\overline{\partial\Omega}$, e_r is chosen very large to realize Dirichlet boundary values, e_r is set to zero for no-flux boundary conditions and it corresponds to a mean for e in case of Robin boundary conditions. Such that, according to (5) and (7) for all Voronoi boxes $K \in \mathscr{V}$ we have to solve

$$\begin{aligned} 0 = \sum_{r\in\mathscr{R}}\Bigg(&\sum_{L\sim K}\frac{|\overline{K}\cap\overline{L}\cap\Omega_r|}{|x_K - x_L|}\kappa_{0,r}F_r(T_{KL})G_{K,L,r}(\varphi)^{p_r-2}(\varphi_L - \varphi_K) \\ &- e_r(\varphi_K - \varphi_r^D)|\overline{K}\cap\overline{\Omega}_r\cap\overline{\partial\Omega}|\Bigg), \end{aligned} \tag{8}$$

$$\begin{aligned} 0 = \sum_{r\in\mathscr{R}}\sum_{L\sim K}\Bigg(&\frac{|\overline{K}\cap\overline{L}\cap\Omega_r|}{|x_K - x_L|}\lambda_r(T_L - T_K) \\ &- \frac{1}{2}\frac{|\overline{K}\cap\overline{L}\cap\Omega_r|}{|x_K - x_L|}\eta_r\kappa_{0,r}F_r(T_{KL})G_{K,L,r}(\varphi)^{p_r-2}(\varphi_L - \varphi_K)^2\Bigg) \\ &+ \sum_{r\in\mathscr{R}}\gamma_r(T_K - T_a)|\overline{K}\cap\overline{\Omega}_r\cap\overline{\partial\Omega}|. \end{aligned} \tag{9}$$

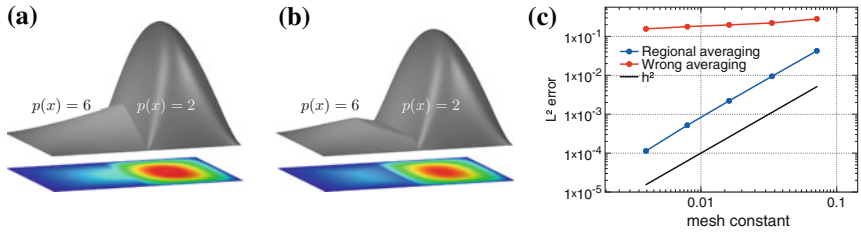


Fig. 2 Discrete solutions of $p(x)$-Laplace equation (10) with constant right-hand side, homogeneous Dirichlet boundary conditions, and piecewise constant $p(x)$: **a** correct approximation, **b** wrong approximation with two local maxima due to gradient averaging ignoring the hetero interface, and **c** L^2 error of approximating solutions for correct and wrong averaging schemes

4 Numerical Tests for the $p(x)$-Laplacian

To justify our discretization ansatz for the gradient norm $|\nabla\varphi|$ in (4), we consider the two-dimensional test case for the $p(x)$-Laplacian

$$-\nabla\cdot(|\nabla\varphi|^{p(x)-2}\nabla\varphi) = f \quad \text{in } \Omega, \quad \varphi = 0 \quad \text{on } \partial\Omega \tag{10}$$

with fixed source term $f = 1$, where Ω is composed of two unit squares Ω_1 and Ω_2 being glued together at one edge and setting $p(x) = 6$ in Ω_1 and $p(x) = 2$ in Ω_2, respectively. The simulations in Fig. 2 illustrate the importance of taking care of the hetero interface when calculating the average of the gradient norm. An averaging over all simplices adjacent to a given edge regardless of the hetero region they belong to leads to an artificial diffusion along the hetero interface. We highlight the appearance of two local maxima (one centered in Ω_1 and one centered in Ω_2) in Fig. 2b. As shown in Fig. 2c, this effect cannot be diminished by grid refinement. Indeed, the L^2 error of the wrong averaging scheme with respect to the correct solution (Fig. 2a) stays above a positive constant.

The validity of our approach (4) and the way it has been implemented (by (8) with right hand side $|K|$ and $\kappa_{0,r} = F_r(T_{KL}) = 1$, $\varphi_r^D = 0$, e_r large) hinges on the fact that all the measures $|\overline{K}\cap\overline{L}\cap\Omega_r|$ can be calculated from contributions from each simplex which at the hetero interfaces have to stay nonnegative. It is guaranteed by the boundary conforming Delaunay property of the underlying triangulation.

5 A Path Following Method for Simulating *S*-shaped Current-Voltage Relations

Our discretization scheme allows to simulate complicated three-dimensional OLED structures, see [6]. As an example we consider a crossbar OLED stack depicted in Fig. 1 (left), given by two stacked cuboids Ω_{anode} and Ω_{org}. The upper one, Ω_{anode},

Table 1 Geometry and material parameters

Domain	p	E_{act} [eV]	κ_0 [1/(Ωm)]	λ [W/(mK)]	η	Thickness [nm]
Ω_{anode}	2.0	0.0	7.4×10^{-6}	1.0×10^{3}	1.0	9
Ω_1	4.07	0.325	7.7×10^{-8}	1.0×10^{3}	1.0	64
Ω_2	6.0	1.588	9.7×10^{-8}	1.0×10^{3}	0.8	20
Ω_3	4.7	0.2	2.1×10^{-7}	1.0×10^{3}	1.0	50

representing the optically transparent anode is overlapping to the left and electrically contacted only on the left side Γ_+. The lower one, Ω_{org}, consisting of the organic semiconducting layers Ω_1, Ω_2, and Ω_3 realizes the actual OLED structure with an active area of 2×2 mm, see Fig. 1 (left). The organic material is contacted by a metal layer. Due to the high conductivity of this layer we assume that the potential is constant here and neglect the metal layer entirely in the simulations by prescribing Dirichlet boundary conditions on the bottom Γ_- of Ω_3.

On Γ_- the potential is set to zero and on Γ_+ to the (spatially constant) externally applied voltage V. We determine the current-voltage relation of the OLED stack (which can be S-shaped) by calculating the current over Γ_+. Then, the Dirichlet boundary is given by $\Gamma^D = \Gamma_+ \cup \Gamma_-$. The ambient temperature is fixed to 293 K, the other essential parameters for the simulation are collected in Table 1.

With the Eqs. (8) and (9) for all Voronoi boxes $K \in \mathcal{V}$ we arrive at a system of $2m$ coupled nonlinear algebraic equations for $u = (\varphi_K, T_K)_{K \in \mathcal{V}}$ of the form

$$g(u, V) = 0, \quad g : \mathbb{R}^{2m} \times \mathbb{R} \to \mathbb{R}^{2m}.$$

To trace a solution branch, starting from a solution (u_0, V_0) of $g(u, V) = 0$ we use a predictor-corrector method [10] adapted to PDE calculations as proposed in [1]. The prediction is obtained by moving forward a step along the tangent t to the branch. First we solve $g_{u,V}(u_0, V_0)t = 0$, $t \in \mathbb{R}^{2m+1}$. To ensure that t points in the forward direction with respect to the tangent t_0 of the last point, we demand $t_0 \cdot t > 0$. In other words, we have to solve

$$\begin{pmatrix} g_{u,V}(u_0, V_0) \\ t_0 \end{pmatrix} t = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Next, we normalize t such that $\|t\| = 1$. Our predictor (u^*, V^*) now is chosen as

$$\begin{pmatrix} u^* \\ V^* \end{pmatrix} = \begin{pmatrix} u_0 \\ V_0 \end{pmatrix} + \frac{\Delta L}{\|t\|_*} t, \quad \text{where } \|t\|_*^2 = \frac{1}{2m} \sum_{i=1}^{2m} t_i^2 + t_{2m+1}^2$$

ensures that a step along the branch gives similar proportion to the unknowns and to the parameter, and, by construction, $\|u^* - u_0, V^* - V_0\|_* = \Delta L$. The corrector

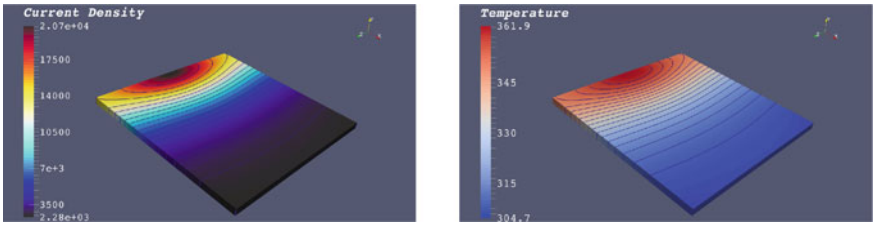


Fig. 3 Simulated current density [A/m^2] (*left*) and temperature distribution [K] (*right*) in horizontal cross section through the middle of the organic emitter layer at an applied voltage of 6.5 V

step consists in solving the nonlinear system

$$\begin{pmatrix} g(u,V) \\ \|u-u_0, V-V_0\|_*^2 - (\Delta L)^2 \end{pmatrix} = 0$$

by Newton's method, where the calculated prediction (u^*, V^*) is used as starting value. If Newton's method does not converge, meaning that the predictor is too far from the desired solution, the step size ΔL (related to the arc length parameter) is locally reduced until the method is convergent. The convergent Newton procedure yields the next point (u_1, V_1) on the solution branch with a distance of ΔL to (u_0, V_0).

Figure 3 contains the simulated current density and the temperature distribution in a horizontal cross section in the emitting layer of the OLED material for an applied voltage of 6.5 V. Have in mind that the temperature and current density maxima appear at the side where the anode voltage is applied. Figure 1 (right) shows simulated S-shaped current voltage relations for test structures with different thermal outcoupling regimes realized by varying heat transfer coefficients γ in (2).

Acknowledgements A.G. and M.L. gratefully acknowledge the funding received via Research Center MATHEON supported by ECMath in project D-SE2.

References

1. Bloch, J., Fuhrmann, J., Gärtner, K.: Bifurcation analysis of nonlinear systems of PDE's. (unpublished report)
2. Bradji, A., Herbin, R.: Discretization of coupled heat and electrical diffusion problems by finite-element and finite-volume methods. IMA J. Numer. Anal. **28**(3), 469–495 (2008)
3. Bulíček, M., Glitzky, A., Liero, M.: Systems describing electrothermal effects with $p(x)$-Laplace like structure for discontinuous variable exponents. SIAM J. Math. Anal. **48**, 3496–3514 (2016)
4. Bulíček, M., Glitzky, A., Liero, M.: Thermistor systems of $p(x)$-Laplace-type with discontinuous exponents via entropy solutions. Discrete Contin. Dyn. Syst. Ser. S **10**(4), 697–713 (2017)
5. Eymard, R., Gallouët, T., Herbin, R.: Discretization of heterogeneous and anisotropic diffusion problems on general nonconforming meshes. IMA J. Numer. Anal. **30**, 1009–1043 (2010)

6. Fischer, A., Koprucki, T., Gärtner, K., Brückner, J., Lüssem, B., Leo, K., Glitzky, A., Scholz, R.: Feel the heat: nonlinear electrothermal feedback in organic LEDs. Adv. Funct. Mater. **24**, 3367–3374 (2014)
7. Fischer, A., Pahner, P., Lüssem, B., Leo, K., Scholz, R., Koprucki, T., Gärtner, K., Glitzky, A.: Self-heating, bistability, and thermal switching in organic semiconductors. Phys. Rev. Lett. **110**, 126,601/1–126,601/5 (2013)
8. Glitzky, A., Liero, M.: Analysis of $p(x)$-Laplace thermistor models describing the electrothermal behavior of organic semiconductor devices. Nonlinear Anal. Real World Appl. **34**, 536–562 (2017)
9. Liero, M., Koprucki, T., Fischer, A., Scholz, R., Glitzky, A.: p-Laplace thermistor modeling of electrothermal feedback in organic semiconductor devices. Z. Angew. Math. Phys. **66**, 2957–2977 (2015)
10. Seydel, R.: Practical Bifurcation and Stability Analysis. Springer, Heidelberg (1994)
11. Si, H., Gärtner, K., Fuhrmann, J.: Boundary conforming Delaunay mesh generation. Comput. Math. Math. Phys. **50**(1), 38–53 (2010)

Finite Volume Scheme for Coupling Two–Phase Flow with Reactive Transport in Porous Media

E. Ahusborde, B. Amaziane and M. El Ossmani

Abstract In this work the numerical solution of a system of coupled partial differential and differential algebraic equations describing two-phase multicomponent flow, transport and chemical reactions is considered. An implicit finite volume scheme is used to descretize a two-phase two-component flow problem, which is then sequentially coupled to a reactive transport problem solved by a direct substitution approach (DSA). More precisely, we used firstly the module $2p2c$ implemented in the parallel open-source simulator DuMu^X to solve a two-phase two-component flow with two dominant species without chemistry. Secondly, the reactive transport is described by advection dispersion equations coupled to differential algebraic equations to deal with the minor species. Again an implicit finite volume method is used to discretize this subsystem using a DSA. In this context, we have developed and integrated a reactive transport module $1pNc - react$ in the DuMu^X framework. Finally, numerical results for a highly complex geochemistry problem are presented to demonstrate the ability of our method to approximate solutions of two-phase flows with reactive transport in heterogeneous porous media.

Keywords Cell centred · Porous media · Reactive transport · Two-phase flow · Kinetic reactions · Code coupling · DuMu^X

MSC (2010): 76S05 · 74S10 · 74F25 · 76T10

E. Ahusborde · B. Amaziane
CNRS/Univ Pau & Pays Adour, Laboratoire de Mathematiques Et de Leurs Applications de Pau, Fédération IPRA, UMR5142, 64000 Pau, France
e-mail: etienne.ahusborde@univ-pau.fr

B. Amaziane
e-mail: brahim.amaziane@univ-pau.fr

M. El Ossmani (✉)
University Moulay Ismaïl, EMMACS-ENSAM, Marjane II, 50000 Meknès, Morocco
e-mail: m.elossmani@ensam.umi.ac.ma

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_43

1 Formulation of the Problem

In this section, we present a geochemical model describing the chemical reactions and the governing equations modelling two-phase multicomponent flow with reactive transport in porous media.

We consider N_s chemical species Y_j $(j = 1, \ldots, N_s)$ involved in N_r reactions (N_e equilibrium reactions and N_k kinetic reactions):

$$\sum_{j=1}^{N_s} \mathbb{S}_{ij} Y_j \rightleftharpoons 0, \quad i = 1, \ldots, N_r \Longleftrightarrow \mathbb{S}Y \rightleftharpoons 0,$$

where $\mathbb{S} \in \mathbb{R}^{N_r \times N_s}$ is the stoichiometric matrix which can be divided as follows: $\mathbb{S} = \begin{pmatrix} \mathbb{S}^e \\ \mathbb{S}^k \end{pmatrix}$, with $\mathbb{S}^e \in \mathbb{R}^{N_e \times N_s}$ and $\mathbb{S}^k \in \mathbb{R}^{N_k \times N_s}$ corresponding respectively to the equilibrium and kinetic reactions.

Each equilibrium reaction gives rise to an algebraic relation called mass action law that links the activities of the species involved in the reaction. In logarithmic form, the mass action law writes as follows:

$$\mathbb{S}^e \log \mathbf{a} = \log \mathbf{K^e}, \tag{1}$$

where $\mathbf{a}$ is a vector of activities of all chemical species, $\mathbf{K^e}$ is a vector of equilibrium constants.

Each kinetic reaction leads to an ordinary differential equation (ODE):

$$\frac{dc}{dt} = -r_k, \tag{2}$$

where c denotes the concentration of a mineral and r_k is the kinetic reaction rate depending on the activities of the species present in the reaction (see for instance [6]).

In the sequel, the index $\alpha \in \{l, g, s\}$ (l for liquid, g for gas and s for solid) refers to the phase, while the superscript i refers to the species. To specify which species belongs to which phase, we define the *phase – species* correspondence by setting α_i to the index of the phase that contains species i.

For each species, we consider the mass balance equation (see for instance [5]):

$$\frac{\partial}{\partial t}(\theta_{\alpha_i} c^i) - \nabla \cdot (\theta_{\alpha_i} D_{\alpha_i} \nabla c^i) + \nabla \cdot (c^i \overrightarrow{q_{\alpha_i}}) = \sum_{j=1}^{N_r} \mathbb{S}_{ji} r_j, \quad i = 1 \ldots N_s, \tag{3}$$

where θ_α [-] denotes the volumetric content of phase α ($\theta_\alpha = \phi S_\alpha$, ϕ [-] being the porosity of the reservoir which is assumed constant for each medium and S_α [-] the saturation of phase α if $\alpha \in \{l, g\}$ and $\theta_s = 1 - \phi$), c^i [mol.m^{-3}] is the molar concentration of species i (in phase α_i), D_α [m^2.s^{-1}] denotes the diffusivity of phase

α, $\overrightarrow{q_\alpha}$ [m.s^{-1}] is the Darcy velocity of phase α, r_j [mol.m^{-3}.s^{-1}] is the rate of the reaction j (it can be equilibrium or kinetic), $\mathbb{S}_{ji}$ [-] is the stoichiometric coefficient of species i in reaction j.

For sake of simplicity even if it is not mandatory in the sequel, we assume that the diffusion coefficient is independent of the chemical species i, i.e. it depends only on the phase α.

The Darcy velocity of phase α is expressed as follows:

$$\overrightarrow{q_\alpha} = -\frac{k_{r\alpha}}{\mu_\alpha}\mathbb{K}(\nabla P_\alpha - \rho_\alpha \overrightarrow{g}), \tag{4}$$

where $k_{r\alpha}(S_l)$ [-] denotes the relative permeability of phase α, μ_α [Pa.s] is the dynamic viscosity of phase α, $\mathbb{K}$ [m^2] is the absolute permeability tensor, P_α [Pa] is the pressure of phase α, ρ_α [kg.m^{-3}] is the mass density of phase α given by an equation of state and $\overrightarrow{g}$ [m.s^{-2}] is the gravitational acceleration.

The phase pressures are connected by the capillary pressure law:

$$P_c(S_l) = P_g - P_l. \tag{5}$$

For sake of simplicity, we introduce the diffusion-advection operator:

$$L_\alpha(c) = -\nabla \cdot (\theta_\alpha D_\alpha \nabla c) + \nabla \cdot (c\overrightarrow{q_\alpha}). \tag{6}$$

Equation (3) can be written in concise notation:

$$\frac{\partial \mathbf{N}}{\partial t} + \mathbf{L} = \mathbb{S}^T \mathbf{r}, \tag{7}$$

with $\mathbf{N} = (\theta_{\alpha_1} c^1, \ldots, \theta_{\alpha_{N_s}} c^{N_s})^T$, $\mathbf{L} = (L_{\alpha_1}(c^1), \ldots, L_{\alpha_{N_s}}(c^{N_s}))^T$, $\mathbf{r} = (r_1, \ldots, r_{N_r})$.

Reaction rates can be eliminated by multiplying Eq. (7) by a $(N_s - N_r) \times N_s$ component matrix $\mathbf{U}$ such that $\mathbf{U}\,\mathbb{S}^T = 0$. This matrix exists because of the full rank assumption on $\mathbb{S}$. In general, the computation of $\mathbf{U}$ can be performed in different ways (see for instance [7]), the simplest being Gaussian elimination.

After multiplication by $\mathbf{U}$, we obtain a new set of $N_s - N_r$ equations:

$$\frac{\partial \mathbf{U}\,\mathbf{N}}{\partial t} + \mathbf{U}\,\mathbf{L} = \mathbf{0}. \tag{8}$$

To retrieve the same number of equations as there are unknowns, we add the N_e mass actions laws defined by (1) corresponding to the equilibrium reactions and N_k ordinary differential equations corresponding to the kinetic reactions given by (2). In the following section, we present our methodology to solve the system that consists of mass conservation laws (8), mass action laws (1) and ODEs (2).

2 Numerical Methodology

All developments in the present study are integrated into DuMu^X framework [4]. DuMu^X (DUNE for Multi-{Phase, Component, Scale, Physics, ...} flow and transport in porous media) is a free and open-source simulator for flow and transport processes in porous media, based on the Distributed and Unified Numerics Environment DUNE [1].

As in [2, 3] we adopt a sequential strategy that consists in splitting the original problem into two sub-problems. This splitting is justified by the fact that among the chemical species there exists one dominant specie within each phase and that the minor species have no significant influence on the mass balance equations of the dominant species. Then, the first subsystem is a simplified two-phase two-component flow devoted to the dominant species. Change of phase happens only in this subsystem. The second one is a reactive transport problem that computes the contribution of the minor species. The numerical approach for the second subsystem is a global implicit approach (GIA) unlike in [2, 3] where a sequential iterative approach (SIA) was used. Both subsystems are discretized by a fully implicit cell-centred finite volume scheme and then sequentially coupled.

- **Two-phase two-component flow model** ($2p2c$)
  This model is obtained using the system (8) for two dominant species in each phase coupled with Darcy, capillary pressure laws, equation of state for the density of each phase and solubility laws. Then the contribution of the minor species is treated explicitly as a source term in the mass conservation laws of two dominant species. To tackle this model, we have used a module implemented in DuMu^X called $2p2c$. The approach is fully implicit. The spatial discretization employs a finite volume method combining a first-order upwinding mobility scheme for the convective terms and a conforming finite element method with piecewise linear elements for the diffusive terms. The time discretization is done by an implicit Euler scheme. The nonlinear system is solved by a Newton method and a preconditioned BiConjugate Gradient STABilized (BiCGSTAB) method is used to solve the linear system. The control of the time-step is based on the number of iterations required by the Newton method to achieve convergence for the last time iteration. The time-step is reduced, if the number of iterations exceeds a specified threshold, whereas it is increased if the method converges within less iterations.
  In this module, the main difficulty is the management of the possible appearance and disappearance of a phase. This process is managed by a phase state dependent variable switch. The phase switch occurs when the equilibrium concentration of a component in a phase is exceeded.
- **Reactive transport model** ($1pNc - react$)
  This model is obtained using the system (8) for other minor species that are only present in liquid phase, mass action laws involved by equilibrium reactions, and ordinary differential equations describing the kinetic reactions. In this context, we have developed and implemented in DuMu^X a one-phase multicomponent transport module. As starting point we used the single-phase two-component module

(called $1p2c$ in DuMu^X) that solves a one-phase flow of a compressible fluid with two components. The primary variables are the pressure p and the mole or mass fraction of dissolved component. In our model, the velocity of the liquid phase is given by the two-phase two-component flow, so we have first removed the liquid pressure from the set of primary variables. Then, we have increased the number of dissolved components from two to N and named this new module $1pNc$ (one-phase N-component). In [3], the reactive transport is solved by a sequential iterative Approach (SIA) using the $1pNc$ module and a home code ChemEqlib [8] to solve the chemistry by taking into account only the equilibrium reactions and a specific treatment to manage the precipitation of the minerals at equilibrium. In [2], still considering a sequential iterative approach, the computation of chemistry problem has been directly introduced in the Dumu^X framework. So the $1pNc$ and ChemEqlib code have been replaced by the $1pNc-react$ module. This new module takes into account the kinetic reactions for minerals and precipitation/dissolution at equilibrium using the Fischer-Burmeister complementarity function. To improve the robustness of the scheme and the accuracy loss due to the time-splitting involved by the sequential iterative approach, in this paper, we switched to a global implicit approach (GIA) for reactive transport subsystem. More precisely, we used a direct substitution approach (DSA) [10]. Again, the spatial discretization is carried out using a finite volume method and the numerical convective and diffusive flux are calculated by the same scheme as for the $2p2c$ module.

Finally, an efficient coupling between the modules $2p2c$ and $1pNc-react$ has been developed and integrated in DuMu^X. The accuracy and effectiveness of this new simulator is demonstrated through numerical investigation.

3 Numerical Results

We have implemented our approach in the framework of DuMu^X. Numerical experiments, to test the method, have been performed on a variety of 2D and 3D problems including gas migration in a nuclear waste repository or long-term fate of injected CO_2 for geological sequestration. The results obtained are satisfactory and the numerical computations for the coupled system have demonstrated that this approach yields physically realistic flow fields in highly heterogeneous media. The SIA and DSA algorithms have been compared for the reactive transport model. It highlighted the advantages of the DSA algorithm. Firstly, it allowed the use of larger time steps and secondly, the errors of mass conservation due to the operator splitting have been reduced. In the sequel, we present a simulation to illustrate our results.

We consider a test case of H_2 gas generation in a nuclear waste repository presented in [9]. In this work the authors use the general purpose reactive transport simulator, TOUGHREACT to solve chemistry model and a two-phase flow model to study H_2 gas generation, pressure build-up, and saturation distribution in a nuclear waste repository. The chemical reactions among aqueous, gaseous and mineral

Table 1 Chemical reactions and thermodynamic data

Aqueous Reactions	$log(K^e)$
$OH^- \rightleftharpoons H_2O$ - H^+	−13.9951
$Fe^{3+} \rightleftharpoons H^+ + Fe^{2+}$-$0.5H_2$(l)	−13.823
$Fe(OH)_2$(l) $\rightleftharpoons 2H_2O$ -$2H^+$ +Fe^{2+}	−20.60
$Fe(OH)_2^+ \rightleftharpoons 2H_2O$ -H^+ +Fe^{2+}-$0.5H_2$(l)	−19.493
$Fe(OH)_3$(l) $\rightleftharpoons 3H_2O$ -$2H^+$ +Fe^{2+}-$0.5H_2$(l)	−25.823
$Fe(OH)_3^- \rightleftharpoons 3H_2O$ -$3H^+$ +Fe^{2+}	−31.00
$Fe(OH)_4^- \rightleftharpoons 4H_2O$ -$3H^+$ +Fe^{2+}-$0.5H_2$(l)	−34.823
Mineral Reactions	$log(K^e)$
$Fe_3O_4 \rightleftharpoons 3Fe^{2+}$ + $4H_2O$-$6H^+$ -H_2(l)	38.814
$Fe \rightleftharpoons Fe^{2+}$-$2H^+$ +H_2(l)	11.6572
Gaseous Reaction	
H_2(g) $\rightleftharpoons H_2$(l)	

Table 2 Initial conditions of major and mineral components

Ion	Init. molality	Mineral	Initial Volume
	[mol.kg$^-$1]		fraction [–]
H^+	1×10^{-7}	Fe	0.72
Fe^{2+}	1×10^{-4}	Fe_3O_4	0.18

phases considered in this test are summarized in Table 1. Homogeneous reactions are assumed at local equilibrium. Iron (Fe) dissolution and magnetite (Fe_3O_4) precipitation are considered under kinetic conditions and the iron corrosion reaction can be expressed as:

$$3Fe + 4H_2O \rightleftharpoons Fe_3O_4 + 4H_2. \tag{9}$$

The initial conditions of major and mineral components are given in Table 2.

The geometry of the repository is represented by a radially-symmetric domain (see Fig. 1). The system is assumed to be initially fully saturated with water with a pressure of 65 bar. The outer boundary at a radial distance of 75 m was prescribed with a constant pressure of 65 bar. Constant aqueous concentrations are specified at the outer boundary and no flow in the rest. A constant temperature of 40 °C was used. The physical parameters used for the different materials are summarized in Table 3. The gas phase is treated as an ideal mixture: steam and hydrogen are assumed as ideal gas. For the liquid phase density, we used the law implemented in DuMu^X which is taken from the International Formulation Committee.

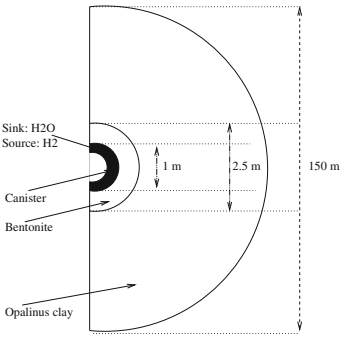


Fig. 1 Radially-symmetric domain representing canister, bentonite and opalinus clay host

Table 3 Physical parameters used for the different materials

	Waste Canister	Bentonite	Opalinus Clay
Porosity [-]	0.1	0.4	0.12
Permeability [m^2]	10^{-19}	10^{-19}	10^{-20}
Two phase parameter model	Van Genuchten	Van Genuchten	Van Genuchten
Residual liquid saturation [-]	0.0	0.3	0.5
Residual gas saturation [-]	0.0	0.0	0.0
Van Genuchten parameter n [-]	2.0	1.82	1.67
Gas entry pressure [Pa]	1.0	1.8×10^7	1.8×10^7
Liquid diffusion/dispersion tensor	$D_l = 1.\ 10^{-9}\ m^2$		

In the kinetic reactions, the dissolution and precipitation rate is expressed by

$$r = k_s A_s \left[1 - \frac{Q}{K^e}\right],$$

with k_s is the kinetic rate constant for the mineral, A_s is the total reactive surface area for the canister, K^e is the equilibrium constant for the mineral, and Q is the reaction quotient. The surface area A_s can be calculated from $A_s = A_0 \frac{V}{V_0} f(S_l)$, where A_0 is the initial surface area (for iron dissolution $A_0 = 121.8\,cm^3/g$), V_0 and V are mineral volume fractions at initial and current times, respectively. The factor f depends on the water saturation S_l as $f(S_l) = (S_l)^n$. A series of simulations was performed to evaluate sensitivity of H_2 generation rates for different values of n ranging from zero ($f = 1$) to two ($f < 1$). Two cases of simulations for a time period of 5000 years based on the two iron dissolution rate constants of $k_s = 2.0 \times 10^{-12}$ and $k_s = 2.0 \times 10^{-11}$ mol/m^2/s are presented in [9]. Here, we present only a simulation based on the lower iron dissolution rate $k_s = 2.0 \times 10^{-12}$. We reproduce in Fig. 2 the various numerical results obtained by using our simulator.

The numerical results presented in Fig. 2 are very close to those in [9]. In the base-case simulation using $n = 0$ which means that the corrosion rate does not depend on

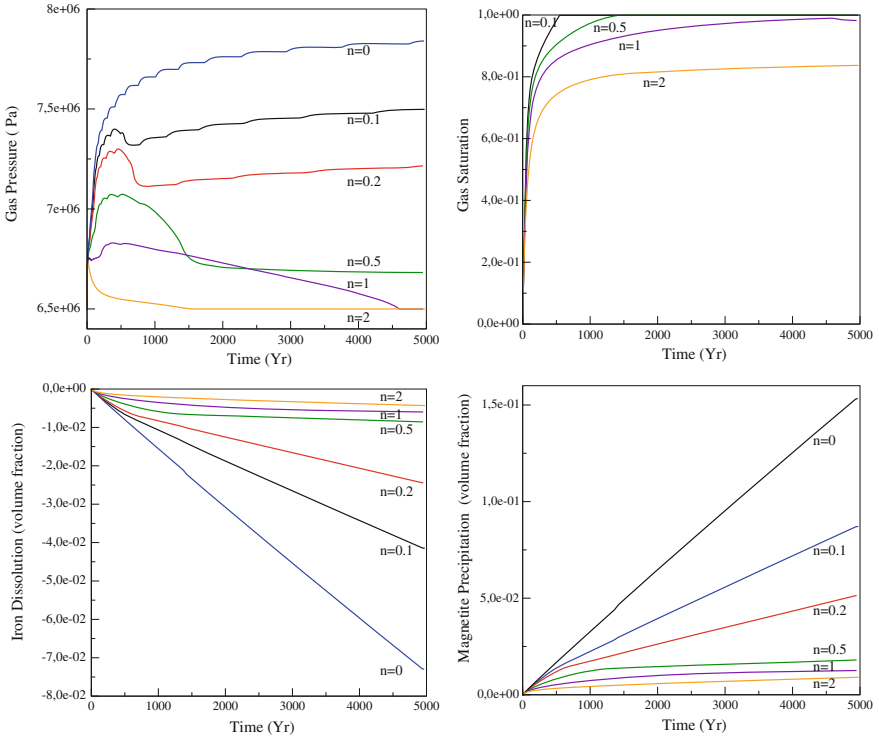


Fig. 2 Gas pressure, gas saturation, iron dissolution and magnetite precipitation at the canister surface for different n values

the water saturation, the pressure increases continuously during the simulation time period of 5000 years to about 80 bar. The same behaviour was captured for $n = 0.1$ and 0.2, the gas pressure increases but slowly compared to the base-case. For other values of n, the corrosion rate and the corresponding H_2 generation rate depend strongly on the water saturation. The pressure increases initially up to a maximum value of 71 bars and decreases to a final stabilized pressure due to the diminution in water saturation and gas generation rate. In the base-case simulation ($n = 0$) the iron volume fraction is reduced about 7% after 5000 years. In the other simulations ($n > 0$), the amounts of dissolved iron and precipitated magnetite are reduced.

Acknowledgements This work was partially supported by the Carnot Institute, ISIFoR project (Institute for the sus-tainable engineering of fossil resources) and CDAPP (Agglomeration Community of Pau-Pyrenees). Their supports are gratefully acknowledged. We also thank the DuMu^X and DUNE teams for their help during the development of our reactive transport module.

References

1. Dune, the distributed and unified numerics environment (2016). http://www.dune-project.org
2. Ahusborde, E., El Ossmani, M.: A sequential approach for numerical simulation of two-phase multicomponent flow with reactive transport in porous media. Math. Comput. Simul. (2016). doi:10.1016/j.matcom.2016.11.007
3. Ahusborde, E., Kern, M., Vostrikov, V.: Numerical simulation of two-phase multicomponent flow with reactive transport in porous media: application to geological sequestration of CO_2. ESAIM Proc. Surv. **50**, 21–39 (2015)
4. Flemisch, B., Darcis, M., Erbertseder, K., Faigle, B., Lauser, A., Mosthaf, K., Mthing, S., Nuske, P., Tatomir, A., Wolff, M., Helmig, R.: Dumu^x, dune for multi-phase, component, scale, physics, ... flow and transport in porous media. Adv. Water Res. **34**(9), 1102–1112 (2011). http://www.dumux.org
5. Helmig, R.: Multiphase Flow and Transport Processes in the Subsurface: A Contribution to the Modeling of Hydrosystems. Springer, Heidelberg (1997)
6. Lasaga, A.C., Soler, J., Ganor, J., Burch, T., Nagy, K.: Chemical weathering rate laws and global geochemical cycles. Geochim. Cosmochim. Acta **58**, 2361–2386 (1994)
7. Saaltink, M., Ayora, C., Carrera, J.: A mathematical formulation for reactive transport that eliminates mineral concentrations. Water Resour. Res. **34**, 1649–1656 (1998)
8. Vostrikov, V.: Simulation numérique d'écoulements diphasiques immiscibles compressibles avec transport réactif en milieux poreux. Université de Pau et des Pays de l'Adour, Thèse (2014)
9. Xu, T., Senger, R., Finsterle, S.: Corrosion-induced gas generation in a nuclear waste repository: Reactive geochemistry and multiphase flow effects. Appl. Geochem. **23**, 3423–3433 (2008)
10. Yeh, G., Tripathi, V.: A critical evaluation of recent developments in hydrogeochemical transport models of reactive multi-chemical components. Water Resour. Res. **25**, 93–108 (1989)

Nonlinear Finite-Volume Scheme for Complex Flow Processes on Corner-Point Grids

Martin Schneider, Dennis Gläser, Bernd Flemisch and Rainer Helmig

Abstract The numerical simulation of subsurface processes requires efficient and robust methods due to the large scales and the complex geometries involved. In this article, a nonlinear finite-volume scheme is presented and applied to non-isothermal two-phase two-component flow in porous media. The idea of the scheme and the model used for the simulations are outlined and a comparison to a standard scheme used in industrial codes is made. Large-scale offshore CO_2 storage in the Johansen formation serves as a benchmark problem, where it is demonstrated that the new scheme can handle highly complex corner-point grids and reproduces the physical processes with a higher accuracy than the standard discretization scheme.

Keywords Finite-volume method · Monotone discretization · Corner-point grid · Challenging grids

MSC (2010): 65M08 · 65N08 · 35Q30

1 Introduction

The overall goal of Carbon Capture and Storage (CCS) is the mineral trapping of anthropogenic CO_2 in suitable geological formations. It can potentially be applied to large stationary point sources, which is why it is widely seen as one of the key technologies towards climate change mitigation [12]. The captured CO_2 can be injected

M. Schneider (✉) · D. Gläser · B. Flemisch · R. Helmig
Institute for Modelling Hydraulic and Environmental Systems (IWS), University of Stuttgart, Pfaffenwaldring 61, 70569 Stuttgart, Germany
e-mail: Martin.Schneider@iws.uni-stuttgart.de

D. Gläser
e-mail: Dennis.Glaser@iws.uni-stuttgart.de

B. Flemisch
e-mail: Bernd.Flemisch@iws.uni-stuttgart.de

R. Helmig
e-mail: Rainer.Helmig@iws.uni-stuttgart.de

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_44

into different types of geological formations as for example depleted oil and gas reservoirs, unmineable coal seams or deep saline aquifers [7], making the technology widely applicable around the globe. This work focuses on the injection into deep saline aquifers, which are the most common and represent the highest storage potential among the above mentioned formation types [7].

Numerical modeling is an important tool for the planning of CO_2 storage projects as it can be used for a first screening of potential storage sites. However, the requirements on these tools are very high. A wide range of physical processes occur in the system with varying importance in space and time. In the injection phase, large pressure gradients and strong viscous and buoyant forces dominate the system. The CO_2 spreads out laterally and rises due to its lower density, eventually accumulating below the caprock and forming a thin layer of CO_2-rich gaseous phase between the brine and the caprock. Over time, the CO_2 slowly dissolves in the brine and is further transported via gravity and diffusion. Geochemical processes can promote the mineralization of the CO_2, which in turn leads to changes in the hydraulic properties of the porous medium.

The extremely large temporal and spatial scales and the strongly nonlinear behavior of the equations require robust and efficient modeling techniques. Furthermore, highly complex corner-point grids are often used to spatially discretize the difficult geometries of geological systems. The numerical techniques therefore have to be capable of handling non-planar, non-matching or degenerated faces. The standard scheme used in industrial codes is the cell-centered finite-volume scheme with two-point flux (TPFA) approximation, an efficient scheme that produces unconditionally monotone solutions. However, large errors in face fluxes are introduced on unstructured grids. In this work, the authors present a nonlinear finite-volume scheme applicable to corner-point grids, which maintains the monotonicity property, but has superior qualities with respect to face-flux accuracy.

The main advantages of monotone schemes in comparison to non-monotone schemes are the reliability of the numerical solution in terms of physical correctness, e.g. the positivity of the solution, and the higher robustness in terms of linear and non-linear solver convergence. With increasing model complexity, robustness is essential to maintain efficiency.

2 Nonlinear Two-Point Flux Approximation

In this section, a nonlinear cell-centered finite-volume scheme, namely the nonlinear two-point flux approximation (NLTPFA), is briefly described, based on the following elliptic equation:

$$-\nabla \cdot (\mathbf{K} \nabla p) = q, \qquad \text{in } \Omega, \tag{1}$$

with the computational domain $\Omega \subset \mathbb{R}^{\dim}$ and the permeability tensor $\mathbf{K}$. Additionally, Dirichlet and Neumann boundary conditions are set on the boundary $\Gamma := \partial\Omega$. Discretization of Eq. (1) reads as follows: Find $\mathbf{p} \in \mathbb{R}^{n_e}$ such that

$$\sum_{\sigma \subset \partial V_i} f_{i,\sigma} = Q_i, \quad f_{i,\sigma} = \sum_{j \in \mathscr{S}_\sigma} t_j p_j, \quad \forall\, V_i \subset \Omega, \tag{2}$$

with faces σ, transmissibility coefficients t_j, integrated source or sink terms Q_i and face stencils $\mathscr{S}_\sigma$. The control volumes V_i form a partition of Ω into $\mathrm{n_e}$ elements, such that $\mathring{V}_i \cap \mathring{V}_j = \emptyset, \forall i \neq j$ and $\Omega = \bigcup_i^{\mathrm{n_e}} V_i$.

Finite-volume schemes differ in the way how the face fluxes $f_{i,\sigma}$ are approximated. For the derivation of the nonlinear two-point flux approximation, we assume that the following approximations have already been constructed

$$\begin{aligned} \tilde{f}_{i,\sigma} &= -\sum_{k \in \mathscr{S}_{i,\sigma}} \alpha_{i,k}(p_k - p_i) \approx -\int_\sigma (\mathbf{K}_i \nabla p) \cdot \mathbf{n}_{ij}\, dS, \\ \tilde{f}_{j,\sigma} &= -\sum_{k \in \mathscr{S}_{j,\sigma}} \alpha_{j,k}(p_k - p_j) \approx -\int_\sigma \left(\mathbf{K}_j \nabla p\right) \cdot \mathbf{n}_{ji}\, dS, \end{aligned} \tag{3}$$

with $\sigma = V_i \cap V_j$, $\mathbf{n}_{ij} = -\mathbf{n}_{ji}$ and the element-wise constant permeability tensors $\mathbf{K}_i, \mathbf{K}_j$. The main steps to calculate these approximations is the decomposition of the conormal $\mathbf{d} := \mathbf{K}\mathbf{n}$ [17] and the usage of additional interpolation rules, e.g. the harmonic averaging point interpolation [2]. The coefficients $\alpha_{i,k}, \alpha_{j,k}$ are determined by the conormal decomposition. The positivity of these coefficients cannot be guaranteed for highly complex grids. We have recently extended these concepts to allow for negative coefficients by using ideas presented in [11] and the combination with optimization techniques, [15]. The final face flux approximation is given by

$$f_{i,\sigma} := \mu_{i,\sigma} \tilde{f}_{i,\sigma} - \mu_{j,\sigma} \tilde{f}_{j,\sigma}, \quad f_{j,\sigma} := -f_{i,\sigma}, \tag{4}$$

$$\mu_{i,\sigma} + \mu_{j,\sigma} = 1, \quad 0 \leq \mu_{i,\sigma} \leq 1, \tag{5}$$

with stencil $\mathscr{S}_\sigma = \mathscr{S}_{i,\sigma} \cup \mathscr{S}_{j,\sigma}$. To end up in a nonlinear two-point approximation, the weights $\mu_{i,\sigma}, \mu_{j,\sigma}$ are chosen such that

$$f_{i,\sigma} = t_{i,\sigma} p_i - t_{j,\sigma} p_j, \tag{6}$$

where the transmissibilities $t_{i,\sigma}, t_{j,\sigma}$ are solution dependent, for a detailed derivation see [15]. Under some assumptions it can be shown that the nonlinear transmissibilities are strictly positive from which one can conclude the monotonicity of the scheme [8, 17], which is one of the main reasons for the development of nonlinear schemes. In general, it seems that this additional nonlinearity does not strongly influence the behavior of the scheme compared to linear ones when solving highly nonlinear partial differential equations [14].

The NLTPFA scheme can be extended to discretize a two-phase, two-component, non-isothermal model. The model considers the two components brine and CO_2 in both the liquid and the gas phase and solves one transport equation per component:

$$\phi \frac{\partial \left(\sum_\alpha \rho_\alpha \frac{M^k}{M^\alpha} x_\alpha^k S_\alpha \right)}{\partial t} + \sum_\alpha \nabla \cdot \{ \rho_\alpha \frac{M^k}{M^\alpha} x_\alpha^k \mathbf{v}_\alpha \} - \sum_\alpha \nabla \cdot \{ \rho_\alpha \frac{M^k}{M^\alpha} D_{\alpha,pm}^k \nabla x_\alpha^k \} - \sum_\alpha q_\alpha^k = 0, \qquad k \in \{\mathrm{b}, \mathrm{CO_2}\}, \alpha \in \{\mathrm{l}, \mathrm{g}\}. \tag{7}$$

In addition to that we solve for the energy balance under the assumption of local thermodynamic equilibrium:

$$\phi \frac{\partial \left(\sum_\alpha \rho_\alpha u_\alpha S_\alpha \right)}{\partial t} + (1-\phi) \frac{\partial \left(\rho_s c_s T \right)}{\partial t} + \sum_\alpha \nabla \cdot \{ \rho_\alpha h_\alpha \mathbf{v}_\alpha \} - \nabla \cdot \{ \lambda_{pm} \nabla T \} - q^h = 0, \qquad \alpha \in \{\mathrm{l}, \mathrm{g}\}. \tag{8}$$

The advective velocity $\mathbf{v}_\alpha$ is calculated using the standard multi-phase Darcy approach for the conservation of momentum:

$$\mathbf{v}_\alpha = -\frac{k_{r\alpha}}{\mu_\alpha} \mathbf{K} \left(\nabla p_\alpha - \rho_\alpha \mathbf{g} \right). \tag{9}$$

The system of equations is closed via the constitutive relations for the capillary pressure $p_c = p_g - p_l$ and the relative permeability $k_{r\alpha}$ and using the fact that $S_l + S_g = 1$ and $x_\alpha^b + x_\alpha^{CO_2} = 1$, as well as assuming the components to be in chemical equilibrium between the phases. A definition of all the quantities used in the above equations can be found in [9] and is not given here again. As primary variables, we use the liquid phase pressure p_l and a second variable, either the CO_2-rich phase saturation S_g or the mole fraction $x_l^{CO_2}$ or x_g^b, depending on the actual phase state [9]. A phase switch occurs when the equilibrium concentration of a component in a phase is exceeded and is evaluated after each Newton iteration.

Note that all the fluid properties appearing in (7) and (8) depend non-linearly on the primary variables, for which the functions listed in Table 2.1 in [9] were used. Furthermore, for the effective diffusion coefficients $D_{\alpha,pm}^k$ and the effective thermal conductivity λ_{pm} of the porous medium the relationships of Millington and Quirk [13] and Somerton [16] were applied respectively.

3 Numerical Results

In this section, we investigate the behavior of the NLTPFA scheme (6) on a CO_2 injection scenario into a geological formation in the Norwegian Sea, namely the *Johansen* formation. The data is provided by *SINTEF*.[1] In this article, we are using

[1] https://www.sintef.no/projectweb/matmora/downloads/johansen/

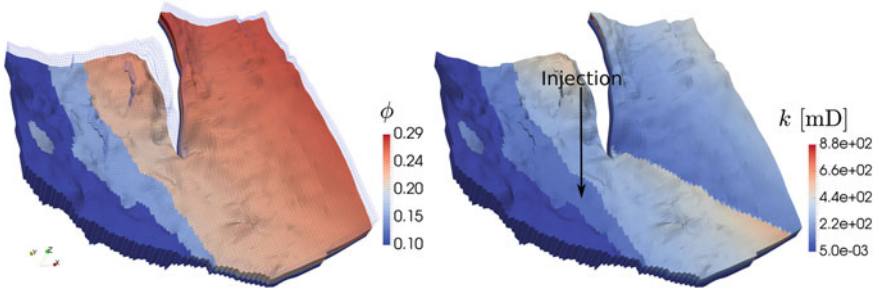


Fig. 1 *Left* porosity distribution and grid of caprock. *Right* permeability distribution and location of injector

Table 1 Parameters used for the simulation of the injection scenario into the Johansen formation

Parameter	Value	Parameter	Value
Simulation time	2000 years	$S_{l,\text{init}}$	1.0
Injection time	100 years	$S_{l,\text{r}}$	0.2
Injection rate	4 $\frac{\text{Mt}}{\text{year}}$	$S_{g,\text{r}}$	0.05
Injection temperature	80°C	λ (Brooks-Corey)	2
Temperature at a depth of 3000 m	100°C	Entry pressure	10^4 Pa

the "NPD5" sector model with porosity and permeability data as shown in Fig. 1. This model consists of eleven layers with five layers representing the highly permeable Johansen formation, situated above a low-permeable shale layer and below five layers describing the caprock (see Fig. 1) [1, 10].

The setting and data are similar to those that have been used on a much simpler grid [6, 14]. It is assumed that the formation is initially fully-saturated with brine and supercritical CO_2 is injected into layers 6,7 and 10 over a period of 100 years with a total injection rate of 4 Mt/year, followed by a period of 1900 years without injection. Important simulation parameters are listed in Table 1. Similar scenarios can be found in [1, 10] but without using a consistent, second-order scheme.

In the following, the linear TPFA, which is the industry standard for solving flow problems on corner-point grids, is compared with the NLTPFA (see (6)). The fully-implicit solution strategy is used for solving the above partial differential equations, namely, the temporal derivatives are discretized with the implicit Euler scheme. Newton's method is applied as nonlinear solver together with a stabilized bi-conjugate

Table 2 Discrete error norms for hydrostatic pressure profile

scheme	NLTPFA	TPFA
e_p	2.52e-14	9.51e-03
e_v	7.50e-10	3.30e+01

gradient (BiCGSTAB) method with an algebraic multigrid preconditioner [4] to solve the occurring linear systems of equations in each Newton iteration. The simulations are performed using our in-house open-source simulator DuMu^x [3], which comes in the form of an additional *DUNE* module [5]. To read in the corner-point grid data, the *opm-grid* module from the *Open Porous Media (OPM) initiative*[2] has been used.

A good accuracy indicator on such complex grids is the linearity-preservation property. For this purpose, we solve Eq. (1) on the Johansen grid with homogeneous porosity and permeability. Furthermore, hydrostatic pressure is chosen as Dirichlet condition on the whole domain boundary. That means that the exact solution in the domain is given by the hydrostatic pressure profile. The discrete relative pressure and velocity L^2-errors, as defined in [15], are listed in Table 2 for the TPFA and the NLTPFA scheme.

It can clearly be seen that the NLTPFA scheme reproduces the exact solution, whereas the errors of the TPFA scheme are approximately eleven orders of magnitude higher.

At the end of this section, we solve the system (7) and (8) on the Johansen formation, where Neumann no-flow conditions are specified at the upper and lower boundaries, elsewhere Dirichlet conditions are set. The solution of the NLTPFA scheme at times $t \in \{100\,\text{a}, 400\,\text{a}, 800\,\text{a}, 2000\,\text{a}\}$ is shown in Fig. 2 (left), Fig. 2 (right) depicts the difference to the TPFA scheme. After 100 years, the CO_2 plume is located near the injection well. Within this time period, transport of CO_2 is caused by high pressure gradients around the injection well. After the injection has ceased, the CO_2 transport is mainly driven by buoyancy forces. Therefore, CO_2 rises in directions of large z-gradients and accumulates below the caprock. It can be observed that the TPFA and the NLTPFA schemes differ at the plume fronts, with differences above 50%. However, it seems that also the TPFA scheme results in an acceptable solution. In general, it is well-known that the TPFA scheme is inconsistent for non-K-orthogonal grids or anisotropic permeability tensors. Therefore, for more complex geological formations, where the non-K-orthogonality is more severe, the difference between these schemes is expected to be more significant [15].

[2] http://opm-project.org/

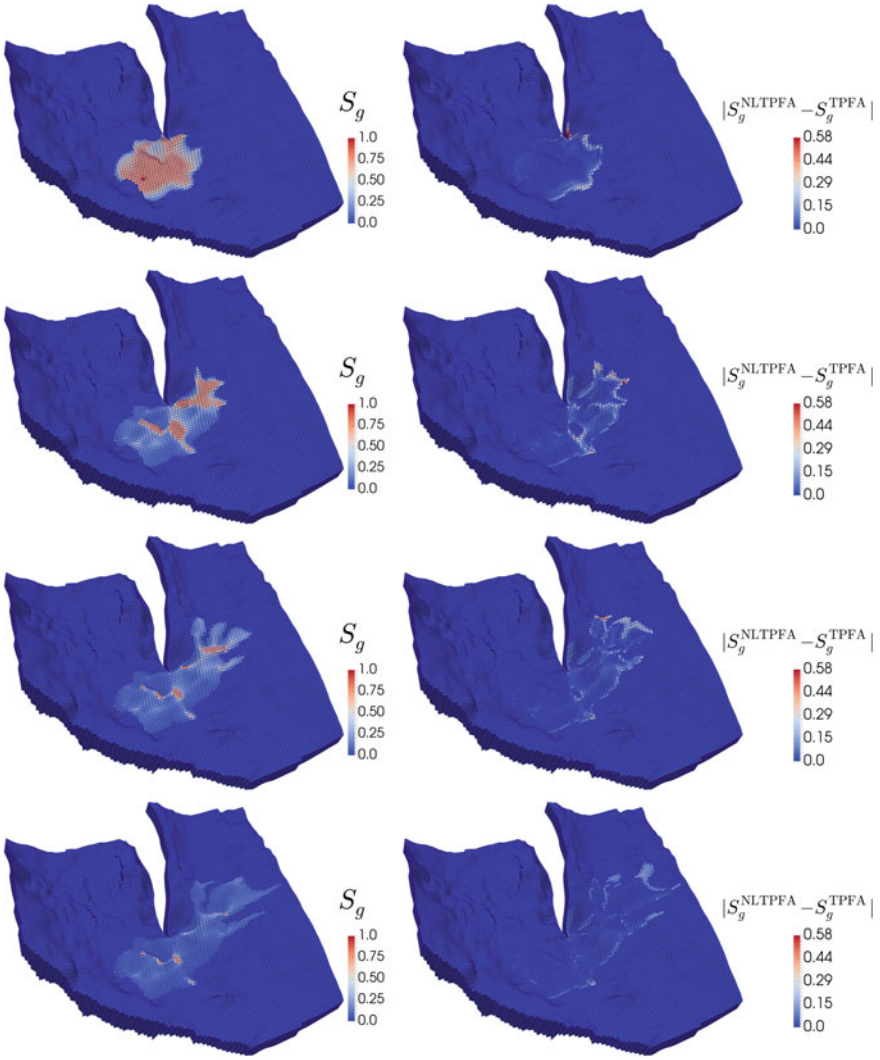


Fig. 2 *Left* solution of NLTPFA scheme after 100, 400, 800 and 2000 years (from *top* to *bottom*). *Right* difference of saturation profiles of NLTPFA and TPFA scheme

4 Conclusion

A monotone nonlinear cell-centered finite-volume scheme has been presented. It can be applied to complex corner-point grids with non-convex and degenerated cell geometries as well as non-matching faces, all of which occur in grids of geological formations. The scheme has been briefly introduced and the main ideas have been summarized. It has been shown that the scheme is exact for linear solutions on such

complex grids, demonstrating the higher accuracy of the NLTPFA in comparison to the linear TPFA scheme. Finally, the method has been applied to a CO_2 storage injection scenario into the Johansen formation. To our knowledge, this is the first cell-centered multi-point flux scheme that is capable of handling such complex grids and processes.

Acknowledgements The authors would like to thank the German Research Foundation (DFG) for financial support of the project within the Cluster of Excellence in Simulation Technology (EXC 310/2) at the University of Stuttgart.

References

1. Afanasyev, A.A.: Application of the reservoir simulator MUFITS for 3D modelling of CO2 storage in geological formations. Energy Proced. **40**, 365–374 (2013)
2. Agélas, L., Eymard, R., Herbin, R.: A nine-point finite volume scheme for the simulation of diffusion in heterogeneous media. Compt. Rendus Math. **347**(11), 673–676 (2009)
3. Becker, B., Beck, M., Fetzer, T., Flemisch, B., Grüninger, C., Hommel, J., Jambhekar, V., Kissinger, A., Koch, T., Schneider, M., Schröder, N., Schwenck, N.: Dumux 2.7.0 (2015). doi:10.5281/zenodo.16722, URL http://dx.doi.org/10.5281/zenodo.16722
4. Blatt, M., Bastian, P.: The Iterative Solver Template Library, pp. 666–675. Springer, Heidelberg (2007)
5. Blatt, M., Burchardt, A., Dedner, A., Engwer, C., Fahlke, J., Flemisch, B., Gersbacher, C., Gräser, C., Gruber, F., Grüninger, C., Kempf, D., Klöfkorn, R., Malkmus, T., Müthing, S., Nolte, M., Piatkowski, M., Sander, O.: The distributed and unified numerics environment, version 2.4. Arch. Numer. Softw. **4**(100), 13–29 (2016)
6. Class, H., Ebigbo, A., Helmig, R., Dahle, H.K., Nordbotten, J.M., Celia, M.A., Audigane, P., Darcis, M., Ennis-King, J., Fan, Y., Flemisch, B., Gasda, S.E., Jin, M., Krug, S., Labregere, D., Naderi Beni, A., Pawar, R.J., Sbai, A., Thomas, S.G., Trenty, L., Wei, L.: A benchmark study on problems related to CO2 storage in geologic formations. Comput. Geosci. **13**(4), 409–434 (2009)
7. Coninck, H., Loos, M., Metz, B., Davidson, O., Meyer, L.: IPCC special report on carbon dioxide capture and storage. Intergovernmental Panel on Climate Change (2005)
8. Danilov, A., Vassilevski, Y.V.: A monotone nonlinear finite volume method for diffusion equations on conformal polyhedral meshes. Rus. J. Numer. Anal. Math. Model. **24**(3), 207–227 (2009)
9. Darcis, M.Y.: Coupling models of different complexity for the simulation of CO2 storage in deep saline aquifers. Ph.D. thesis (2013). URL http://elib.uni-stuttgart.de/opus/volltexte/2013/8141
10. Eigestad, G.T., Dahle, H.K., Hellevang, B., Riis, F., Johansen, W.T., Øian, E.: Geological modeling and simulation of CO2 injection in the Johansen formation. Comput. Geosci. **13**(4), 435–450 (2009)
11. Gao, Z., Wu, J.: A second-order positivity-preserving finite volume scheme for diffusion equations on general meshes. SIAM J. Scient. Comput. **37**(1), A420–A438 (2015)
12. Intergovernmental panel on climate change: climate change 2014: mitigation of climate change, vol. 3, Cambridge University Press (2015)
13. Millington, R.J., Quirk, J.P.: Permeability of porous solids. Trans. Faraday Soc. **57**, 1200–1207 (1961)
14. Schneider, M., Flemisch, B., Helmig, R.: Monotone nonlinear finite-volume method for non-isothermal two-phase two-component flow in porous media. Int. J. Numer. Methods Fluids (2016)

15. Schneider, M., Flemisch, B., Helmig, R., Terekhov, K., Tchelepi, H.: Monotone nonlinear finite-volume method for challenging grids (2017). (SimTech preprint). URL http://www.simtech.uni-stuttgart.de/publikationen/prints.php?ID=1507
16. Somerton, W., El-Shaarani, A., Mobarak, S.: High temperature behavior of rocks associated with geothermal type reservoirs. In: SPE California Regional Meeting. Society of Petroleum Engineers (1974)
17. Yuan, G., Sheng, Z.: Monotone finite volume schemes for diffusion equations on polygonal meshes. J. Comput. Phys. **227**(12), 6288–6312 (2008)

Consistent Nonlinear Solver for Solute Transport in Variably Saturated Porous Media

Daniil Svyatskiy and Konstantin Lipnikov

Abstract We propose a new Jacobian-free solver for the system of nonlinear equations describing transport of a nonreactive solute in porous media. Maximum principles (MPs) are important properties of the solution and impose severe requirements on the discretization and nonlinear solver. Exact local water balance is needed to show the MP for the solute concentration. The proposed solver guarantees the discrete MPs for the solute with the tolerance of linear solvers (typically 10^{-10}) even when the tolerance of the nonlinear solver is not tight (typically 10^{-5}). The proposed technique combines the stable discretization of the Fréchet derivative of the continuous functional describing the system with the slope-limiting algorithm for water retention models.

Keywords Jacobian-free solver · Richards equation · Solute transport · Discrete maximum principle

MSC (2010): 65N08 · 65N06 · 65H10 · 35B50

1 Introduction

Physically based modeling of coupled flow and transport is important for developing soil remediation strategies and for assessment of potential impact of pollutant on groundwater ecosystems. In these applications, Richards' equation [8] is often used to model variably saturated subsurface flow. It is coupled with a convection-diffusion equation for transport which results in a coupled system of nonlinear parabolic PDEs.

Historically, many important contributions to the field were made using orthogonal meshes where simple two-point flux approximation (TPFA) schemes provide the

D. Svyatskiy (✉) · K. Lipnikov
LANL, Theoretical Division, Los Alamos, NM 87544, USA
e-mail: dasvyat@lanl.gov

K. Lipnikov
e-mail: lipnikov@lanl.gov

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_45

second-order accuracy. Accurate modeling of complex surface topography, soil and bedrock horizonts that are not typically parallel to this topography, and various engineering subsurface constructions require unstructured meshes and hence advanced discretization schemes [2]. These schemes preserve important mathematical and physical properties of the underlying PDEs at the expense of additional complexity. Since the exact Jacobian may be hard to calculate for the advanced discretizations, and more complex physical models, nonlinear solvers with an approximate Jacobian or even Jacobian-free solvers are desired.

In the considered model, both water pressure and solute concentration satisfy the maximum principles (MPs), but in this paper we focus on the later. It has been shown in [3] that discrete MPs for the solute concentration relies on the exact water balance. Earlier it was shown in [1] that for Richards' equation the low-order implicit discretization in time of its mixed-form, in concert with a finite volume discretization in space, conserves mass if the nonlinear equation is solved exactly. Violations of this balance are often on the order of a nonlinear residual, typically about 10^{-5}; hence, more iterations of a nonlinear solver is needed to improve this balance. Most nonlinear solvers with an approximate Jacobian converge only geometrically which increases significantly the number of excessive iterations.

Since performance of a nonlinear solver becomes crucial not only for accuracy of a numerical solution but also for its monotonicity, we propose to control solution properties via the specially designed nonlinear solver. We describe the consistent nonlinear solver (CNLS) that guarantees the water balance up to tolerance of linear solvers, typically about 10^{-10}.

The paper outline is as follows. In Sect. 1, we introduce the governing equations. In Sect. 2, we discretize them and show how the exact water balance leads to the discrete MPs for a nonreactive solute. In Sect. 3, we describe the CNLS. In Sect. 4, we apply this solver to a challenging problem of water infiltration into a heterogeneous porous medium.

2 Model Description

Let Ω be a two-dimensional polygonal domain with the Lipschitz boundary $\Gamma = \Gamma_D \cup \Gamma_N$ and $t \in (0, T)$, $T > 0$. The dispersive transport of a nonreactive solute is modeled by the system of equations for the unknown water pressure, p, and solute concentration, C. Under the assumption of the constant water density, we have

$$\frac{\partial \theta(p)}{\partial t} + \operatorname{div}(\mathbf{q}) = 0, \quad \mathbf{q} = -\mathbf{K}k(p)\big(\nabla p - \rho \mathbf{g}\big), \tag{1}$$

$$\frac{\partial \big(\theta(p)\, C\big)}{\partial t} + \operatorname{div}(\mathbf{q}C) - \operatorname{div}(\theta \mathbf{D}(\mathbf{q})\nabla C) = 0 \tag{2}$$

where $\mathbf{q}$ the Darcy velocity, $\mathbf{g}$ the gravity vector, and θ the water content. The mechanical properties of soil are described by the absolute permeability tensor $\mathbf{K}$, the relative

permeability $k(p)$, and the dispersion tensor $\mathbf{D}$. Typically, the relative permeability is a non-decreasing function of pressure, $\frac{\partial k}{\partial p} \geq 0$. The water retention model (WRM) defines nonlinear dependencies $k(p)$ and $\theta(p)$ for different soils. We consider the water retention models defined by van Genuchten and Mualem equations [4]. Let

$$p = g_D(\mathbf{x}) \quad \text{on} \quad \Gamma_D, \qquad \mathbf{q} \cdot \mathbf{n} = g_N(\mathbf{x}) \quad \text{on} \quad \Gamma_N. \tag{3}$$

We denote by Γ_{out} the outflow part of Γ where $\mathbf{q} \cdot \mathbf{n} \geq 0$ and set $\Gamma_{in} = \Gamma / \Gamma_{out}$. Then,

$$C = g_{in}(\mathbf{x}) \quad \text{on} \quad \Gamma_{in}, \qquad \mathbf{n} \cdot (\mathbf{q}\, C - \theta \mathbf{D}\, \nabla C) = g_{out}(\mathbf{x}) \quad \text{on} \quad \Gamma_{out}. \tag{4}$$

The system (1)–(4) is closed with appropriate initial conditions.

2.1 Discretization and the Maximum Principles

We consider a time step from t^n to $t^{n+1} = t^n + \Delta t^n$. Let Ω_h be a polytopal mesh with cells c and faces f. We use $|c|$ and $|f|$ to denote the cell volume and face area, respectively. We employ the staggered discretization where the Darcy velocity is defined on faces, one number q_f per face, and all other variables are defined on cells, one number (p_c, C_c, and θ_c) per cell. Advance discretizations may introduce additional degrees of freedom on faces, p_f. Let $\mathbf{q}_h$, p_h and C_h be the global vectors which combine the corresponding degrees of freedom for all cells and faces.

The governing equations will be solved sequentially on each time step. First, we apply the implicit backward-Euler time discretization to the flow equation (1):

$$\frac{\theta_c^{n+1} - \theta_c^n}{\Delta t^n} + \mathrm{div}_c^h(\mathbf{q}_h^{n+1}) = 0, \quad \mathrm{div}_c^h(\mathbf{q}_h^{n+1}) = \frac{1}{|c|} \sum_{f \in \partial c} \sigma_{c,f}\, q_f^{n+1}, \tag{5}$$

where $\sigma_{c,f} = \pm 1$ depending on the mutual orientation of the fixed and exterior normals on face f of cell c. We use the MFD scheme [5] to calculate the Darcy fluxes.

We use the computed flux and water content to discretize the transport Eq. (2). Its semi-implicit discretization reads

$$\frac{\theta_c^{n+1} C_c^{n+1} - \theta_c^n C_c^n}{\Delta t^n} + \frac{1}{|c|} \sum_{f \in \partial c} \sigma_{c,f}\, q_f^{n+1} C_{c,f}^n + \frac{1}{|c|} \sum_{f \in \partial c} \sigma_{c,f}\, u_f^{n+1} = 0, \tag{6}$$

where $C_{c,f}^n$ denotes the concentration value in an upwind cell, and u_f^{n+1} is the dispersive flux. The direction of upwind is defined by the Darcy flux q_f^{n+1}.

Both the pressure and concentration satisfy the MPs; but here we focus on the discrete MPs for C_h. The scheme (6) can be implemented and analyzed as two separate steps: advection and dispersion. Let us consider a cell c and assume that $\sigma_{c,f} = 1$ for all faces of c. To simplify notations, we write q_f instead of q_f^{n+1}. Using the first-order upwind scheme, the advection step can be written as follows:

$$\frac{\theta_c^{n+1}\widetilde{C}_c^{n+1} - \theta_c^n C_c^n}{\Delta t^n} + \frac{1}{|c|}\sum_{q_f \geq 0} q_f\, C_c^n + \frac{1}{|c|}\sum_{q_f < 0} q_f\, C_{c,f}^n = 0.$$

Eliminating θ_c^{n+1} using formula (5), we obtain

$$\left(\theta_c^n - \frac{\Delta t}{|c|}\sum_{f\in\partial c} q_f\right)\widetilde{C}_c^{n+1} = \left(\theta_c^n - \frac{\Delta t}{|c|}\sum_{q_f>0} q_f\right) C_c^n - \frac{\Delta t}{|c|}\sum_{q_f<0} q_f\, C_{c,f}^n.$$

Thus, $\widetilde{C}_c^{n+1}$ is the convex combination of C_c^n and $C_{c,f}^n$ under the CFL condition

$$|c|\theta_c^n - \Delta t \sum_{q_f>0} q_f > 0.$$

Due to the properties of a WRM and the upwind discretization, outfluxes, $q_f > 0$, tend to zero if θ_c^n tends to zero, so the CFL condition is not degenerate. Moreover, in many applications a WRM is defined such that $\theta_c^n \geq \theta_{min} > 0$, so a soil can not be completely dry. Thus, $\widetilde{C}_c^{n+1}$ has no internal extrema provided that the flow Eq. (5) is solved exactly and the CFL condition is satisfied. Typically, (5) is solved by an iterative nonlinear solver, so the discrete water balance holds up to the tolerance of this solver. Later, we show how to improve significantly the discrete water balance.

Finally, the dispersion step reads:

$$\frac{\theta_c^{n+1}C_c^{n+1} - \theta_c^{n+1}\widetilde{C}_c^{n+1}}{\Delta t^n} + \frac{1}{|c|}\sum_{f\in\partial c} u_f^{n+1} = 0.$$

Any discretization scheme for the dispersive flux that preserves the MPs can be used here [6]. A detailed presentation of such a scheme is beyond the scope of this paper.

3 New Consistent Nonlinear Solver

Earlier, we emphasized importance of the exact water balance (5) for the discrete MPs for the concentration. In conventional nonlinear solvers, the Darcy flux is calculated approximately and the water balance holds up to the tolerance of this solver, typically about 10^{-5}. To guarantee a monotone concentration field C_h, this tolerance should

much tighter, about 10^{-10}. Thus, the performance of a nonlinear solver becomes crucial aspect for the overall quality of the numerical solution.

On orthogonal meshes, where TPFA method is accurate, the Newton-Raphson method with the exact Jacobian may double accuracy of each iteration if an iteration is close enough to the solution. Geometric complexity of the subsurface environment introduces challenges which can be handled only with advanced discretization schemes. The complex structure of these schemes makes computation of the exact Jacobian matrix either quite costly or extremely complex. These issues have generated a significant interest in development of Jacobian-free, inexact Newton methods and Picard-type method. Unfortunately, simplification of a solver reduces its convergence properties from quadratic to linear, which in turn leads to significant growth of nonlinear iterations to achieve the tolerance of about 10^{-10}.

We propose the new approach for building the approximate Jacobian for flow Eq. (1). This is the only approach which guarantees the local water balance on the order of linear solvers, even when the nonlinear tolerance is not tight, and hence provides the discrete MP for the solute concentration in the transport Eq. (2).

The system of nonlinear Eq. (5) can be written formally as $\mathscr{F}_h(p_h) = 0$, where

$$\mathscr{F}_h(p_h) = \frac{\theta_h(p_h) - \theta_h^n}{\Delta t^n} + \mathrm{div}^h(\mathbf{q}_h(k_h, p_h)).$$

We can write $\mathscr{F}_h(p_h) = \mathscr{A}_h(p_h) + \mathbb{D}_h(p_h)\, p_h$, where $\mathscr{A}_h(p_h)$ is the accumulation term and $\mathbb{D}_h(p_h)$ is the matrix operator representing diffusion term $\mathrm{div}^h(\mathbf{q}_h(k_h, p_h))$.

Following the framework of inexact Newton solvers, the solution increment on the s-th nonlinear iteration is calculated as follows:

$$p_h^{s+1} = p_h^s + \Delta p_h^s \qquad \Delta p_h^s = -\left(\mathbb{P}^s\right)^{-1} \mathscr{F}_h(p_h^s), \tag{7}$$

where $\mathbb{P}^s$ is a preconditioner which depends on iterate p_h^s. A good preconditioner is designed to improve the contraction properties of $\mathscr{F}_h$. The closer (in some metric) $\mathbb{P}^s(p_h^s)$ to the exact Jacobian matrix of $\mathscr{F}_h(p_h^s)$, the faster the asymptotic convergence rate. To define such a preconditioner, we follow the strategy proposed in [7] which suggests to define $\mathbb{P}^s$ as a stable discretization of the Fréchet derivative of the continuous functional $\mathscr{F}$ given by Eq. (1). Recall that the Fréchet derivative $\mathscr{J}(p^s)$ of $\mathscr{F}$ at point p^s (superscript s emphasizes connection with an iterative solver) acting on a small increment function δp satisfies

$$\|\mathscr{F}(p^s + \delta p) - \mathscr{F}(p^s) - \mathscr{J}(p^s)\delta p\|_Y = o(\|\delta p\|_X)$$

in appropriate Banach spaces X and Y. To avoid technical details, we assume that all functions are sufficiently smooth. Then, X and Y are the spaces of continuously differentiable (in space and time) functions. Let

$$\mathscr{J}(p^s)\,\delta p = \frac{\partial \theta}{\partial p}(p^s)\,\frac{\partial\, \delta p}{\partial t} - \mathrm{div}\left(\mathbf{K}\, k(p^s)\, \nabla \delta p\right) + \mathrm{div}\left(\mathbf{V}^s \delta p\right), \tag{8}$$

where

$$\mathbf{V}^s = -\mathbf{K}\frac{\partial k}{\partial p}(p^s)\big(\nabla p^s - \rho\mathbf{g}\big) = \mathbf{q}^s\,\frac{\partial k}{\partial p}(p^s)\frac{1}{k(p^s)}.$$

Using the Taylor expansion of the water content and relative permeability functions around point p^s we can easily verify that

$$\mathscr{F}(p^s + \delta p) - \mathscr{F}(p^s) - \mathscr{J}(p^s)\delta p = \Psi((\delta p)^2),$$

where Ψ is the bounded operator from X to Y.

The preconditioner $\mathbb{P}^s$ is defined as the stable discretization of the continuous operators in (8). We use the backward Euler scheme for the first term. In this case the discretization matrix coincides with the Jacobian matrix corresponding to functional $\mathscr{A}_h$. For the the second term, we use the MFD scheme, the same one we used to derive functional $\mathscr{F}_h$. For the last term, the advection operator, we use the cell-centered upwind finite volume (FV) scheme. Thus, the preconditioner is defined by this matrix operator

$$\begin{aligned}\mathbb{P}^s\, w_h &= \frac{\partial\theta}{\partial p}(p_h^s)\frac{w_h}{\Delta t^n} + \mathrm{div}^h(\mathbf{q}_h(k_h^s, w_h)) + \mathrm{div}^h_{FV}(V_h\, w_h)\\ &\equiv \big(\mathbb{J}_A(p_h^s) + \mathbb{D}(p_h^s) + \mathbb{C}(p_h^s)\big)\, w_h,\end{aligned}$$

where V_h is the first-order upwind flux corresponding to velocity $\mathbf{V}^s$. The direction of upwind is defined by the advective flux, V_f, and the upwinded pressure value, $p^s_{c,f}$, on face f:

$$V_f = q_f^s\,\frac{\partial k}{\partial p}(p^s_{c,f})\frac{1}{k(p^s_{c,f})}.$$

Note that this formula uses physical quantities available at the current iteration.

With this definition of the preconditioner, we can reformulate one step of the nonlinear solver (7) as the solution the following equation for the next iterates p_h^{s+1}:

$$\left[\frac{\partial\theta}{\partial p}(p_h^s)\frac{\Delta p_h^s}{\Delta t^n} + \frac{\theta(p_h^s) - \theta(p_h^n)}{\Delta t^n}\right] - \mathrm{div}^h\big(\mathbf{q}_h(k_h^s, p_h^{s+1})\big) + \mathrm{div}^h_{FV}\big(V_h\,\Delta p_h^s\big) = 0. \tag{9}$$

After calculating the new iterate p_h^{s+1}, we could define new values $\widehat{\theta}_h^{n+1}$ and $\widehat{\mathbf{q}}_h^{n+1}$, that are close to the values provided by the corresponding WRM, as follows:

$$\widehat{\theta}_c^{s+1} = \theta_c^s + \frac{\partial\theta}{\partial p}(p_c^s)(p_c^{s+1} - p_c^s), \qquad \widehat{q}_f^{s+1} = q_f^{s+1} - V_f\,(p^{s+1}_{c,f} - p^s_{c,f}). \tag{10}$$

For these values, the local mass conservation property is satisfied almost exactly, up to the tolerance of a linear solver. Unfortunately, formula (10) may result in non-physical saturation $s = \theta/\phi$ which must be between the residual saturation $s_{\min}$ and $s_{\max} \le 1$ that may vary from cell to cell. To avoid unphysical overshoots and

undershoots, we limit the derivative in formula (10) as follows:

$$\widehat{\theta}_c^{s+1} = \theta_c^s + \alpha_c^s \frac{\partial \theta}{\partial p}(p_c^s)(p_c^{s+1} - p_c^s), \qquad 0 \le \alpha \le 1.$$

If $||\Delta p_h^{s+1}||_\infty \le \varepsilon_{tol}$, then

$$\alpha_c^s = \begin{cases} \min\left(1, \dfrac{s_c^s - s_{c,min}}{\frac{\partial s}{\partial p}(p_c^s)\,\varepsilon_{tol}}, \dfrac{s_{c,max} - s_c^s}{\frac{\partial s}{\partial p}(p_c^s)\,\varepsilon_{tol}}\right) & \text{if } \dfrac{\partial s}{\partial p}(p_c^s) \neq 0, \\ 1 & \text{otherwise.} \end{cases}$$

The straightforward conclusion from this definition is that $s_{c,min} \le \widehat{s}_c^{s+1} \le s_{c,max}$. To compute coefficients α_c^s an estimate of $||\Delta p_h^{s+1}||_\infty$ is required. Finally, to restore the algebraic consistency between $\widehat{s}_h^{s+1}$, $\widehat{\mathbf{q}}_h^{s+1}$ and p_h^{s+1}, the preconditioner $\mathbb{P}^s$ has to include this limiter. The new solver, referred to as the *consistent nonlinear solver* (CNLS), defined by the following preconditioner

$$\mathbb{P}^s_{CNLS} = \mathrm{diag}\{\alpha_c^s\}\,\mathbb{J}_A(p_h^s) + \mathbb{D}(p_h^s) + \mathbb{C}(p_h^s).$$

Remark 1 It is possible to apply preconditioner $\mathbb{P}_{CNLS}$ and compute θ^{n+1}, $\mathbf{q}_h^{n+1}$ as a post-processing step. However, this requires one auxiliary linear solver:

$$p_h^{s+1} = p_h^s + \Delta p^s, \quad \Delta p^s = -(\mathbb{P}^s_{CNLS})^{-1}\mathscr{F}_h(p_h^s),$$

$$\theta_c^{s+1} = \theta_c^s + \alpha_c^s \left(\frac{\partial \theta}{\partial p}\right)^s \Delta p^s, \quad q_f^{s+1} = \widehat{q}_f^{s+1} = \mathbf{q}_f(k^s, p_h^{s+1}) - V_f\,\Delta p^s.$$

4 Numerical Experiments

In this section we apply the proposed method to solute transport simulation in a heterogeneous variably saturated domain. Let $\Omega = [0, 216] \times [0, 107.52]$ be the computational domain consisting of four soils, Fig. 1 (left). Each soil has different horizontal and vertical permeabilities, porosities and WRM parameters. On the top boundary we defined subset $\Gamma_{I,1} = \{(x, z):\ 75 \le x \le 140, z = 107.52\}$ where the nonreactive solute (the tracer) is infiltrated inside the reservoir with the inward flux $q_I = 0.003$. By default, all values are given in the SI system of units. On the rest of the top boundary, the inward background flux is set to $q_b = 1.5 \cdot 10^{-6}$. The infiltration stops at 0.1 year. The boundary concentration is set to 10^{-6} for the infiltration domain and zero for the rest of the top boundary. On the bottom boundary the pressure is set to the atmospheric value. No flow conditions are set on the remaining boundary.

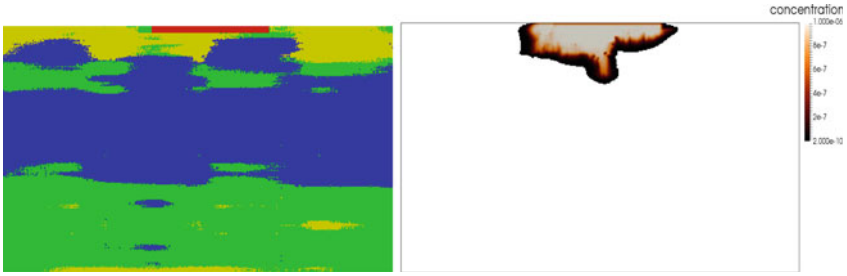


Fig. 1 *Left panel* the computation domain with four different soils. *Right panel* the solute concentration 0.3 year

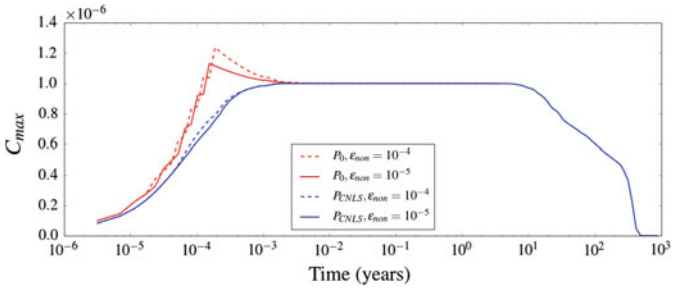


Fig. 2 Maximum of the tracer concentration as a function of time

Since we do not focus on discretization of dispertion fluxes, the dispersion is set to zero.

We compare two solution strategies with preconditioners $\mathbb{P}^s$ and $\mathbb{P}^s_{CNLS}$. For infiltration problems, the convergence criterion is typically given using the maximum norm, in order to capture accurately the fine-scale solution dynamics around the infiltration basin. Our criterion is the combination of the infinity norms of the normalized nonlinear residual and the solution increment:

$$\max\left(\left\|\frac{|\Delta p_h|}{|p_h - p_{atm}| + p_{atm}}\right\|_\infty, \left\|\frac{\Delta t}{\phi}\mathscr{F}_h\right\|_\infty\right) \leq \varepsilon_{non}.$$

In Fig. 2 we present the maximum of the solute concentration as the function of time. According to the MPs, the concentration should be between 0 and 10^{-6}. We observe that the preconditioner $\mathbb{P}^s$ produces large (10–20%) overshoots when the infiltration starts that gradually dissipate. The proposed preconditioner $\mathbb{P}^s_{CNLS}$ provides non-oscillatory solution without overshoots even if when the nonlinear tolerance is not very tight. The overall number of nonlinear iterations is about the same in both strategies.

Acknowledgements This work was carried out under the auspices of the NNSA of the U.S. Department of Energy at LANL under Contract No. DE-AC52-06NA25396. The authors acknowledge

the support of the US DOE Office of Science Advanced Scientific Computing Research (ASCR) Program in Applied Mathematics Research.

References

1. Celia, M.A., Bouloutas, E.T.: A general mass-conservative numerical solution for the unsaturated flow equation. Water Resour. Res. **26**(7), 1483–1496 (1990)
2. Coon, E., Moulton, J., Berndt, M., Manzini, G., Garimella, R., Lipnikov, K., Painter, S.: Coupling surface and subsurface flow using mimetic finite differences. Water Res. (2017)
3. Frolkoviĉ, P.: Maximum principle and local mass balance for numerical solutions of transport equation coupled with variable density flow. Acta Math. Univ. Com. LXVI **I**(1), 137–157 (1998)
4. van Genuchten, M.: A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci. Soc. Am. J. **44**, 892–898 (1980)
5. Gyrya, V., Lipnikov, K., Manzini, G., Svyatskiy, D.: M-adaptation and the mimetic finite difference method. M3AS: Math. Mod. Meth. Appl. Sci. **24**(8), 1621–1663 (2014)
6. Lipnikov, K., Manzini, G., Svyatskiy, D.: Analysis of the monotonicity conditions in the mimetic finite difference method for elliptic problems. J. Comput. Phys. **230**(7), 2620–2642 (2011)
7. Lipnikov, K., Moulton, D., Svyatskiy, D.: New preconditioning strategy for jacobian-free solvers for variably saturated flows with richards' equation. Adv. Water Resour. **94**, 11–22 (2016)
8. Richards, L.: Capillary conduction of liquids through porous mediums. Physics **1**(5), 318–333 (1931)

A Two-Dimensional Complete Flux Scheme in Local Flow Adapted Coordinates

Jan ten Thije Boonkkamp, Martijn Anthonissen and Ruben Kwant

Abstract We present a formulation of the two-dimensional complete flux (CF) scheme in terms of local orthogonal coordinates adapted to the flow, i.e., one coordinate axis is aligned with the local velocity field and the other one is perpendicular to it. This approach gives rise to an advection-diffusion-reaction boundary value problem (BVP) for the flux component in the local flow direction. For the other (diffusive) flux component we use central differences. We will demonstrate the performance of the scheme for several examples.

Keywords Conservation laws · Finite volume method · Numerical flux · Complete flux scheme · Local orthogonal coordinates.

MSC (2010): 65N08 · 65N99

1 Introduction

Conservation laws occur frequently in science and engineering, modeling a wide variety of phenomena, for example laminar flames or gas discharges in plasmas. These conservation laws are often of advection-diffusion-reaction type, describing the interplay between different processes such as advection or drift, diffusion or conduction and (chemical) reactions or impact ionization. We restrict ourselves to stationary two-dimensional conservation laws.

J. ten Thije Boonkkamp (✉) · M. Anthonissen · R. Kwant
Department of Mathematics and Computer Science, Eindhoven University of Technology,
PO Box 513, 5600MB Eindhoven, The Netherlands
e-mail: j.h.m.tenthijeboonkkamp@tue.nl

M. Anthonissen
e-mail: m.j.h.anthonissen@tue.nl

R. Kwant
e-mail: r.j.kwant@student.tue.nl

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_46

The prototypical conservation law reads

$$\nabla\cdot\big(\boldsymbol{u}\varphi-\varepsilon\nabla\varphi\big)=s, \tag{1}$$

where, for example, $\boldsymbol{u}=u\boldsymbol{e}_x+v\boldsymbol{e}_y$ is a flow velocity, $\varepsilon\geq\varepsilon_{\min}>0$ a diffusion coefficient and s a reaction rate. The unknown φ might be the mass fraction of one of the constituent species in a laminar flame or a plasma. Associated with (1) we introduce the flux (vector) $\boldsymbol{f}=f_x\boldsymbol{e}_x+f_y\boldsymbol{e}_y$, given by

$$\boldsymbol{f}=\boldsymbol{u}\varphi-\varepsilon\nabla\varphi. \tag{2}$$

Consequently, the conservation law can be concisely written as $\nabla\cdot\boldsymbol{f}=s$. Integrating this equation over a fixed domain Ω and applying Gauss's law we obtain the integral form of the conservation law, i.e.,

$$\oint_{\partial\Omega}\boldsymbol{f}\cdot\boldsymbol{n}\,\mathrm{d}s=\int_{\Omega}s\,\mathrm{d}A, \tag{3}$$

where $\boldsymbol{n}$ is the unit outward normal on the positively oriented boundary $\partial\Omega$.

For space discretization of (1) we employ the finite volume method (FVM). To that purpose, we introduce grid points $\boldsymbol{x}_{i,j}=(x_i,y_j)$ where φ has to be approximated and control volumes $\Omega_{i,j}=(x_{i-1/2},x_{i+1/2})\times(y_{j-1/2},y_{j+1/2})$ covering the domain. Here $x_{i\pm1/2}=\frac{1}{2}(x_i+x_{i\pm1})$ etc. Taking $\Omega=\Omega_{i,j}$ in (3) and approximating all integrals involved with the midpoint rule, we find

$$\Delta y\big(F_{x,i+1/2,j}-F_{x,i-1/2,j}\big)+\Delta x\big(F_{y,i,j+1/2}-F_{y,i,j-1/2}\big)=\Delta x\,\Delta y\,s_{i,j}, \tag{4}$$

where $F_{x,i+1/2,j}$ is the approximation of f_x at the interface point $(x_{i+1/2},y_j)$ etc. and $s_{i,j}=s(\boldsymbol{x}_{i,j})$. The FVM has to be completed with numerical approximations of all fluxes.

For the numerical flux approximation we employ the complete flux (CF) scheme introduced in [4]. The basic idea of this scheme is to compute the numerical flux from a local *one-dimensional* boundary value problem for the conservation law, including the source term. For one-dimensional problems the scheme gives excellent results and is proven to be uniformly second order convergent [1]. The generalization of this approach to two-dimensional problems is tedious. Instead, the one-dimensional CF scheme is often applied componentwise ignoring the cross-flux terms. For dominant diffusion this version of the scheme is still adequate, however, for dominant advection the scheme suffers from significant numerical diffusion. To remedy this problem, we have included the cross flux as an artificial source term in the local one-dimensional BVPs, virtually eliminating diffusion. This modified scheme is able to reproduce very steep layers in the solution of (1).

However, for three-dimensional conservation laws this approach is rather cumbersome and adds too much anti-diffusion. Therefore we adopt another approach which is expected to be more suitable. Inspired by the skew upstream differencing

schemes introduced in [2], we define a local (ξ, η)-coordinate system adapted to the local velocity $\boldsymbol{u}$ and compute the flux components in this coordinate system. This way, we obtain a ξ-component of the flux parallel to $\boldsymbol{u}$ and an η-component perpendicular to $\boldsymbol{u}$, which can be combined to the normal component of the numerical flux. For both components we use a one-dimensional flux approximation scheme. For the ξ-component we take into account the full advection-diffusion-reaction balance, and include the easy to compute (diffusive) cross flux term as an additional source. On the other hand, for the η-component, we restrict ourselves to the homogeneous flux scheme [4]. The resulting scheme exhibits uniform second order convergence.

We have organized our paper as follows. In Sect. 2 we outline the one-dimensional complete flux scheme, and subsequently in Sect. 3, we present the two-dimensional scheme in local, flow adapted coordinates. Next, in Sect. 4 we demonstrate the performance of the scheme for two examples. Concluding remarks are given in Sect. 5.

2 One-Dimensional Complete Flux Scheme

In this section we outline the one-dimensional version of the complete flux scheme; for more details see [4].

The one-dimensional conservation law can be concisely written as $\mathrm{d}f/\mathrm{d}x = s$ with $f = u\varphi - \varepsilon \mathrm{d}\varphi/\mathrm{d}x$. The integral representation of the flux $f_{j+1/2} = f(x_{j+1/2})$ at the cell edge $x_{j+1/2}$ is based on the following two-point BVP:

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}\left(u\varphi - \varepsilon\frac{\mathrm{d}\varphi}{\mathrm{d}x}\right) = s, \quad x_j < x < x_{j+1}, \tag{5a}$$

$$\varphi(x_j) = \varphi_j, \quad \varphi(x_{j+1}) = \varphi_{j+1}, \tag{5b}$$

consequently, the flux will be the superposition of a homogeneous flux, corresponding to the advection-diffusion operator, and an inhomogeneous flux, taking into account the effect of the source term. Let us first introduce the following variables/notation:

$$a = \frac{u}{\varepsilon}, \quad P = a\Delta x, \quad A(x) = \int_{x_{j+1/2}}^{x} a(z)\,\mathrm{d}z, \quad \langle p, q\rangle = \int_{x_j}^{x_{j+1}} p(x)q(x)\,\mathrm{d}x. \tag{6}$$

We refer to P and A as the (grid) Péclet function and integral, respectively, generalizing the well-know (grid) Péclet number. Using the integrating factor formulation of (5a) we can derive the following representation for the flux:

$$f_{j+1/2} = f^{\mathrm{h}}_{j+1/2} + f^{\mathrm{i}}_{j+1/2}, \tag{7a}$$

$$f^{\mathrm{h}}_{j+1/2} = \left(\mathrm{e}^{-A(x_j)}\varphi_j - \mathrm{e}^{-A(x_{j+1})}\varphi_{j+1}\right)/\langle \varepsilon^{-1}, \mathrm{e}^{-A}\rangle, \tag{7b}$$

$$f^{\mathrm{i}}_{j+1/2} = \Delta x \int_0^1 G(\sigma)s(x(\sigma))\,\mathrm{d}\sigma, \tag{7c}$$

where $f^{\mathrm{h}}_{j+1/2}$ and $f^{\mathrm{i}}_{j+1/2}$ are the homogeneous and inhomogeneous part of the flux, respectively. In (7c) the function G, depending on the scaled coordinate $\sigma(x) = (x - x_j)/\Delta x$, is the so-called the Green's function for the flux, since it relates the source to the flux, different from the usual Green's function, which relates the source to the solution.

To determine the numerical flux $F_{j+1/2}$ we restrict ourselves to constant u and ε. Moreover, we take s piecewise constant, i.e., $s(x) = s_j$ for $x_{j-1/2} < x < x_{j+1/2}$. In this case we can evaluate all integrals involved exactly and find:

$$F_{j+1/2} = F^{\mathrm{h}}_{j+1/2} + F^{\mathrm{i}}_{j+1/2}, \tag{8a}$$

$$F^{\mathrm{h}}_{j+1/2} = \frac{\varepsilon}{\Delta x}\big(B(-P)\varphi_j - B(P)\varphi_{j+1}\big), \tag{8b}$$

$$F^{\mathrm{i}}_{j+1/2} = \Delta x\big(C(-P)s_j - C(P)s_{j+1}\big), \tag{8c}$$

where $P = u\Delta x/\varepsilon$ is the constant (grid) Péclet number. In the expressions in (8) we have introduced the functions B and C defined by $B(z) = z/\big(\mathrm{e}^z - 1\big)$ and $C(z) = \big(\mathrm{e}^{z/2} - 1 - z/2\big)/\big(z\big(\mathrm{e}^z - 1\big)\big)$. In the special case that $P = 0$, i.e., there is no flow, the fluxes in (8) reduce to

$$F^{\mathrm{h}}_{j+1/2} = \frac{\varepsilon}{\Delta x}\big(\varphi_j - \varphi_{j+1}\big), \quad F^{\mathrm{i}}_{j+1/2} = \tfrac{1}{8}\Delta x(s_j - s_{j+1}). \tag{9}$$

The homogeneous flux in (9) is the central difference approximation of the diffusive flux $f^{\mathrm{d}} = -\varepsilon \mathrm{d}\varphi/\mathrm{d}x$, and $F^{\mathrm{i}}_{j+1/2} = \mathscr{O}\big(\Delta x^2\big)$ for $\Delta x \to 0$, provided s is sufficiently smooth. In this case we may omit the inhomogeneous component.

3 Two-Dimensional Complete Flux Scheme

In this section we present an extension of the one-dimensional numerical flux to two-dimensional conservation laws. The basic idea is to decompose the normal component of the numerical flux vector at a cell interface into a component aligned with the (local) velocity field and a component perpendicular to it. We assume that ε is constant.

Consider as an example the computation of the numerical flux component $F_{x,\mathrm{e}} = F_{x,i+1/2,j}$ at the eastern cell interface of the control volume, see Fig. 1 where we adopt the compass notation to denote the location of interface/grid points. Suppose, the basis vector $\boldsymbol{e}_x$ and the local flow velocity $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{x}_{\mathrm{e}})$ enclose an angle $\alpha = \alpha(\boldsymbol{x}_{\mathrm{e}})$ $(-\pi < \alpha \leq \pi)$, oriented counter-clockwise, given by $\tan(\alpha) = v(\boldsymbol{x}_{\mathrm{e}})/u(\boldsymbol{x}_{\mathrm{e}})$. Based on the flow velocity at $\boldsymbol{x}_{\mathrm{e}}$ we introduce a local orthogonal coordinate system, denoted by (ξ, η), and corresponding basis $\{\boldsymbol{e}_\xi, \boldsymbol{e}_\eta\}$ according to

$$\boldsymbol{e}_\xi = \cos(\alpha)\,\boldsymbol{e}_x + \sin(\alpha)\,\boldsymbol{e}_y, \quad \boldsymbol{e}_\eta = -\sin(\alpha)\,\boldsymbol{e}_x + \cos(\alpha)\,\boldsymbol{e}_y. \tag{10}$$

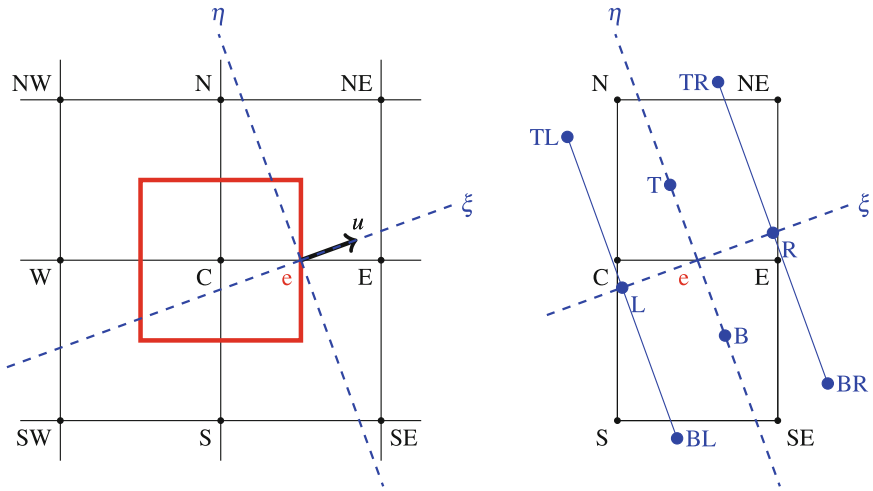


Fig. 1 Control volume and local coordinate system for the computation of $F_{x,\mathrm{e}}$ (*left*) and stencil (*right*). Function values at locations L, R, B, T, TL, BL, TR, BR are found by interpolation/extrapolation, see (18)

The transformation between the position vectors $\boldsymbol{x} = x\boldsymbol{e}_x + y\boldsymbol{e}_y$ in Cartesian coordinates and $\boldsymbol{\xi} = \xi\boldsymbol{e}_\xi + \eta\boldsymbol{e}_\eta$ in local coordinates is given by

$$\boldsymbol{x} - \boldsymbol{x}_\mathrm{e} = \boldsymbol{R}(\alpha)\boldsymbol{\xi}, \quad \boldsymbol{R}(\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}. \tag{11}$$

Note that in the (ξ, η)-coordinate system the interface velocity $\boldsymbol{u}(\boldsymbol{x}_\mathrm{e}) = U\boldsymbol{e}_\xi$ with $U = |\boldsymbol{u}(\boldsymbol{x}_\mathrm{e})| \geq 0$, elsewhere $\boldsymbol{u} = u_\xi\boldsymbol{e}_\xi + u_\eta\boldsymbol{e}_\eta$.

We introduce $\varphi^*(\xi, \eta) = \varphi(x, y)$. Reformulated in the local, orthogonal (ξ, η)-coordinate system, the conservation law (1) and the expression (2) for the flux read

$$\nabla \cdot \boldsymbol{f} = \frac{\partial f_\xi}{\partial \xi} + \frac{\partial f_\eta}{\partial \eta} = s, \tag{12a}$$

$$\boldsymbol{f} = f_\xi\boldsymbol{e}_\xi + f_\eta\boldsymbol{e}_\eta = \left(u_\xi\varphi^* - \varepsilon\frac{\partial\varphi^*}{\partial\xi}\right)\boldsymbol{e}_\xi + \left(u_\eta\varphi^* - \varepsilon\frac{\partial\varphi^*}{\partial\eta}\right)\boldsymbol{e}_\eta. \tag{12b}$$

The expression for the flux at the interface reduces to

$$\boldsymbol{f}(\boldsymbol{x}_\mathrm{e}) = \left(U\varphi^* - \varepsilon\frac{\partial\varphi^*}{\partial\xi}\right)\boldsymbol{e}_\xi - \varepsilon\frac{\partial\varphi^*}{\partial\eta}\boldsymbol{e}_\eta. \tag{13}$$

In the following we omit the asterisk (∗). Note that $f_x = \cos(\alpha)\, f_\xi - \sin(\alpha)\, f_\eta$, thus to compute $F_{x,\mathrm{e}}$ we need numerical approximations $F_{\xi,\mathrm{e}}$ and $F_{\eta,\mathrm{e}}$ of the flux components $f_\xi(\boldsymbol{x}_\mathrm{e})$ and $f_\eta(\boldsymbol{x}_\mathrm{e})$, respectively.

First, consider the computation of the component $F_{\xi,\mathrm{e}}$. Similar to the derivation of the one-dimensional flux, we determine $F_{\xi,\mathrm{e}}$ from the following local quasi-one-dimensional BVP:

$$\frac{\partial f_\xi}{\partial \xi} = \frac{\partial}{\partial \xi}\Big(U\varphi - \varepsilon \frac{\partial \varphi}{\partial \xi}\Big) = s_\xi, \quad -\tfrac{1}{2}h < \xi < \tfrac{1}{2}h, \eta = 0, \tag{14a}$$

$$\varphi(-\tfrac{1}{2}h, 0) = \varphi_\mathrm{L}, \quad \varphi(\tfrac{1}{2}h, 0) = \varphi_\mathrm{R}, \tag{14b}$$

where we choose $h = \min(\Delta x, \Delta y)$. Equation (14a) is a reformulation of the conservation law (12a), defined on the line segment connecting $\boldsymbol{x}_\mathrm{L} = \boldsymbol{x}_\mathrm{e} - \frac{1}{2}h\boldsymbol{e}_\xi$ and $\boldsymbol{x}_\mathrm{R} = \boldsymbol{x}_\mathrm{e} + \frac{1}{2}h\boldsymbol{e}_\xi$, with the flow velocity $\boldsymbol{u}$ replaced by $U\boldsymbol{e}_\xi$ and where the right hand side function s_ξ is a modified source term containing an approximation of the cross flux f_η. It is given by

$$s_\xi = s(\boldsymbol{x}(\xi, 0)) + \varepsilon \delta_{\eta\eta}\varphi(\boldsymbol{x}(\xi, 0)), \tag{15}$$

with $\delta_{\eta\eta}\varphi$ the standard central difference approximation of $\partial^2\varphi/\partial\eta^2$. Note that the boundary values $\varphi_\mathrm{L} = \varphi(\boldsymbol{x}_\mathrm{L})$ (left) and $\varphi_\mathrm{R} = \varphi(\boldsymbol{x}_\mathrm{R})$ (right) are not grid point values and need to be approximated by interpolation. We will specify this shortly. Analogous to (8) we find the following expressions for the flux:

$$F_{\xi,\mathrm{e}} = F^\mathrm{h}_{\xi,\mathrm{e}} + F^\mathrm{i}_{\xi,\mathrm{e}}, \tag{16a}$$

$$F^\mathrm{h}_{\xi,\mathrm{e}} = \frac{\varepsilon}{h}\big(B(-P)\varphi_\mathrm{L} - B(P)\varphi_\mathrm{R}\big), \tag{16b}$$

$$\begin{aligned} F^\mathrm{i}_{\xi,\mathrm{e}} &= h\big(C(-P)s_{\xi,\mathrm{L}} - C(P)s_{\xi,\mathrm{R}}\big) \\ &= h\big(C(-P)\big(s_\mathrm{L} + \varepsilon\delta_{\eta\eta}\varphi_\mathrm{L}\big) - C(P)\big(s_\mathrm{R} + \varepsilon\delta_{\eta\eta}\varphi_\mathrm{R}\big)\big), \end{aligned} \tag{16c}$$

with $P = Uh/\varepsilon > 0$ the (local) grid Péclet number. Analogous to the previous, the numerical flux $F_{\xi,\mathrm{e}}$ is the sum of the homogeneous flux $F^\mathrm{h}_{\xi,\mathrm{e}}$, corresponding to the advection-diffusion operator in ξ-direction, and the inhomogeneous flux $F^\mathrm{i}_{\xi,\mathrm{e}}$, depending on source and cross flux.

Next, for the (diffusive) component $F_{\eta,\mathrm{e}}$ we adopt the standard central difference scheme for the homogeneous part and discard the inhomogeneous part; cf. (9). Introducing the auxiliary points $\boldsymbol{x}_\mathrm{B} = \boldsymbol{x}_\mathrm{e} - \frac{1}{2}h\boldsymbol{e}_\eta$ (bottom) and $\boldsymbol{x}_\mathrm{T} = \boldsymbol{x}_\mathrm{e} + \frac{1}{2}h\boldsymbol{e}_\eta$ (top), it is given by

$$F_{\eta,\mathrm{e}} = F^\mathrm{h}_{\eta,\mathrm{e}} = \frac{\varepsilon}{h}\big(\varphi_\mathrm{B} - \varphi_\mathrm{T}\big), \tag{17}$$

where $\varphi_\mathrm{B} = \varphi(\boldsymbol{x}_\mathrm{B})$ and $\varphi_\mathrm{T} = \varphi(\boldsymbol{x}_\mathrm{T})$ need to be determined by interpolation.

To determine the auxiliary function values in (16) and (17) we need interpolation; see Fig. 1. Since $\boldsymbol{x}_\mathrm{L}, \boldsymbol{x}_\mathrm{R}, \boldsymbol{x}_\mathrm{B}$ and $\boldsymbol{x}_\mathrm{T}$ are all located in the rectangle $\mathscr{R}(\boldsymbol{x}_\mathrm{e}) = [x_\mathrm{C}, x_\mathrm{E}] \times$

$[y_S, y_N]$ with vertices NE, N, S and SE, centered around $\boldsymbol{x}_e$, we use linear interpolation in x-direction and quadratic interpolation in y-direction for $(x, y) \in \mathscr{R}(\boldsymbol{x}_e)$. Let p be the interpolation polynomial, then we have for example $\varphi_L = p(\boldsymbol{x}_L)$. Introducing the scaled coordinates σ_x $(0 \le \sigma_x \le 1)$ and σ_y $(-1 \le \sigma_y \le 1)$ according to $\sigma_x(x) = (x - x_C)/\Delta x$, $\sigma_y(y) = (y - y_C)/\Delta y$, the interpolation polynomial p can be written as

$$p(x, y) = \begin{bmatrix} 1 - \sigma_x & \sigma_x \end{bmatrix} \begin{bmatrix} \varphi_S & \varphi_C & \varphi_N \\ \varphi_{SE} & \varphi_E & \varphi_{NE} \end{bmatrix} \begin{bmatrix} -\frac{1}{2}\sigma_y\left(1 - \sigma_y\right) \\ \left(1 - \sigma_y^2\right) \\ \frac{1}{2}\sigma_y\left(1 + \sigma_y\right) \end{bmatrix} = \sum_{Q \in \mathscr{N}(\boldsymbol{x}_e)} a_Q(\boldsymbol{x})\varphi_Q, \tag{18}$$

where $\mathscr{N}(\boldsymbol{x}_e) = \{N, NE, C, E, S, SE\}$. Applying this interpolation formula to all φ-values in (16) and (17) and rearranging terms, we find the following expressions for the numerical flux:

$$F^h_{\xi,e} = \frac{\varepsilon}{h} \sum_{Q \in \mathscr{N}(\boldsymbol{x}_e)} \left(B(-P)a_Q(\boldsymbol{x}_L) - B(P)a_Q(\boldsymbol{x}_R)\right)\varphi_Q, \tag{19a}$$

$$F^i_{\xi,e} = h\varepsilon \sum_{Q \in \mathscr{N}(\boldsymbol{x}_e)} \left(C(-P)\delta_{\eta\eta}a_Q(\boldsymbol{x}_L) - C(P)\delta_{\eta\eta}a_Q(\boldsymbol{x}_R)\right)\varphi_Q + h\left(C(-P)s_L - C(P)s_R\right), \tag{19b}$$

$$F_{\eta,e} = \frac{\varepsilon}{h} \sum_{Q \in \mathscr{N}(\boldsymbol{x}_e)} \left(a_Q(\boldsymbol{x}_B) - a_Q(\boldsymbol{x}_T)\right)\varphi_Q. \tag{19c}$$

Note that the central difference approximations $\delta_{\eta\eta}a_Q(\boldsymbol{x}_L)$ and $\delta_{\eta\eta}a_Q(\boldsymbol{x}_R)$ in the expression for the inhomogeneous flux $F^i_{\xi,e}$ contain function values in the points $\boldsymbol{x}_L \pm h\boldsymbol{e}_\eta$ and $\boldsymbol{x}_R \pm h\boldsymbol{e}_\eta$, which are outside $\mathscr{R}(\boldsymbol{x}_e)$. For these values we still apply (18) (extrapolation). Finally, the numerical flux is then given by $F_{x,e} = \cos(\alpha)\, F_{\xi,e} - \sin(\alpha)\, F_{\eta,e}$, and depends on the six grid point values φ_Q with $Q \in \mathscr{N}(\boldsymbol{x}_e)$. A similar procedure can be applied to all other numerical fluxes, and substituting these in (4) gives rise to the 9-point stencil shown in Fig. 1.

4 Numerical Results

We have applied the complete flux scheme to equation (1) for two different flow velocities, viz. a constant velocity field $\boldsymbol{u}(x, y) = u\boldsymbol{e}_x + v\boldsymbol{e}_y$ not aligned with the grid and $\boldsymbol{u}(x, y) = 2y\left(1 - x^2\right)\boldsymbol{e}_x - 2x\left(1 - y^2\right)\boldsymbol{e}_y$, corresponding to rotating flow.

For the first case, we set $s(x, y) = (x - 1)(x + 1)y(y - 1)$, take several values of the angle α and vary ε from 1 to 10^{-8}. To determine the order of convergence, we choose $\Delta x = \Delta y = h$ and apply Richardson extrapolation to the numerical solution at location $(\frac{1}{2}, \frac{1}{2})$. We always obtain order 2 in the limit $h \to 0$, uniformly in ε, see the results in Table 1.

Table 1 Values for $r_h := (\varphi_{h/2} - \varphi_h)/(\varphi_{h/4} - \varphi_{h/2}) \approx 2^p$ with φ_h the numerical approximation of $\varphi(\frac{1}{2}, \frac{1}{2})$ computed with grid size h and p the convergence order. *Left* $u = -1$, $v = \sqrt{2}/2$. *Right* $u = 1$, $v = 1$

h^{-1}	$\varepsilon = 1$	$\varepsilon = 10^{-2}$	$\varepsilon = 10^{-8}$		h^{-1}	$\varepsilon = 1$	$\varepsilon = 10^{-2}$	$\varepsilon = 10^{-8}$
40	3.7156	1.0432	0.0705		40	4.0706	1.3677	3.5803
80	3.8493	2.9941	2.2333		80	4.0351	2.2128	3.7470
160	3.9230	3.7650	3.0759		160	4.0175	2.9808	3.9053
320	3.9611	3.9790	3.5871		320	4.0087	3.7679	3.9603
640	3.9805	4.0127	3.8014		640	4.0044	3.7679	3.9821

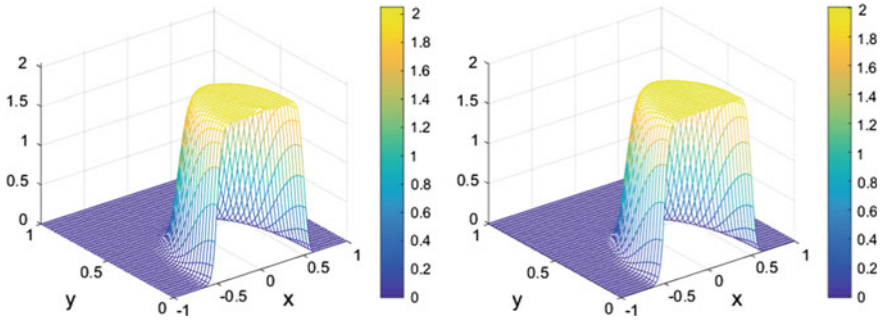


Fig. 2 Numerical solution of the rotating flow problem, computed with the CF (*left*) and HF scheme (*right*). The grid size $\Delta x = \Delta y = 2.5 \times 10^{-2}$

The second flow velocity comes from a benchmark problem by Smith and Hutton [3]; see also Example 3 in [4]. In this problem an inlet profile with steep layer defined for $-1 \le x \le 0$ and $y = 0$ is convected around a 180° bend to the outlet ($0 < x \le 1$ and $y = 0$). There is no source. We have computed numerical solutions for $\varepsilon = 10^{-8}$ on a coarse 80×40 grid using the CF-scheme as well as the homogeneous flux (HF) scheme, ignoring the inhomogeneous flux (16c); see Fig. 2. Both schemes produce a sharp resolution of the interior layer. Comparing these solutions to the standard HF solution presented in [4], which flattens out the profile, we conclude that the schemes in local (ξ, η)-coordinates do not suffer from numerical diffusion.

5 Concluding Remarks

We have presented a version of the two-dimensional complete flux scheme in terms of local, orthogonal flow adapted coordinates. The scheme involves a flux component F_ξ parallel to the local velocity field and a component F_η perpendicular to it. For F_ξ we employ the one-dimensional complete flux approximation, including the cross flux, whereas for F_η the homogeneous flux scheme suffices. The resulting

finite volume scheme exhibits uniform second order convergence and does not suffer from numerical diffusion. This approach is readily generalized to three-dimensional problems, which will be presented in future work.

References

1. Liu, L., van Dijk, J., ten Thije Boonkkamp, J., Mihailova, D., van de Mullen, J.: The complete flux scheme—error analysis and application to plasma simulation. J. Comput. Appl. Math. **250**, 229–243 (2013)
2. Raithby, G.: Skew upstream differencing schemes for problems involving fluid flow. Comput. Methods Appl. Mech. Eng. **9**, 153–164 (1976)
3. Smith, R., Hutton, A.: The numerical treatment of advection: a performance comparison of current methods. Numer. Heat Trans. **5**, 439–461 (1982)
4. ten Thije Boonkkamp, J., Anthonissen, M.: The finite volume-complete flux scheme for advection-diffusion-reaction equations. J. Sci. Comput. **46**, 47–70 (2011)

hp–Adaptive Discontinuous Galerkin Methods for Porous Media Flow

Birane Kane, Robert Klöfkorn and Christoph Gersbacher

Abstract We present an adaptive Discontinuous Galerkin discretization for the solution of porous media flow problems. The considered flows are immiscible and incompressible. The adaptive approach implemented allows for refinement/coarsening in both the element size and the polynomial degree. The method is evaluated using homogeneous and heterogeneous test cases.

Keywords *hp*-Adaptivity · Fully implicit · Discontinuous galerkin · DUNE

MSC (2010): 65M08 · 65M60 · 65M50

1 Introduction

The modeling and simulation of flow in porous media is essential in many environmental problems such as groundwater flow and petroleum engineering. The inherent geological complexity and the strong heterogeneity of the soil parameters require locally conservative methods such as Discontinuous Galerkin (DG) methods in order to be able to follow small concentrations [1].

The first *h*-adaptive DG framework for porous media two-phase flow was introduced by Klieber and Riviére [11]. The authors used a decoupled formulation with continuous capillary pressure functions, only $2d$ flow on non-conforming simplicial grids were considered and they implemented an error indicator obtained from

B. Kane
University of Stuttgart, Stuttgart, Germany
e-mail: birane.kane@ians.uni-stuttgart.de

R. Klöfkorn (✉)
International Research Institute of Stavanger, Stavanger, Norway
e-mail: robert.kloefkorn@iris.no

C. Gersbacher
University of Freiburg, Freiburg, Germany
e-mail: gersbach@mathematik.uni-freiburg.de

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_47

transient linear convection diffusion problems [9]. More recently, Kane [10] implemented a higher order h-adaptive scheme for $2d$ and $3d$ two-phase flow problem with strong heterogeneity, discontinuous capillary pressure functions and gravity effects. The results in [10] show that an increase of the polynomial degree gives a considerable improvement of the solution with sharper fronts and the oscillations appearing in the vicinity of the front are reduced with the local mesh refinement. A discretization scheme independent abstract framework allowing for a more rigorous a-posteriori estimator for porous media two-phase flow problem was introduced by [13]. This paved the way for a h-adaptive strategy for a homogeneous two-phase flow problem. However it has only been applied so far to Finite Volume methods [6].

The first contribution of this work is to provide a first and second order Adam-Moulton time discretization combined with the Interior Penalty DG methods. This implicit space time discretization leads to a fully coupled nonlinear system requiring to build a Jacobian matrix at each time step for the Newton-Raphson method. The second contribution of this work is providing a hp-adaptive strategy extending the previous work of [10], this is the first porous media two-phase flow fully adaptive scheme allowing for adaptivity for both the element size and the polynomial degree. This hp-adaptive strategy allows to refine the mesh when the solution is estimated to be rough and increase the polynomial degree when the solution is estimated to be smooth hence compensating the increased computational cost for complex models.

The rest of this document is organised as follows. In the next section, we describe the two-phase flow model. The DG discretization is introduced in Sect. 3. The adaptive strategy in space is outlined in Sect. 4. Numerical examples are provided in Sect. 5. Finally concluding remarks are provided in the last section.

2 Governing Equations

We consider an open and bounded domain $\Omega \in \mathbb{R}^d, d \in \{1, 2, 3\}$ and the time interval $\mathcal{J} = (0, T)$, $T > 0$. The flow of the wetting-phase and the nonwetting-phase is described by the Darcy's law and the continuity equation for each phase, namely, with $\sum_\alpha s_\alpha = 1$ and $p_n - p_w = p_c(s_{w,e})$,

$$v_\alpha = -\lambda_\alpha K(\nabla p_\alpha - \rho_\alpha g) \text{ and } \quad \phi \frac{\partial \rho_\alpha s_\alpha}{\partial t} + \nabla \cdot (\rho_\alpha v_\alpha) = \rho_\alpha q_\alpha. \tag{1}$$

Here, we search for the phase pressures p_α and the phase saturations s_α, $\alpha \in \{w, n\}$. We denote with subscript w the wetting-phase and with subscript n the nonwetting-phase. K is the permeability of the porous medium, ρ_α is the phase density, q_α is a source/sink term and g is the constant gravitational vector. We assume the porosity ϕ is time independent and there exist $\phi_1, \phi_2 > 0$ such that $0 < \phi_1 \leq \phi \leq \phi_2$ and with the phase mobilities $\lambda_\alpha = \frac{k_{r\alpha}}{\mu_\alpha}$, $\alpha \in \{w, n\}$, where μ_α is the phase viscosity and

$k_{r\alpha}$ is the relative permeability of phase α. The relative permeabilities are functions that depend on the phase saturation in nonlinear fashion (i.e. $k_{r\alpha} = k_{r\alpha}(s_\alpha)$). For example, in the Brooks-Corey model [5], $k_{rw}(s_{w,e}) = s_{w,e}^{\frac{2+3\theta}{\theta}}$, $k_{rn}(s_{n,e}) = (s_{n,e})^2(1 - (1 - s_{n,e})^{\frac{2+\theta}{\theta}})$, where the effective saturation $s_{\alpha,e}$ is $s_{\alpha,e} = \frac{s_\alpha - s_{\alpha,r}}{1 - s_{w,r} - s_{n,r}}$, $\forall \alpha \in \{w, n\}$. Here, $s_{\alpha,r}, \alpha \in \{w, n\}$ are the phase residual saturations. The parameter $\theta \in [0.2, 3.0]$ is a result of the inhomogeneity of the medium. The capillary pressure $p_c = p_c(s_{w,e})$ is a function of the phase saturation $p_c(s_{w,e}) = p_d s_{w,e}^{-1/\theta}$ where $p_d \geq 0$ is the constant entry pressure.

From the constitutive relations $s_w + s_n = 1$ and $p_n - p_w = p_c(s_{w,e})$, we can rewrite the two-phase flow problem as a system of equations with two unknowns p_w and s_n,

$$\begin{aligned} -\nabla \cdot (\lambda_t K \nabla p_w + \lambda_n K \nabla p_c - (\rho_w \lambda_w + \rho_n \lambda_n) K g) &= q_w + q_n, \\ \phi \frac{\partial s_n}{\partial t} - \nabla \cdot (\lambda_n K (\nabla p_w - \rho_n g)) - \nabla \cdot (\lambda_n K \nabla p_c) &= q_n. \end{aligned} \tag{2}$$

Here, $\lambda_t = \lambda_w + \lambda_n$ denotes the total mobility.

The first equation of (2) is of elliptic type with respect to the pressure p_w. The type of the second equation of (2) is either nonlinear hyperbolic if $\frac{\partial p_c(s_n)}{\partial s_n} \equiv 0$ or degenerate parabolic if the capillary pressure is not neglected. The diffusion term might degenerate if $\lambda_n(s_n = 0) = 0$. In order to have a complete system we add appropriate boundary and initial conditions. Thus, we assume that the boundary of the system is divided into disjoint open sets $\partial\Omega = \bar{\Gamma}_D \cup \bar{\Gamma}_N$. We define the total inflow $J_t = J_w + J_n$ as the sum of the phases inflow on the Neumann boundary $\bar{\Gamma}_N$.

3 Discretization

Let $\mathcal{T}_h = \{E\}$ be a family of non-degenerate, quasi-uniform, possibly non-conforming partitions of Ω consisting of N_h elements (quadrilaterals or triangles in 2d, tetrahedrons or hexahedrons in 3d) of maximum diameter h. Let Γ^h be the union of the open sets that coincide with internal interfaces of elements of $\mathcal{T}_h$. Dirichlet and Neumann boundary interfaces are collected in the set Γ_D^h and Γ_N^h. Let e denote an interface in Γ^h shared by two elements E_- and E_+ of $\mathcal{T}_h$; we associate with e a unit normal vector n_e directed from E_- to E_+. We also denote by $|e|$ the measure of e. The discontinuous finite element space is $\mathscr{D}_r(\mathcal{T}_h) = \{v \in \mathbb{L}^2(\Omega) : v_{|E} \in \mathscr{P}_r(E) \ \forall E \in \mathcal{T}_h\}$, where $\mathscr{P}_r(E)$ denotes $\mathbb{Q}_r$ (resp. $\mathbb{P}_r$) the space of polynomial functions of degree at most $r \geq 1$ on E (resp. the space of polynomial functions of total degree $r \geq 1$ on E). We approximate the pressure and the saturation by discontinuous polynomials of total degrees r_p and r_s respectively.

For any function $q \in \mathscr{D}_r(\mathscr{T}_h)$, we define the jump operator $[\![\cdot]\!]$ and the average operator $\{\cdot\}$ over the interface e: $\forall e \in \Gamma^h, \quad [\![q]\!] := q_{E_-} - q_{E_+}, \quad \{q\} := \frac{1}{2}q_{E_-} + \frac{1}{2}q_{E_+}$, and $\forall e \in \partial\Omega, \quad [\![q]\!] = \{q\} := q_{E_-}$.
In order to treat the strong heterogeneity of the permeability tensor, we follow [8] and introduce a weighted average operator $\{\cdot\}_\omega$:

$$\forall e \in \Gamma^h, \quad \{q\}_\omega = \omega_{E_-} q_{E_-} + \omega_{E_+} q_{E_+} \text{ and } \forall e \in \partial\Omega, \quad \{q\}_\omega = q_{E_-}.$$

The weights are $\omega_{E_-} = \frac{\delta_K^{E_+}}{\delta_K^{E_+} + \delta_K^{E_-}}, \quad \omega_{E_+} = \frac{\delta_K^{E_-}}{\delta_K^{E_+} + \delta_K^{E_-}}$ with $\delta_K^{E_-} = n_e^T K_{E_-} n_e$ and $\delta_K^{E_+} = n_e^T K_{E_+} n_e$. Here, K_{E_-} and K_{E_+} are the permeability tensors for the elements E_- and E_+.

3.1 Semi Discretization in Space

The derivation of the semi-discrete DG formulation is standard (see [11]). First, we multiply each equation of (2) by a test function and integrate over each element, then we apply Green formula to obtain the semi-discrete weak DG formulation. Hence, the aforementioned formulation consists in finding the continuous in time approximations $p_{w,h}(\cdot, t) \in \mathscr{D}_{r_p}(\mathscr{T}_h)$, $s_{n,h}(\cdot, t) \in \mathscr{D}_{r_s}(\mathscr{T}_h)$ such that:

$$\begin{aligned} \mathscr{B}_h(p_{w,h}, \varphi; s_{n,h}) &= l_h(\varphi) \qquad \forall \varphi \in \mathscr{D}_{r_p}(\mathscr{T}_h), \forall t \in \mathscr{J}, \\ (\Phi \partial_t s_{n,h}, \psi) + c_h(p_{w,h}, \psi; s_{n,h}) + d_h(s_{n,h}, \psi) &= r_h(\psi) \qquad \forall \psi \in \mathscr{D}_{r_s}(\mathscr{T}_h), \forall t \in \mathscr{J}. \end{aligned} \tag{3}$$

The bilinear form $\mathscr{B}_h$ in the total fluid conservation equation of the system (3) is:

$$\mathscr{B}_h(p_{w,h}, \varphi; s_{n,h}) = \mathscr{B}_{bulk,h} + \mathscr{B}_{cons,h} + \mathscr{B}_{sym,h} + \mathscr{B}_{stab,h}. \tag{4}$$

The first term $\mathscr{B}_{bulk,h} := \mathscr{B}_{bulk,h}(p_{w,h}, \varphi; s_{n,h})$ of (4) is the volume term:

$$\mathscr{B}_{bulk,h} = \sum_{E \in \mathscr{T}_h} \int_E (\lambda_t K \nabla p_{w,h} + \lambda_n K \nabla p_{c,h}) \cdot \nabla \varphi - \sum_{E \in \mathscr{T}_h} (\rho_n \lambda_n + \rho_w \lambda_w) K g \cdot \nabla \varphi. \tag{5}$$

The second term $\mathscr{B}_{cons,h} := \mathscr{B}_{cons,h}(p_{w,h}, \varphi; s_{n,h})$, is the consistency term:

$$\begin{aligned} \mathscr{B}_{cons,h} = &- \sum_{e \in \Gamma^h \cup \Gamma_D^h} \int_e \{\lambda_t K \nabla p_{w,h}\}_\omega \cdot n_e [\![\varphi]\!] - \sum_{e \in \Gamma^h \cup \Gamma_D^h} \int_e \{\lambda_n K \nabla p_{c,h}\}_\omega \cdot n_e [\![\varphi]\!] \\ &+ \sum_{e \in \Gamma^h \cup \Gamma_D^h} \int_e \{(\rho_n \lambda_n + \rho_w \lambda_w) K g)\}_\omega \cdot n_e [\![\varphi]\!]. \end{aligned} \tag{6}$$

The term $\mathscr{B}_{sym,h} := \mathscr{B}_{sym,h}(p_{w,h}, \varphi; s_{n,h})$, is the symmetry term. Depending on the choice of ε we get different DG methods ($\varepsilon = -1$ SIPG, $\varepsilon = 1$ NIPG, $\varepsilon = 0$ IIPG):

$$\mathscr{B}_{sym,h} = \varepsilon \sum_{e\in\Gamma^h\cup\Gamma_D^h} \int_e \{\lambda_t K\nabla\varphi\}_\omega [\![p_{w,h}]\!] + \varepsilon \sum_{e\in\Gamma^h\cup\Gamma_D^h} \int_e \{\lambda_n K\nabla\varphi\}_\omega [\![s_{n,h}]\!]. \tag{7}$$

$\mathscr{B}_{stab,h} := \mathscr{B}_{stab,h}(p_{w,h}, \varphi) = \sum_{e\in\Gamma^h\cup\Gamma_D^h} \gamma_e^p \int_e [\![p_{w,h}]\!][\![\varphi]\!]$ is the stability term.

Following [3], the penalty formulation is: $\gamma_e^p = \sigma_p \frac{r_p(r_p+d-1)|e|}{min(|E_-|,|E_+|)}$, $\sigma_p \geq 0$.

The right hand side of the total fluid conservation equation of the system (3) is a linear form including the boundary conditions and the source terms.

$$\begin{aligned} l_h(\varphi) = &\int_\Omega (q_w + q_n)\varphi - \sum_{e\in\Gamma_N} \int_e J_t\varphi + \varepsilon \sum_{e\in\Gamma_D^h} \int_e \lambda_t K\nabla\varphi \cdot n_e p_D \\ &+ \varepsilon \sum_{e\in\Gamma_D^h} \int_e \lambda_n K\nabla\varphi \cdot n_e s_D + \sum_{e\in\Gamma_D^h} \gamma_e^p \int_e p_D\varphi, \qquad \forall\varphi \in \mathscr{D}_{r_p}(\mathscr{T}_h). \end{aligned} \tag{8}$$

The second equation of the system (3) is the discrete weak formulation of the nonwetting-phase conservation equation where the convection term $-\nabla\cdot(\lambda_n K(\nabla p_w - \rho_n g))$ might be approximated by an upwind discretization technique.

$$\begin{aligned} c_h(p_{w,h}, \psi; s_{n,h}) = &\sum_{E\in\mathscr{T}_h} \int_E (K\lambda_n(\nabla p_{w,h} - \rho_n g))\cdot\nabla\psi - \sum_{e\in\Gamma^h\cup\Gamma_D^h} \int_e \{K\lambda_n^\# \nabla p_{w,h}\}_\omega [\![\psi]\!] \\ &+ \sum_{e\in\Gamma^h\cup\Gamma_D^h} \int_e \{\rho_n K\lambda_n^\# g\}_\omega [\![\psi]\!] + \varepsilon \sum_{e\in\Gamma^h\cup\Gamma_D^h} \int_e \{K\lambda_n^\#\nabla\psi\}_\omega [\![p_{w,h}]\!], \end{aligned} \tag{9}$$

where $\lambda_n^\# = (1-\rho)\lambda_{n,E} + \rho\lambda_n^\uparrow$ and $\lambda_n^\uparrow$ is the upwind mobility:

$$\forall e \in \partial E_- \cap \partial E_+, \quad \lambda_n^\uparrow = \begin{cases} \lambda_{n,E_-} & \text{if } -K(\nabla p_w + \nabla p_c - \rho_n g)\cdot n \geq 0, \\ \lambda_{n,E_+} & \text{else.} \end{cases}$$

Hence depending on the value of $\rho \in \{0, 1\}$, we might use central differencing or upwinding of the mobility for internal interfaces.
The diffusion term $-\nabla\cdot(\lambda_n K\nabla p_c)$ is discretized by a bilinear form similar to that of (4). A more detailed expression can be found in [10].

3.2 *Fully Coupled/Fully Implicit DG Scheme*

The time interval $[0, T]$ is divided into N intervals $\Delta t_i = t_{i+1} - t_i$ as $0 = t_0 \leq t_1 \leq \cdots \leq t_{N-1} \leq t_N = T$. Let p_w^i and s_n^i be the numerical solutions at time t^i. The approximation $s_{n,h}^0$ is chosen as the L^2 projection of the saturation $s_n(0)$. For the

sake of simplicity and easier reading, we apply a first order Adams-Moulton time discretization and Interior Penalty DG for space discretization to the semi-discrete system (3):

$$\begin{aligned} \mathscr{B}_h(p_{w,h}^{i+1}, \varphi; s_{n,h}^{i+1}) &= l_h(\varphi), \qquad \forall \varphi \in \mathscr{D}_{r_p}(\mathscr{T}_h), \\ (\Phi \frac{s_{n,h}^{i+1} - s_{n,h}^{i}}{\Delta t}, \psi) + c_h(p_{w,h}^{i+1}, \psi; s_{n,h}^{i+1}) + d_h(s_{n,h}^{i+1}, \psi) &= r_h(\psi), \qquad \forall \psi \in \mathscr{D}_{r_s}(\mathscr{T}_h), \\ (s_{n,h}^{0}, \zeta) &= (s_n^0, \zeta), \qquad \forall \zeta \in \mathscr{D}_{r_s}(\mathscr{T}_h). \end{aligned} \tag{10}$$

4 Adaptivity

The first approach considered, *GradIndicator*, is based on a heuristic indicator which depends on the local gradient of the DG solution measured in the L^2 norm. We define on each element E of the mesh, the indicator η_E^i at time step i, such that: $\eta_E^i = \|\nabla s_n^i\|_{L^2(E)}$, $\forall E \in \mathscr{T}_h$. Each element whose indicator η_E^i is greater than a threshold value $\eta_{Tol} \geq 0$ is refined.

For the second approach, the choice between h-adaptivity and p-adaptivity depends heavily on the value of a smoothness indicator ς_E. Given an error indicator η_E, $E \in \mathscr{T}_h$, we define η_E^{r-1} the L^2 projection into a lower order polynomial space $\mathscr{D}_{r-1}(\mathscr{T}_h)$. The derivation of this L^2 projection is quite straightforward due to the hierarchical aspect of the modal DG bases implemented. The indicator $\eta_E = \|s_n\|_{H^1(E)}$ allows to refine the mesh when the solution is estimated to be rough and increase the polynomial degree when the solution is estimated to be smooth. The use of heuristic error indicators requires a maximum level of allowed h-refinement $maxlevel$ to be specified to avoid overly aggressive refinement. Whenever an element is selected for h-refinement it is also selected for p-coarsening in order to reduce the oscillations in the vicinity of the front of the propagation. An hp-adaptive strategy of this type called *PRIOR2P* as in [12] is implemented. In that approach, the smoothness indicator is $\varsigma_E \sim 1 - \frac{log((\eta_E^{r-1})/(\eta_E^{r-2}))}{log((r-1)/(r-2))}$, where r is the local polynomial degree.

5 Numerical Simulations

In this section we present some numerical tests for the adaptive DG scheme. All test cases are implemented with the Interior Penalty methods. In order to ensure second order accuracy, we employ a central differencing of the mobility for internal interfaces thus following a similar approach to that of Rivière et al. [7].

5.1 2d Flow Problem

We consider here a two dimensional test case that admits an exact solution from [2] aiming to examine the L^2 error of the DG methods. The problem depicts the transport of a Gaussian pulse in a rotating flow field. Considering $\Omega = (-0.5, 0.5)^2$ and $J = (0, T)$, we search for (S) such that: $\frac{\partial S}{\partial t} + \nabla \cdot (uS + K\nabla S) = 0$ in $\Omega \times J$.

The problem boundary and initials conditions derive from the exact solution $S(x, y, t) = 2\sigma^2/(2\sigma^2 + 4Kt) \quad e^{(-\frac{(\bar{x}-x_c)^2+(\bar{y}-y_c)^2}{2\sigma^2+4Kt})}$ where, $u = (-4y, 4x)^T$, $\bar{x} = xcos(4t) + ysin(4t)$, $\bar{y} = -xsin(4t) + ycos(4t)$, $x_c = -0.25$, $y_c = 0$, $K = 10^{-4}$ and $2\sigma^2 = 0.004$.

The domain is subdivided uniformly into square elements. The coarsest mesh consist of 8×8 elements. The solutions are approximated by piecewise polynomials of order k, $k \in \{1, 2, 3, 4\}$. The penalty parameter $\sigma_p = 10^{-10}$. Figures 1, 2 and 3 provide contours of the solution for the IIPG scheme combined with second order Adams-Moulton method time discretization. The numerical analysis in Table 1 shows that the PRIOR2P indicator yields a smaller L^2 error.

5.2 3d Heterogeneous Problem

In this section, we focus on a three-dimensional case. We also consider different sand types with different permeabilities and different entry pressures (Table 2).

The bottom of the reservoir is impermeable for both phases. Hydrostatic conditions for the pressure p_w and homogeneous Dirichlet conditions for the saturation s_n are prescribed at the lateral boundaries. A flux of 0.25 Kg $s^{-1}m^{-2}$ of

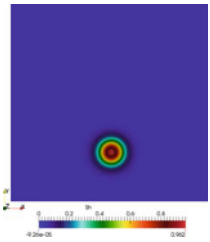


Fig. 1 Rotating pulse, solution at T = 0.4

Table 1 L^2 error for solution at $T = 0.4$

	$\|S - S_h\|_{L^2(\Omega)}$	Final nb of DOFs	Avg nb of lin it/ Newton cycle	Avg inv time/ Newton [s]	Avg assem time / Newton [s]
GradIndicator	3.06×10^{-04}	19230	28.63	0.0525	0.44
PRIOR2P	3.494×10^{-06}	46750	33.36	0.096	0.96

the DNAPL is infiltrated from the top into a domain of depth of 1 m. The initial `ALUCubeGrid` mesh consist of $17 \times 17 \times 17$ hexahedral elements and resolves the interfaces between regions with different permeabilities. 150 time steps of length $\Delta t = 20$ s are computed (final time $T = 3000$ s). This grid is locally adapted (non-conforming). We also set $r_p = r_s$ for the problem.

Figure 4 illustrates the evolution of the nonwetting saturation during the simulation. The effects of the *hp* algorithm are reflected in the mesh distribution showing an intense refinement and lower polynomial degree in the parts of the domain where the value of the indicator is above the threshold value. The second row of Fig. 4 shows the drastic improvement of the front shape and the reduction of the oscillations in the vicinity of the front when h and hp-adaptive methods are used. Table 3 provides more details concerning the computational effort.

6 Conclusion & Outlook

In this work, we have introduced an adaptive discontinuous Galerkin scheme for incompressible, immiscible two-phase flow in strongly heterogeneous porous media with gravity forces and discontinuous capillary pressures. We considered as a $3d$ test case a DNAPL infiltration in an initially water saturated reservoir. The oscillations appearing in the vicinity of the front of the propagation are reduced with the local

Table 2 3d problem parameters

	Ω_1	Ω_2	$\Omega \setminus \Omega_1 \cap \Omega \setminus \Omega_2$
Φ [-]	0.39	0.39	0.40
k [m^2]	6.64×10^{-16}	6.64×10^{-15}	6.64×10^{-11}
S_{wr} [-]	0.1	0.1	0.12
S_{nr} [-]	0.00	0.00	0.00
θ [-]	2.0	2.0	2.70
p_d [Pa]	5000	5000	755

Table 3 L^2 error for solution at $T = 0.4$

	Final nb of DOFs	Avg nb of lin it/ Newton cycle	Avg inv time/ Newton [s]	Avg assem time / Newton [s]	Total CPU time [s]
No-adapt deg=2	196520	96.19	2.97	33.9	7360.3
h-adapt deg=1	171232	127.157	20.38	9.86	6740.4
h-adapt deg=2	398680	509.56	78.0	16.26	19812.1
hp-adapt deg2	296680	468	70.69	28.3	19294.8

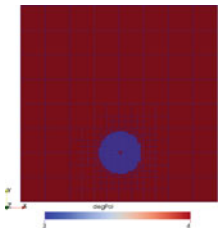


Fig. 2 Polynomial degrees at T = 0.4 for GradIndicator

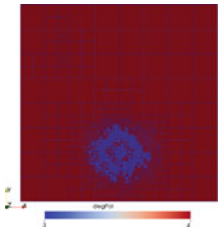


Fig. 3 Polynomial degrees at T = 0.4 for PRIOR2P

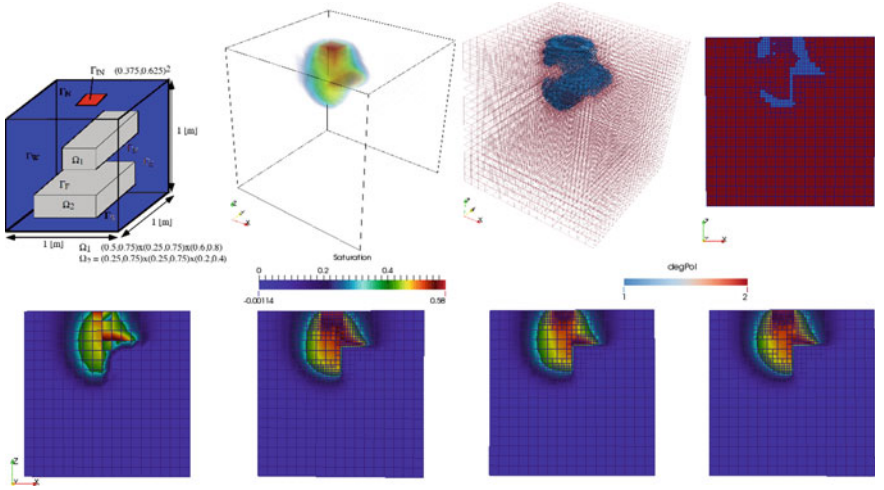


Fig. 4 *First row* from *left* to *right*, domain geometry, contour plot of saturation distribution after 3000 s of injection, mesh distribution, polynomial degree distribution along the slice y = 0.45. *Second row* saturation profile along the slice y = 0.45; from *left* to *right*, non-adaptive with $r_p = r_s = 2$, h-adaptive with $r_p = r_s = 1$, h-adaptive with $r_p = r_s = 2$, hp-adaptive with $\max\{r_p, r_s\}$=2

mesh refinement and the decrease of the local polynomial order. Future work will be concerned with the derivation of robust anisotropic *hp*-adaptive methods and the extension to other DG methods such as the Compact Discontinuous Galerkin 2 (CDG2) [4].

Acknowledgements Birane Kane acknowledges the Cluster of Excellence in Simulation Technology (SimTech) at the University of Stuttgart for financial support. Robert Klöfkorn acknowledges the Research Council of Norway and the industry partners – ConocoPhillips Skandinavia AS, BP Norge AS, Det Norske Oljeselskap AS, Eni Norge AS, Maersk Oil Norway AS, DONG Energy A/S, Denmark, Statoil Petroleum AS, ENGIE E&P NORGE AS, Lundin Norway AS, Halliburton AS, Schlumberger Norge AS, Wintershall Norge AS – of The National IOR Centre of Norway for financial support.
Both authors would like to thank the reviewers for helpful comments to improve this work.

References

1. Bastian, P.: Numerical computation of multiphase flow in porous media. Ph.D. thesis, habilitationsschrift Univeristät Kiel (1999)
2. Bastian, P.: Higher order discontinuous galerkin methods for flow and transport in porous media. In: Challenges in Scientific Computing-CISC 2002, pp. 1–22. Springer (2003)
3. Bastian, P.: A fully-coupled discontinuous galerkin method for two-phase flow in porous media with discontinuous capillary pressure. Comput. Geosci. **18**(5), 779–796 (2014)
4. Brdar, S., Dedner, A., Klöfkorn, R.: Compact and stable discontinuous galerkin methods for convection-diffusion problems. SIAM J. Sci. Comput. **34**(1), 263–282 (2012)
5. Brooks, R.H., Corey, A.T.: Hydraulic properties of porous media and their relation to drainage design. Trans. ASAE **7**(1), 26–0028 (1964)
6. Cancès, C., Pop, I., Vohralík, M.: An a posteriori error estimate for vertex-centered finite volume discretizations of immiscible incompressible two-phase flow. Math. Comput. **83**(285), 153–188 (2014)
7. Epshteyn, Y., Rivière, B.: Fully implicit discontinuous finite element methods for two-phase flow. Appl. Numer. Math. **57**(4), 383–401 (2007)
8. Ern, A., Mozolevski, I., Schuh, L.: Discontinuous galerkin approximation of two-phase flows in heterogeneous porous media with discontinuous capillary pressures. Comput. Methods Appl. Mech. Eng. **199**(23), 1491–1501 (2010)
9. Ern, A., Proft, J.: A posteriori discontinuous galerkin error estimates for transient convection-diffusion equations. Appl. Math. Lett. **18**(7), 833–841 (2005)
10. Kane, B.: Using dune-fem for adaptive higher order discontinuous galerkin methods for two-phase flow in porous media. Arch. Numer. Softw. (submitted)
11. Klieber, W., Riviere, B.: Adaptive simulations of two-phase flow by discontinuous galerkin methods. Comput. Methods Appl. Mech. Eng. **196**(1), 404–419 (2006)
12. Mitchell, W.F., McClain, M.A.: A comparison of hp-adaptive strategies for elliptic partial differential equations. ACM Trans. Math. Softw. (TOMS) **41**(1), 2 (2014)
13. Vohralík, M., Wheeler, M.F.: A posteriori error estimates, stopping criteria, and adaptivity for two-phase flows. Comput. Geosci. **17**(5), 789–812 (2013)

A Nonlinear Flux Approximation Scheme for the Viscous Burgers Equation

N. Kumar, J.H.M. ten Thije Boonkkamp, B. Koren and A. Linke

Abstract We present a nonlinear flux approximation scheme for the spatial discretization of the viscous Burgers equation. We derive the numerical flux function from a local two-point boundary value problem (BVP), which results in a nonlinear equation that depends on the local boundary values and the diffusion constant. The flux scheme is consistent and stable (does not introduce any spurious oscillations), as demonstrated by the numerical results.

Keywords Numerical flux · Nonlinear local BVP · Viscous burgers equation

MSC (2010): 65M08 · 34B15

1 Introduction

In this contribution we present a nonlinear flux approximation scheme for the spatial discretization of the viscous Burgers equation. The Burgers equation is an ideal test problem, as its spatial discretization can be carried over to the convective and viscous fluxes involved in the Navier–Stokes equations. The expression for the flux is derived from a local two-point BVP and is inspired by [5], where a local BVP is solved to

N. Kumar (✉) · J.H.M. ten Thije Boonkkamp · B. Koren
Department of Mathematics and Computer Science, Eindhoven University of Technology, PO Box 513, 5600MB Eindhoven, The Netherlands
e-mail: n.kumar@tue.nl

J.H.M. ten Thije Boonkkamp
e-mail: j.h.m.tenthijeboonkkamp@tue.nl

B. Koren
e-mail: b.koren@tue.nl

A. Linke
Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstr. 39, 10117 Berlin, Germany
e-mail: alexander.linke@wias-berlin.de

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_48

derive an integral representation of the flux for the convection-diffusion-reaction equation. The resulting numerical flux is expressed as a sum of a *homogeneous* part, which depends on the Péclet number (local balance of convection and diffusion) and an *inhomogeneous* part depending on the effects of the source term (associated with the reaction). Note that the homogeneous flux approximation is similar to the approximation methods described in [1, 3]. In this contribution, we extend the homogeneous approximation to nonlinear problems.

In the vanishing viscosity limit, the viscous Burgers equation is a singularly perturbed problem. Moreover, the nonlinearity of the flux does not allow us to express the homogeneous flux as linear combination of the convective and the viscous part, which makes it cumbersome to have a consistent numerical flux. In this paper, we extend the local BVP method to nonlinear problems, such that the resulting numerical flux is consistent, i.e., reduces to the correct flux in the limit case. A discussion on nonlinear local two-point BVPs can be found in [2], where the authors show (i) the solvability of some auxiliary local nonlinear two-point BVPs, and (ii) the convergence of the discrete scheme to a weak solution of the continuous problem.

The paper is organized as follows: in Sect. 2 we formulate the local BVP for the flux approximation. Sect. 3 gives details of the derivation for the numerical fluxes. In Sect. 4 we compare the nonlinear scheme with the linearized homogeneous flux scheme described in [5] as well as with other standard methods. Sect. 5 gives the concluding remarks.

2 Flux from Local Two-Point BVP

Consider the one-dimensional viscous Burgers equation

$$u_t + f(u, u_x)_x = 0, \qquad f(u, u_x) := \tfrac{1}{2}u^2 - \nu u_x, \tag{1}$$

defined on $\Omega(\subset \mathbb{R}) \times (0, T)$, where $\nu(\geq 0)$ is the diffusion coefficient. The spatial discretization of the Burgers equation using a finite-volume method requires the approximation of the flux function $f(u, u_x)$ at each interface between two control volumes. The semi-discrete formulation of Eq. (1) is given by

$$\Delta x\, \dot{u}_j + F_{j+1/2} - F_{j-1/2} = 0, \qquad \dot{u} := u_t, \tag{2}$$

where $F_{j+1/2} \approx f(u, u_x)|_{x=x_{j+1/2}}$, see Fig. 1. The derivation of the flux $F_{j+1/2}$ is based on the following model BVP, in which we ignore the time dependence of the unknown:

$$f_x = \left(\tfrac{1}{2}u^2 - \nu u_x\right)_x = 0, \quad x \in (x_j, x_{j+1}), \tag{3a}$$

$$u(x_j) = u_j = u_L, \qquad u(x_{j+1}) = u_{j+1} = u_R. \tag{3b}$$

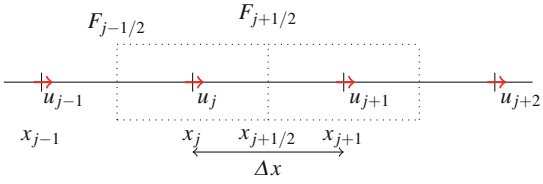


Fig. 1 Spatial discretization for the one-dimensional Burgers equation

The solution of the nonlinear BVP (3) provides us the numerical flux function $\mathscr{F}(u_L, u_R, \nu/\Delta x)$, which is constant on the interval (x_j, x_{j+1}). Thus, the numerical flux at the interface of the control volume $F_{j+1/2} = \mathscr{F}(u_L, u_R, \nu/\Delta x)$. Using the normalized coordinate $\sigma, \sigma \in [0, 1]$ and the parameter ε, defined by

$$\sigma := \frac{x - x_j}{\Delta x}, \quad \varepsilon := \frac{\nu}{\Delta x},$$

the BVP (3) can be expressed as

$$\left(\tfrac{1}{2}u^2 - \varepsilon u_\sigma\right)_\sigma = 0, \quad \sigma \in (0, 1), \tag{4a}$$

$$u(0) = u_L, \quad u(1) = u_R. \tag{4b}$$

Further, it can be shown that the above BVP has a monotonic solution.

Lemma 1 *The nonlinear local boundary value problem (4) has a strictly monotonic solution.*

Proof Any solution u of the problem can be represented as

$$u(\sigma) = u_L + (u_R - u_L)\frac{\Lambda(\sigma)}{\Lambda(1)}, \quad u'(\sigma) = (u_R - u_L)\frac{\lambda(\sigma)}{\Lambda(1)}, \quad ()' := \frac{\mathrm{d}}{\mathrm{d}\sigma},$$

for $\sigma \in [0, 1]$, where the functions $\lambda, \Lambda : [0, 1] \to \mathbb{R}$ are given by

$$\lambda(\sigma) := \exp\left(\frac{1}{\varepsilon}\int_0^\sigma u(\eta)\mathrm{d}\eta\right) \quad \text{and} \quad \Lambda(\sigma) := \int_0^\sigma \lambda(\xi)\mathrm{d}\xi.$$

For $u_L > u_R$, $u'(\sigma) < 0$ causing $u(\sigma)$ to be a monotonically decreasing function. Similarly, if $u_L < u_R$, then $u'(\sigma) > 0$ and u is monotonically increasing. □

3 The Numerical Flux Function

We now derive expressions for the numerical flux function using the BVP (4). As a consequence of Lemma 1 we consider the cases: $u_L > u_R$ and $u_L < u_R$.

3.1 The Case $u_L > U_R$

The solution of the BVP (4) in this case results in a (strictly) decreasing function, i.e., $u_\sigma < 0$. Using the left boundary condition $u(0) = u_L$, we get that the numerical flux at the interface, $F_{j+1/2}$, is given by

$$F_{j+1/2} = f(0) = \frac{1}{2}u_L^2 - \varepsilon u_\sigma(0). \tag{5}$$

Alternatively, the flux can be determined using the right boundary condition

$$F_{j+1/2} = \frac{1}{2}u_R^2 - \varepsilon u_\sigma(1). \tag{6}$$

Since $u_\sigma < 0$, we conclude that $F_{j+1/2} > 0$, therefore there exists a $c \in \mathbb{R}$, such that

$$F_{j+1/2} = \frac{1}{2}u^2 - \varepsilon u_\sigma = \frac{1}{2}u_L^2 - \varepsilon u_\sigma(0) = \tfrac{1}{2}u_R^2 - \varepsilon u_\sigma(1) = \frac{1}{2}c^2, \tag{7}$$

with $|c| \geq \max(|u_L|, |u_R|)$. The above relation gives us the first-order differential equation

$$\frac{\mathrm{d}u}{\mathrm{d}\sigma} = \frac{1}{2\varepsilon}(u^2 - c^2), \quad \sigma \in (0, 1), \tag{8}$$

which needs to satisfy both $u(0) = u_L$ and $u(1) = u_R$. Integrating the differential equation and connecting the left boundary condition with the right boundary condition results in the following nonlinear equation for the unknown c with parameters u_L, u_R and ε

$$H^+(c) := \log\left|\frac{(u_L + c)(u_R - c)}{(u_L - c)(u_R + c)}\right| - \frac{c}{\varepsilon} = 0. \tag{9}$$

Thus, $F_{j+1/2}$ is given by the non-trivial roots of the function $H^+(c)$, which is an odd function. We restrict ourselves to $c > 0$. Note that the nonlinear equation (9) can also be expressed as

$$e^{-c/2\varepsilon}\left|(u_L + c)(u_R - c)\right| - e^{c/2\varepsilon}\left|(u_L - c)(u_R + c)\right| = 0. \tag{10}$$

Let $s = (u_L + u_R)/2$, then for $s \geq 0$, we get that $u_L \geq |u_R|$ and the non-trivial solution of Eq. (10) satisfies $c \geq u_L \geq |u_R|$. In the inviscid limit $\varepsilon \to 0$, for $s \geq 0$ Eq. (10) reduces to

$$e^{c/2\varepsilon}(c - u_L)(c + u_R) = 0 \Rightarrow c = u_L.$$

Similarly for $s < 0$, we have $u_R < 0$, implying $c \geq -u_R \geq |u_L|$ and the limit case solution is then given by $c = -u_R (> 0)$. Thus, the numerical flux in the inviscid limit is given by

$$F_{j+1/2} = \begin{cases} \frac{1}{2}u_L^2, & \text{if } s \geq 0, \\ \frac{1}{2}u_R^2, & \text{if } s < 0, \end{cases} \tag{11}$$

which is actually the Godunov flux for the inviscid Burgers equation. Moreover, if $u_L = u_R = u$, then $u_\sigma = 0$ and the numerical flux is given by $F_{j+1/2} = \mathscr{F}(u, u) = \frac{1}{2}u^2 = f(u)$, for constant u. Hence the numerical flux function $\mathscr{F}$ is *consistent* with the continuous flux function f.

3.2 The Case $u_L < U_R$

From Lemma 1 we conclude that $u_\sigma > 0$ for $u_L < u_R$. Thus $F_{j+1/2} = u^2/2 - \varepsilon u_\sigma$ is positive if $\varepsilon u_\sigma < u^2/2$ and negative if $\varepsilon u_\sigma > u^2/2$. Therefore, we split the derivation of the numerical flux into two cases, depending on the sign of the flux.

Case 1: **Positive flux**

If the flux is positive, then the numerical flux is evaluated as for the case $u_L > u_R$ and is given by roots of the function $H^+(c)$, defined in Eq. (9), with $c \in (0, M)$, $M := \min(|u_L|, |u_R|)$.

Case 2: **Negative flux**

If the flux is negative, then there exists a $c \in \mathbb{R}$, such that

$$F_{j+1/2} = \frac{1}{2}u^2 - \varepsilon u_\sigma = -\frac{1}{2}c^2.$$

This relation gives rise to the first-order differential equation

$$\frac{\mathrm{d}u}{\mathrm{d}\sigma} = \frac{1}{2\varepsilon}(u^2 + c^2), \qquad \sigma \in (0, 1), \tag{12}$$

with the boundary conditions (4b). Integrating the first-order differential equation and connecting the left boundary condition with the right boundary condition gives us another nonlinear equation for c, i.e.,

$$H^-(c) := \arctan\left(\frac{u_R}{c}\right) - \arctan\left(\frac{u_L}{c}\right) - \frac{c}{2\varepsilon} = 0. \tag{13}$$

As before, the numerical value of $F_{j+1/2} = -c^2/2$ is given by the non-trivial roots of the function $H^-(c)$. We restrict ourselves to the case $0 < u_L < u_R$.

We now formulate the conditions for which $H^+(c)$ and $H^-(c)$ have non-trivial roots.

Lemma 2 *For $0 < u_L < u_R$, if the inequality*

$$\frac{1}{u_L} - \frac{1}{u_R} > \frac{1}{2\varepsilon}, \tag{14}$$

holds then $H^-(c)$ has a non-trivial solution, otherwise $H^+(c)$ has a non-trivial solution.

Proof Let $\alpha_-(c) := \arctan(u_R/c) - \arctan(u_L/c)$ and $\beta_-(c) := c/2\varepsilon$, such that $H^-(c) := \alpha_-(c) - \beta_-(c)$. Using $\arctan(1/z) = -\arctan(z) + \pi\,\mathrm{sgn}(z)/2$, $\alpha_-(c)$ can be expressed as

$$\alpha_-(c) = \arctan\left(\frac{c}{u_L}\right) - \arctan\left(\frac{c}{u_R}\right) + \frac{\pi}{2}\big(\mathrm{sgn}(u_R) - \mathrm{sgn}(u_L)\big).$$

Using the fact that $\alpha_-(c)$ is an odd function we restrict ourselves to the case $c > 0$. For $0 < u_L < u_R$, $\alpha_-(c)$ has a maximum at $c = \sqrt{u_L u_R}(< u_R)$. Clearly $H^-(c)$ has a non-trivial root whenever $\alpha_-(c) = \beta_-(c)$, i.e., the two functions intersect for $c > 0$, which is possible only if $\alpha'_-(0) > \beta'_-(0)$, or,

$$\alpha'_-(0) = \frac{1}{u_L} - \frac{1}{u_R} > \beta'_-(0) = \frac{1}{2\varepsilon}, \quad (.)' = \frac{\mathrm{d}}{\mathrm{d}c}.$$

Thus, if the above condition holds then $H^-(c)$ has a non-trivial solution, which satisfies $\sqrt{u_L u_R} < c < u_R$.

Next, we investigate the condition under which $H^+(c)$ has a non-trivial root. Let

$$z(c) := \frac{(u_R - u_L)c}{u_L u_R - c^2},$$

such that $z \in [0, 1]$ with $z(0) = 0, z(u_L) = 1$. Clearly, for $c \in (0, u_L)$ we have $z(c) > 0$. Using $z(c)$ we can rewrite Eq. (9) as

$$H^+(c) = \log\left(\frac{1 + z(c)}{1 - z(c)}\right) - \frac{c}{\varepsilon} = 2\,\mathrm{Artanh}(z(c)) - \frac{c}{\varepsilon}, \quad c \in (0, u_L).$$

Further, let $\alpha_+(c) := 2\mathrm{Artanh}(z(c))$ and $\beta_+(c) := c/\varepsilon$, such that $H^+(c) = \alpha_+(c) - \beta_+(c)$. For $c \in (0, u_L)$, $\alpha_+(c)$ is an increasing function, thus $H^+(c)$ has a non-trivial root only if $\alpha'_+(0) < \beta'_+(0)$. The derivative $\alpha'_+(c)$ is given by

$$\alpha'_+(c) = 2(u_R - u_L)\frac{1}{1 - z^2(c)}\frac{u_L u_R + c^2}{(u_L u_R - c^2)^2}.$$

The condition $\alpha'_+(0) < \beta'_1(0)$ translates to

$$\frac{1}{u_L} - \frac{1}{u_R} < \frac{1}{2\varepsilon}.$$

Lastly, integrating the differential equation (8) with $c = 0$ gives us the condition for *zero-flux*: $1/u_L - 1/u_R = 1/2\varepsilon$, which is in agreement with the above criteria for positive or negative flux. □

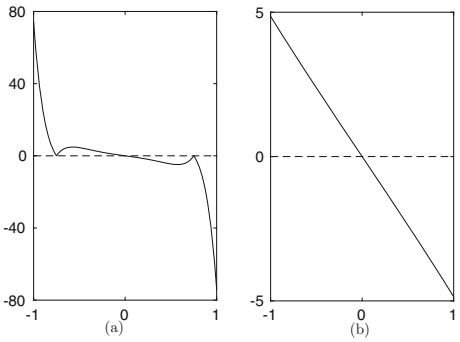


Fig. 2 $H^+(c)$ (a) and $H^-(c)$ (b), for $u_L = 0.75$, $u_R = 1.0$ and $\varepsilon = 0.1$

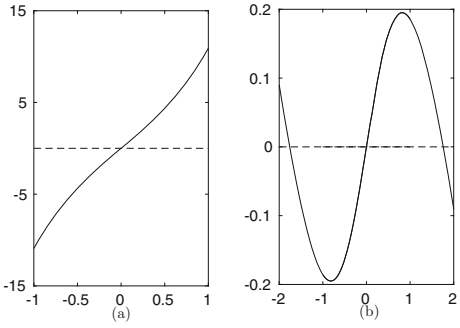


Fig. 3 $H^+(c)$ **a** and $H^-(c)$ **b**, for $u_L = 1$, $u_R = 10$ and $\varepsilon = 1$

Figure 2 shows the plots of the functions $H^+(c)$ and $H^-(c)$, for $u_L = 0.75$, $u_R = 1$ and $\varepsilon = 0.1$ not satisfying (14). Hence, $H^-(c)$ does not have a non-trivial root, unlike $H^+(c)$ which has a non-trivial root at $c = 0.749(\approx u_R)$. In Fig. 3, for $u_L = 1$, $u_R = 10$ and $\varepsilon = 1$, condition (14) is satisfied. Thus, $H^+(c)$ does not have a non-trivial root, whereas $H^-(c)$ has a non-trivial root, at $c = 1.7597$.

4 Numerical Results

We compare the proposed nonlinear local BVP scheme with the upwind scheme and the homogeneous flux scheme described in [4, 5]. In the homogeneous flux scheme, the numerical flux $F^{\text{lin}}_{j+1/2}$ is derived from a *linearized* homogeneous local two-point BVP and is given by

$$F^{\text{lin}}_{j+1/2} = \varepsilon\big(B(-P)u_L - B(P)u_R\big),$$

where $B(z) := z/(e^z - 1)$ is the Bernoulli function and $P := U_{j+1/2}/2\varepsilon$, is the grid Péclet number. The interface velocity $U_{j+1/2} = (u_L + u_R)/2$ is given by the central approximation. The availability of an analytical solution to the viscous Burgers

equation defined on $(0, 1) \times (0, T)$, $T \in (0, 1]$ provides us a reference solution to compare the schemes:

$$u^{\text{ref}}(x, t) = 1 + \frac{1}{2} \tanh\left(\frac{1}{4\nu}\left(x - 0.1 - \frac{1}{2}t\right)\right). \tag{15}$$

We use the explicit fourth-order Runge–Kutta scheme for the temporal discretization with $\Delta t = 10^{-3}$. For this test case, we have $0 < u_R < u_L$ throughout the computational domain, and the numerical flux is given by the roots of $H^+(c)$. The Newton solver converges in 2–8 iteration steps (depending on the tolerance, ranging from 10^{-3} to 10^{-8}), for a good initial guess. A fairly accurate initial guess can be derived using the bounds on the derivative u_σ, that can be obtained using Lemma 1. Figure 4 shows the convergence of the error $\mathbf{e_u} := |\mathbf{u} - \mathbf{u}^{\text{ref}}|_1$ for $\nu = 10^{-3}$ over a family of uniform grids. Grid refinement (for fixed ν) causes ε to increase (for the test case, $\varepsilon = 2^i \times 10^{-2}$, $i = 1, 2, \ldots, 6$). Moreover, it is observed that the root-finder converges faster for higher values of ε (≈ 1) than for smaller values of ε. On coarse grids all three schemes exhibit first-order accuracy, with the local BVP schemes being slightly more accurate than the upwind scheme. However, on grid refinement (increasing ε) the nonlinear BVP scheme is found to be more accurate compared to the upwind and the linearized local BVP scheme (Fig. 4). Further, Richardson extrapolation shows that for finer grids the nonlinear local BVP scheme exhibits second-order convergence.

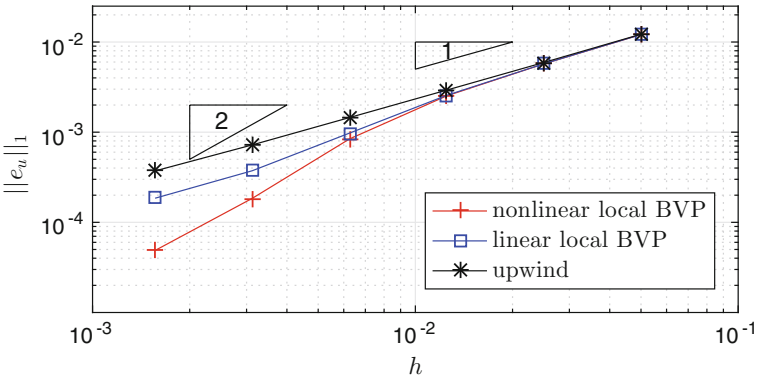


Fig. 4 Convergence of the 1-norm of the error e_u for $\nu = 10^{-3}$ for the proposed nonlinear local BVP scheme, homogeneous linear local BVP scheme and the upwind scheme for a family of grids ($\Delta x = 0.1 \times 2^{-i}$; $i = 1, 2, 3, 4, 5, 6$)

5 Conclusion

In this paper, we have presented a flux approximation scheme for the viscous Burgers equation, in which the numerical flux function is given by the solution of a local nonlinear two-point BVP, resulting in a locally exact approximation that corresponds with the nonlinearity of the flux function. The resulting numerical flux is shown to be consistent with the Godunov method in the inviscid limit and is more accurate than the linearized homogeneous approximation scheme in [4, 5].

In the future, we plan to extend the scheme by including source terms and also the time derivative into the local BVP, and by then solving the inhomogeneous BVP, to get the nonlinear complete-flux scheme.

Acknowledgements This work is part of the Industrial Partnership Programme (IPP) Computational Sciences for Energy Research of the Foundation for Fundamental Research on Matter (FOM), which is part of the Netherlands Organization for Scientific Research (NWO).

References

1. Allen, D.N.G., Southwell, R.V.: Relaxation methods applied to determine the motion, in two dimensions, of a viscous fluid past a fixed cylinder. Q. J. Mech. Appl. Math. **8**, 129–145 (1955)
2. Eymard, R., Fuhrmann, J., Gärtner, K.: A finite volume scheme for nonlinear parabolic equations derived from one-dimensional local Dirichlet problems. Numeri. Math. **102**, 463–495 (2006)
3. Il'in, A.M.: Differencing scheme for a differential equation with a small parameter affecting the highest derivative. Math. Notes Acad. Sci. USSR **6**, 596–602 (1969)
4. Kumar, N., ten Thije Boonkkamp, J., Koren, B.: Flux approximation cheme for the incompressible Navier–Stokes equations using local boundary value problems. In: Lecture Notes in Computational Science and Engineering, vol. 112, pp. 43–51. Springer, Heidelberg (2016)
5. ten Thije Boonkkamp, J.H.M., Anthonissen, M.J.H.: The finite volume-complete flux scheme for advection-diffusion-reaction equations. J. Sci. Comput. **46**, 47–70 (2011)

Mimetic Staggered Discretization of Incompressible Navier–Stokes for Barycentric Dual Mesh

René Beltman, Martijn J. H. Anthonissen and Barry Koren

Abstract A staggered discretization of the incompressible Navier–Stokes equations is presented for polyhedral non orthogonal nonsmooth meshes admitting a barycentric dual mesh. The discretization is constructed by using concepts of discrete exterior calculus. The method strictly conserves mass, momentum and energy in the absence of viscosity.

Keywords Mimetic finite-volume discretizations · Barycentric dual mesh

MSC (2010): 65M08 · 65N08 · 76D05

1 Introduction

In staggered methods the incompressible Navier–Stokes equations are discretized in terms of the normal velocity components at the cell faces and pressure variables in the cell-centers. Staggered mesh methods were introduced for Cartesian meshes by Harlow and Welch in the form of the MAC scheme [6]. The staggering allows for an efficient discretization of the divergence-free condition, leading to exact conservation of mass. It was subsequently shown [7] that, besides momentum and mass, the staggered Cartesian discretization also conserves the secondary quantities vorticity and kinetic energy, in the inviscid case.

The staggered mesh method was subsequently extended to unstructured meshes [5, 9]. In this formulation the orthogonality properties of a Delaunay-Voronoi dual

R. Beltman (✉) · M.J.H. Anthonissen · B. Koren
Department of Mathematics and Computer Science, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands
e-mail: r.beltman@tue.nl

M.J.H. Anthonissen
e-mail: m.j.h.anthonissen@tue.nl

B. Koren
e-mail: b.koren@tue.nl

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_49

mesh were exploited. Perot [10] showed that, on unstructured meshes, both for a discretization of the momentum equation in divergence-form and a discretization in rotation-form, kinetic energy is conserved. However, for the divergence-form conservation of momentum was proved but not conservation of vorticity, and, for the rotation-form conservation of vorticity was shown to be satisfied, but not conservation of momentum. Recently, a variational formulation of the MAC scheme was generalized to nonconforming meshes and proved to converge [3].

In most of the aforementioned cases, the primal mesh admits a circumcentric dual mesh. The circumcentric dual mesh has desirable orthogonality properties that allow for simple interpolation between the primal and dual meshes. For many meshes a circumcentric dual mesh does not exist. Such a situation is encountered in, for example, cut-cell methods [4]. In such cases a barycentric dual mesh can be used, although this type of dual mesh is harder to deal with, because it lacks orthogonality properties.

In this work we will present a barycentric discretization. We will discretize the divergence-form of the Navier–Stokes equations, given by

$$\frac{\partial \underline{u}}{\partial t} = (\underline{u} \cdot \nabla)\underline{u} - \frac{1}{\rho}\nabla p + \nu \Delta \underline{u}, \tag{1a}$$

$$0 = \nabla \cdot \underline{u}. \tag{1b}$$

The discretization presented in [12] will be generalized to polyhedral meshes. This will be done by showing that the mimetic inner product matrices [2] can be interpreted as discrete Hodge operators and by using them as such. They allow for polyhedral volumes with a varying number of faces. Furthermore, we will complete the dual mesh to a cell-complex and show that a discretization on this cell-complex leads to a method that conserves mass, momentum and energy (in the absence of viscosity) in the interior of the domain and also reproduces accurate boundary fluxes of these quantities. The discretization has a narrow stencil for the orthogonal part of the mesh.

2 Primal-Dual Mesh Structure and Discrete Exterior Calculus

In the continuous setting, conservation statements, like conservation of energy, can be derived from the primary equations by using fundamental properties of the continuous differential operators. Examples of this are that the curl of a gradient is always zero and the divergence of a curl is always zero. If the fundamental properties of the continuous differential operators can be transferred to the discrete setting, then it is possible to derive discrete conservation properties by similar arguments as in the continuous case. We will discretize in such a way that properties of the continuous differential operators are transferred to the discrete setting as much as possible.

2.1 The Primal Mesh and Incidence Matrices

The mesh $\mathscr{G} := \{\mathscr{P}, \mathscr{L}, \mathscr{S}, \mathscr{V}\}$ consists of a set of points $\mathscr{P}$, lines $\mathscr{L}$, surfaces $\mathscr{S}$ and volumes $\mathscr{V}$. The volumes $\mathscr{V}$ are polyhedra that exactly fill the flow domain Ω. The intersection of two volumes in $\mathscr{V}$ is either empty or it is a polygon part of the boundary of both. The set $\mathscr{S}$ is the union of the polygons making up the boundary of all the volumes in $\mathscr{V}$. Similarly, $\mathscr{L}$ is the union of the different line segments making up the boundaries of the polygons in $\mathscr{S}$ and $\mathscr{P}$ is the union of all endpoints of lines in $\mathscr{L}$.

We discretize the velocity field by integrating over the polygonal faces $s \in \mathscr{S}$:

$$u_s^{(2)} := \int_s \underline{u} \cdot \mathrm{d}\underline{A},$$

where $\mathrm{d}\underline{A}$ is an infinitesimal oriented surface area and the superindex indicates the dimension of s. The numbers $u_s^{(2)}$, $s \in \mathscr{S}$, are the discrete variables and, if we number the elements in $\mathscr{S}$, we can order them in a vector $\mathbf{u}^{(2)}$. We denote the space of possible $\mathbf{u}^{(2)}$ by $C^{(2)} = \mathbb{R}^{N_s}$, where N_s is the number of elements in $\mathscr{S}$. In a similar vein we can discretize a vector field by integrating it along the lines in $\mathscr{L}$ and define a space $C^{(1)}$. By integrating and evaluating a scalar function on the elements of $\mathscr{V}$ and $\mathscr{P}$, respectively, we can discretely represent this function on the volumes or points of our mesh and analogously define the spaces $C^{(3)}$ and $C^{(0)}$.

The integrated discrete variables are also known as discrete forms [8] or cochains [1]. They allow to discretize the divergence, gradient and curl (or really the exterior derivative) in such a way that the generalized Stokes theorem is valid in the finite number of situations provided by the mesh. To be able to define the discretizations of the divergence, gradient and curl we need to give an orientation to the elements in $\mathscr{G}$. The choice of orientation is arbitrary. Examples of oriented mesh elements are shown in Fig. 1. Suppose we are given the surface fluxes $u_s^{(2)}$. Using the divergence theorem we can determine the divergence of $\underline{u}$ integrated over all $v \in \mathscr{V}$:

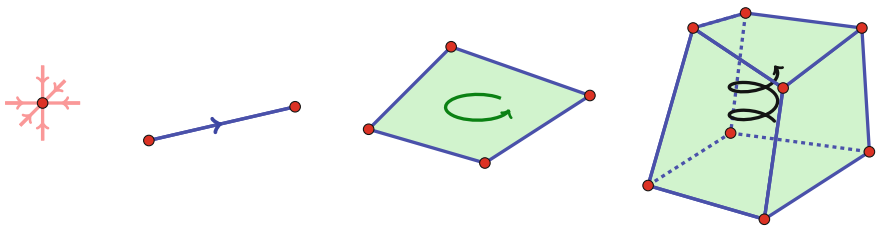


Fig. 1 A point $p \in \mathscr{P}$ is either classified as a sink (ingoing *arrows*) or a source (outgoing *arrows*), a *line* $l \in \mathscr{L}$ is oriented by a direction along the *line*, a face $s \in \mathscr{S}$ by a sense of rotation in its plane and a volume $v \in \mathscr{V}$ by a right- or left-hand-rule

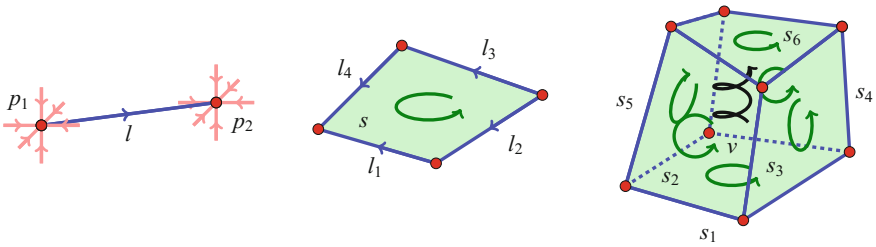


Fig. 2 For the line the orientations of p_2 and l agree, therefore $\alpha_l(p_2) = 1$. Similarly we have $\alpha_l(p_1) = -1$, $\alpha_s(l_1) = \alpha_s(l_2) = -1$, $\alpha_s(l_3) = \alpha_s(l_4) = 1$. Volume v has right-handed orientation indicated by a helix and we find $\alpha_v(s_1) = \alpha_v(s_4) = -1$ and $\alpha_v(s_2) = \alpha_v(s_3) = \alpha_v(s_5) = \alpha_v(s_6) = 1$

$$\int_v \nabla \cdot \underline{u} \, \mathrm{d}V = \sum_{s \in \partial v} \alpha_v(s) u_s^{(2)},$$

where $\alpha_v(s) = 1$ if the orientations of s and v agree and $\alpha_v(s) = -1$ otherwise. Examples are given in Fig. 2. In matrix notation we can express this for all $v \in \mathscr{V}$ at once as $\mathbf{d}^{(3)} = \mathbb{D}^{(3,2)}\mathbf{u}^{(2)}$, where the entry of $\mathbf{d}^{(3)}$ corresponding to $v \in \mathscr{V}$ is $d_v^{(3)} := \int_v \nabla \cdot \underline{v} \, \mathrm{d}V$ and

$$\mathbb{D}_{v,s}^{(3,2)} := \begin{cases} +1 & \text{if } s \in \partial v \text{ and the orientations of } s \text{ and } v \text{ agree,} \\ -1 & \text{if } s \in \partial v \text{ and the orientations of } s \text{ and } v \text{ disagree,} \\ 0 & \text{if } s \notin \partial v. \end{cases}$$

We see that the incidence matrix $\mathbb{D}^{(3,2)}$ gives the integral of the divergence of a vector field $\underline{u}$ over the volumes when it is applied to the discrete representation of this vector field on the surfaces. Similarly, using the fundamental theorem of calculus, the discrete representation of a function on $\mathscr{P}$ can be used to determine the integral of the gradient over the lines in $\mathscr{L}$. Suppose $\mathbf{q}^{(0)} \in C^{(0)}$ is this discrete representation, i.e., $q_p^{(0)} := q\big|_p$ then $[\mathbb{D}^{(1,0)}\mathbf{q}^{(0)}]_l = \int_l \nabla q \cdot \mathrm{d}\underline{l}$, where $\mathbb{D}_{l,p}^{(1,0)} = +1$, if $p \in \partial l$ and their orientations agree, $\mathbb{D}_{l,p}^{(1,0)} = -1$ if $p \in \partial l$ and their orientations disagree and $\mathbb{D}_{l,p}^{(1,0)} = 0$ if $p \notin \partial l$. A similarly defined matrix $\mathbb{D}^{(2,1)}$ returns the integral of the curl over surfaces in $\mathscr{S}$ of a vector field when applied to a discretization of this vector field on $\mathscr{L}$, representing an exact discretization of the Kelvin-Stokes theorem. The incidence matrices have the properties $\mathbb{D}^{(2,1)}\mathbb{D}^{(1,0)}\mathbf{a}^{(0)} = \mathbf{0}^{(2)}$ for all $\mathbf{a}^{(0)} \in C^{(0)}$ and $\mathbb{D}^{(3,2)}\mathbb{D}^{(2,1)}\mathbf{b}^{(1)} = \mathbf{0}^{(3)}$ for all $\mathbf{b}^{(1)} \in C^{(1)}$ representing the fact that the curl of a gradient and the divergence of a curl are zero [1].

2.2 The Dual Mesh and Discrete Hodge Operators

Primal mesh and incidence matrices alone are not sufficient to discretize the equations. We introduce a dual mesh which allows to conveniently interpolate between discrete variables defined on mesh elements of dimension k and dual mesh elements of dimension $3-k$. Using these interpolations we can, for example, apply the incidence matrix corresponding to the curl twice, once on the primal mesh and, after interpolation to the dual mesh, once on the dual mesh. This then allows for the construction of discretizations of higher order differential operators like the Laplacian. Moreover, the interpolation between the primal and dual mesh introduces the metric aspects of the differential equation in the discrete setting. The discrete operations on only the primal (or dual) mesh, as indicated by the ones and zeros of the incidence matrices, only depend on the topology of the mesh and not on the lengths, areas or volumes of the mesh elements. These metrical notions only play a role in the interpolation between the primal and dual mesh. This interpolation is also the place where the discretization error enters the method.

The dual mesh $\tilde{\mathscr{G}} := \{\tilde{\mathscr{V}}, \tilde{\mathscr{S}}, \tilde{\mathscr{L}}, \tilde{\mathscr{P}}\}$ consists of the set of points $\tilde{\mathscr{V}}$ which are dual to the volumes in $\mathscr{V}$, the set of lines $\tilde{\mathscr{S}}$ dual to the surfaces in $\mathscr{S}$, etc. Everything related to the dual mesh will be given a tilde. In the introduction we mention the circumcentric dual mesh, constructed by connecting the circumcenters of neighboring primal volumes, and the barycentric dual mesh, constructed by connecting the barycenters of every primal volume with the barycenters of the primal surfaces that constitute the boundary of that primal volume. It should be noted that the line elements $\tilde{\mathscr{S}}$ of the barycentric dual mesh consist of two straight line segments. The dual mesh elements are given an outer orientation, i.e. an orientation of their complement in the ambient Euclidean three-dimensional space. This orientation will be chosen to coincide with the orientation of their corresponding primal cells. For the dual mesh, analogously to the primal mesh, we define discrete spaces $C^{(\tilde{k})}$, $k = 0, 1, 2, 3$, and incidence matrices $\mathbb{D}^{\widetilde{(k+1)},\tilde{k}}$, $k = 0, 1, 2$. Note that $C^{(k)} = C^{\widetilde{(3-k)}}$, because of the bijection between $\tilde{\mathscr{G}}$ and $\mathscr{G}$. If the dual mesh elements are numbered in the same way as the primal mesh elements and the outer orientation for the dual mesh is chosen to coincide with the primal mesh, then $\mathbb{D}^{\widetilde{(k+1)},\tilde{k}} = \mathbb{D}^{(3-k,3-k-1),T}$, $k = 0, 1, 2$, see [13].

The primal mesh is a so-called cell-complex, because the boundary of every element in $\mathscr{G}$ is either a union of lower dimensional elements in $\mathscr{G}$ or empty. This property implies that a discrete divergence theorem holds for the complete mesh. The dual mesh $\tilde{\mathscr{G}}$ is not a cell-complex, because at the boundary $\partial\Omega$ boundary cells are missing. To be able to derive conservation statements up to the boundary we need to complete the dual mesh to a cell-complex. This can be done in the following way. Let us denote the mesh elements of the primal mesh that make up the boundary $\partial\Omega$ by $\mathscr{G}_b$. We take the dual mesh to $\mathscr{G}_b$ within $\partial\Omega$, which we denote by $\tilde{\mathscr{G}}_b$. The $(n-1)$-dimensional dual mesh $\tilde{\mathscr{G}}_b$ is a cell-complex because $\partial\partial\Omega = \emptyset$, and, moreover, the

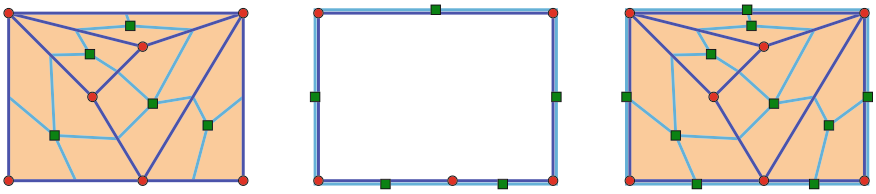


Fig. 3 On the *left*, we show a primal mesh $\mathcal{G}$ (*blue*/*red*) and its barycentric dual $\tilde{\mathcal{G}}$ (*light blue*/*green*). In the *middle*, we show the boundary part $\mathcal{G}_b$ of the primal mesh and the corresponding barycentric dual $\tilde{\mathcal{G}}_b$ within $\partial\Omega$. On the *right*, we show how $\tilde{\mathcal{G}}$ and $\tilde{\mathcal{G}}_b$ combine to form the cell-complex $\bar{\tilde{\mathcal{G}}}$

set $\tilde{\mathcal{G}}_b$ is exactly the set that completes $\tilde{\mathcal{G}}$ to a cell-complex $\bar{\tilde{\mathcal{G}}} := \tilde{\mathcal{G}} \cup \tilde{\mathcal{G}}_b$. This is illustrated in Fig. 3. We will subsequently partly define the variables and discretization on $\bar{\tilde{\mathcal{G}}}$, the fact that this is a cell-complex allows us to derive conservation statements for momentum and energy neatly up to the boundary.

The interpolation matrices between $C^{(k)}$ and $C^{\widetilde{(3-k)}}$ are called discrete Hodge operators, because they map discrete k-forms to discrete $(3-k)$-forms, like the continuous Hodge operators map differential k-forms to differential $(3-k)$-forms. When the circumcentric dual mesh is used, the orthogonality of the primal and dual mesh allow for consistent diagonal discrete Hodge operators [8]. When the barycentric dual mesh is used, diagonal discrete Hodge operators are still available for interpolation between $C^{(0)}$ and $C^{(\tilde{3})}$, and, $C^{(3)}$ and $C^{(\tilde{0})}$, by using the volume averages. However, for interpolation between primal lines and dual surfaces, and, primal surfaces and dual lines, no consistent diagonal Hodge operators exist for general polyhedra in the barycentric case.

As barycentric Hodge operators we will use the mimetic inner product matrices used in the mimetic finite difference method. It can be shown that the mimetic inner product matrices $M_{\mathscr{L}}$ and $M_{\mathscr{S}}$ presented in [2], which define the mimetic inner product on the spaces $C^{(1)}$ and $C^{(2)}$, respectively, can be interpreted as discrete Hodge operators mapping from $C^{(k)}$ to $C^{(3-k)}$, for $k = 1, 2$ for a barycentric dual mesh. It is shown in [14] that these matrices are consistent and stable for polyhedral meshes with only minor assumptions. We use the notation $\mathbb{H}^{(\tilde{2},1)}$ and $\mathbb{H}^{(\tilde{1},2)}$ for, respectively, $M_{\mathscr{L}}$ and $M_{\mathscr{S}}$ to indicate that they map from the primal mesh to the dual mesh. In the parts of the mesh where the primal and dual mesh are orthogonal we will use the diagonal Hodge operator instead.

These Hodge operators are symmetric and positive definite for mesh cells with varying number of faces and therefore allow for a generalization of the method described in [11] from simplicial meshes to polyhedral meshes.

3 Barycentric Discretization

We use a barycentric dual mesh and a discretization similar to [11], but we discretize the viscous term differently by employing the aforementioned Hodge operators. Moreover, we define the pressure at all the points in the dual cell-complex $\bar{\tilde{\mathscr{G}}}$ and solve an extra equation for the faces in $\mathscr{G}_b$, which allows conservation statements that hold up to the boundary. Furthermore we introduce extra vorticity variables only for the primal edges in the regions where the Hodge operator $\mathbb{H}^{(\tilde{2},1)}$ is not diagonal. This results in an efficient treatment of the nonorthogonal part of the mesh by avoiding the inversion of $\mathbb{H}^{(\tilde{2},1)}$ while still allowing for nonorthogonal polyhedral meshes. We discretize (1a) by approximating its line integral over dual line elements.

To define the convection operator we need a primal volume vector reconstruction operator that approximates the velocity vector in the barycenter of the primal volumes v by using the flux variables $u_s^{(2)}$ for $s \in \partial v$. The integral of the velocity field over the primal cells, can be approximated [12] according to

$$\int_v \underline{u}\, \mathrm{d}V \cong \sum_{s\in\partial v} \alpha_v(s)(\underline{x}_s - \underline{x}_v)u_s^{(2)},$$

where $\underline{x}_s$ and $\underline{x}_v$ are barycenters of s and v, respectively. This is a first order approximation that holds for arbitrary polygons and is a second order approximation when the mesh is uniform [12]. It can be written as $\mathbb{R}\mathbf{u}^{(2)}$, where $\mathbb{R}$ is the $3N_v \times N_s$ matrix with $\mathbb{R}^i_{v,s} = \alpha_v(s)(x_s^i - x_v^i)$, where $i = 1, 2, 3$ indicates the vector components. Let $\underline{\underline{\mathbb{H}}}^{(\tilde{0},3)}$ be the $3N_v \times 3N_v$ matrix that divides the vector components by the cell volume of the corresponding cell. Then $\underline{\underline{\mathbb{H}}}^{(\tilde{0},3)}\mathbb{R}\mathbf{u}^{(2)}$ gives an, in general first order, approximation of the vector field $\underline{u}$ in the dual points $\tilde{v} \in \tilde{\mathscr{V}}$, denoted by $\underline{u}_{\tilde{v}}^{(\tilde{0})}$. Given a vector field $\underline{c}$ discretized in dual mesh points $\tilde{v} \in \tilde{V}$ as $\underline{c}^{(\tilde{0})}$, an approximation of the line integral of this vector field over the dual lines $\tilde{s} \in \tilde{\mathscr{S}}$ is then given by [12]

$$\int_{\tilde{s}} \underline{c} \cdot \mathrm{d}\underline{l} \cong \sum_{\tilde{v}\in\partial\tilde{s}} \alpha_{\tilde{s}}(\tilde{v})\underline{c}_{\tilde{v}}^{(\tilde{0})} \cdot (\underline{x}_s - \underline{x}_v). \tag{2}$$

Outer orientations of $\tilde{s}$ and $\tilde{v}$ agree if and only if the orientations of s and v agree, hence (2) is written for all dual mesh lines at once as $\mathbb{R}^T\underline{\mathbf{c}}^{(\tilde{0})}$. These approximation properties of $\mathbb{R}$ and $\mathbb{R}^T$ will be used in the discretization of the convection term.

Integrating the convective term over primal cells and applying the divergence theorem we obtain

$$\int_v (\underline{u}\cdot\nabla)\underline{u}\, \mathrm{d}V = \sum_{s\in\partial v} \alpha_v(s) \int_s \underline{u}\, \underline{u}\cdot \mathrm{d}\underline{A}. \tag{3}$$

The surface integrals will be approximated as

$$\int_s \underline{u}\,\underline{u}\cdot \mathrm{d}\underline{A} \cong \sum_{\{v|s\in\partial v\}} \frac{1}{2}\underline{u}_{\tilde{v}}^{(\tilde{0})} u_s^{(2)}.$$

Let $\underline{\mathbf{u}}^{(\tilde{0})} = \underline{\mathbb{H}}^{(\tilde{0},3)}\mathbb{R}\mathbf{u}^{(2)}$ and let $\mathbb{A}[\underline{\mathbf{u}}^{(\tilde{0})}]$ be the $3N_s \times N_s$ matrix consisting of the three $N_s \times N_s$ diagonal blocks with on the diagonal element corresponding to s, respectively the x-, y- or z-component of $\sum_{\{v|s\in\partial v\}} \underline{u}_{\tilde{v}}^{(\tilde{0})}/2$. Using this matrix we can write the approximation of (3) for all $v \in \mathscr{V}$ at once as $\underline{\mathbb{D}}^{(3,2)}\mathbb{A}[\underline{\mathbf{u}}^{(\tilde{0})}]\mathbf{u}^{(2)}$, where $\underline{\mathbb{D}}^{(3,2)}$ is a componentwise version of $\mathbb{D}^{(3,2)}$. Subsequently applying $\underline{\mathbb{H}}^{(\tilde{0},3)}$ and $\mathbb{R}^T$ gives $\mathbb{C}[\mathbf{u}^{(2)}]\mathbf{u}^{(2)} := \mathbb{R}^T\underline{\mathbb{H}}^{(\tilde{0},3)}\underline{\mathbb{D}}^{(3,2)}\mathbb{A}[\underline{\mathbf{u}}^{(\tilde{0})}]\mathbf{u}^{(2)}$, which is an approximation of the convection term integrated over the dual line integrals.

For the primal lines $l \in \mathscr{L}$ for which $\mathbb{H}^{(\tilde{2},1)}$ contains off-diagonal terms we introduce the vorticity variables $\omega_l^{(1)}$. We collect them in a vector $\boldsymbol{\omega}^{(b1)}$, where b in the superindex indicates that the complete barycentric Hodge operator applies. To split the primal lines in the set where $\mathbb{H}^{(\tilde{2},1)}$ is diagonal and the set where it is non-diagonal we introduce matrices $\mathbb{I}_b : C^{(1)} \to C^{(b1)}$ and $\mathbb{I}_d : C^{(1)} \to C^{(d1)}$, where $\mathbb{I}_b$ is the identity with the rows corresponding to the lines where $\mathbb{H}^{(\tilde{2},1)}$ is the diagonal eliminated and $\mathbb{I}_d$ the same but then with the rows corresponding to lines where $\mathbb{H}^{(\tilde{2},1)}$ is non-diagonal eliminated.

For the pressure term we use a straightforward discretization by applying the discrete gradient operator $\bar{\mathbb{D}}^{(\tilde{1},\tilde{0})}$, where the bar indicates that this is the extension of $\mathbb{D}^{(\tilde{1},\tilde{0})}$ to the dual cell-complex. The complete semi-discrete system is given by

$$\begin{bmatrix} \mathbb{H}^{(\tilde{1},2)}\partial_t + \mathbb{C}[\mathbf{u}^{(2)}] + \nu\mathbb{L} & \mathbb{H}^{(\tilde{1},2)}\mathbb{D}^{(2,1)}\mathbb{I}_b^T & -\bar{\mathbb{D}}^{(\tilde{1},\tilde{0})} \\ \mathbb{I}_b\mathbb{D}^{(\tilde{2},\tilde{1})}\mathbb{H}^{(\tilde{1},2)} & \nu^{-1}\mathbb{I}_b\mathbb{H}^{(\tilde{2},1)}\mathbb{I}_b^T & 0 \\ -\bar{\mathbb{D}}^{(\tilde{1},\tilde{0}),T} & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u}^{(2)} \\ \boldsymbol{\omega}^{(b1)} \\ \mathbf{p}^{(\tilde{0})} \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1^{(\tilde{1})} \\ \mathbf{r}_2^{(b1)} \\ \mathbf{r}_3^{(3)} \end{bmatrix},$$

where $\mathbb{L} := \mathbb{H}^{(\tilde{1},2)}\mathbb{D}^{(2,1)}\mathbb{I}_d^T(\mathbb{I}_d\mathbb{H}^{(\tilde{2},1)}\mathbb{I}_d^T)^{-1}\mathbb{I}_d\mathbb{D}^{(\tilde{2},\tilde{1})}\mathbb{H}^{(\tilde{1},2)}$ and it should be noted that $\mathbb{I}_d\mathbb{H}^{(\tilde{2},1)}\mathbb{I}_d^T$ is diagonal. The right-hand side vector incorporates the Dirichlet boundary condition on the velocity. In this semi-discrete system the viscous term in the momentum equation consists of two contributions $\nu\mathbb{L}\mathbf{u}^{(2)}$ and $\mathbb{H}^{(\tilde{1},2)}\mathbb{D}^{(2,1)}\mathbb{I}_b\boldsymbol{\omega}^{(b1)}$ corresponding to the orthogonal and non-orthogonal parts of the mesh, respectively.

From these discrete equations we can derive the discrete conservation laws that correspond to

$$\partial_t \int_\Omega \underline{u}\,\mathrm{d}V = -\int_{\partial\Omega} \underline{u}(\underline{u}\cdot\underline{n}) + p\underline{n} - \nu\nabla\underline{u}\cdot\underline{n}\,\mathrm{d}A,$$

$$\partial_t \int_\Omega \frac{1}{2}(\underline{u}\cdot\underline{u})\,\mathrm{d}V = -\int_{\partial\Omega} \frac{1}{2}\underline{u}\cdot\underline{u}(\underline{u}\cdot\underline{n}) + p(\underline{u}\cdot\underline{n}) + \nu(\underline{\omega}\times\underline{u})\cdot\underline{n}\,\mathrm{d}A - \nu\int_\Omega \underline{\omega}\cdot\underline{\omega}\,\mathrm{d}V.$$

Details will be given at FVCA8 and in a forthcoming publication.

4 Future Work

We have presented a mimetic staggered discretization of the incompressible Navier–Stokes equations. The method uses the barycentric dual mesh and discrete exterior calculus. This discretization will be used to develop a cut-cell method for modeling flow around complex objects by using a Cartesian mesh. The resulting efficient method, despite using a mesh not aligned with the objects, is anticipated to still be physically accurate as a result of its many conservation properties.

References

1. Bochev, P.B., Hyman, J.M.: Principles of mimetic discretizations of differential operators. In: Compatible Spatial Discretizations, pp. 89–119. Springer, Heidelberg (2006)
2. Brezzi, F., Buffa, A., Manzini, G.: Mimetic scalar products of discrete differential forms. J. Comput. Phys. **257**, 1228–1259 (2014)
3. Chénier, E., Eymard, R., Gallouët, T., Herbin, R.: An extension of the MAC scheme to locally refined meshes: convergence analysis for the full tensor time-dependent Navier–Stokes equations. Calcolo **52**(1), 69–107 (2015)
4. Cheny, Y., Botella, O.: The LS-stag method: a new immersed boundary/level-set method for the computation of incompressible viscous flows in complex moving geometries with good conservation properties. J. Comput. Phys. **229**, 1043–1076 (2010)
5. Hall, C., Cavendish, J., Frey, W.: The dual variable method for solving fluid flow difference equations on Delaunay triangulations. Comput. Fluids **20**, 145–164 (1991)
6. Harlow, F.H., Welch, J.E.: Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface. Phys. Fluids **8**, 2182 (1965)
7. Lilly, D.K.: On the computational stability of numerical solutions of time-dependent non-linear geophysical fluid dynamics problems. Mon. Weather Rev. **93**, 11–26 (1965)
8. Mohamed, M.S., Hirani, A.N., Samtaney, R.: Discrete exterior calculus discretization of incompressible Navier–Stokes equations over surface simplicial meshes. J. Comput. Phys. **312**, 175–191 (2016)
9. Nicolaides, R.: Flow discretization by complementary volume techniques. In: Proceedings of the 9th AIAA CFD Meeting. AIAA Paper 89-1978 (1989)
10. Perot, B.: Conservation properties of unstructured staggered mesh schemes. J. Comput. Phys. **159**, 58–89 (2000)
11. Perot, B., Nallapati, R.: A moving unstructured staggered mesh method for the simulation of incompressible free-surface flows. J. Comput. Phys. **184**, 192–214 (2003)
12. Perot, J.B., Vidovic, D., Wesseling, P.: Mimetic reconstruction of vectors. In: Compatible Spatial Discretizations, pp. 173–188. Springer, Heidelberg (2006)
13. Tonti, E.: The Mathematical Structure of Classical and Relativistic Physics. Springer, Heidelberg (2013)
14. da Veiga, L.B., Lipnikov, K., Manzini, G.: The Mimetic Finite Difference Method for Elliptic Problems, vol. 11. Springer, Heidelberg (2014)

A Reduced-Basis Approach to Two-Phase Flow in Porous Media

Sébastien Boyaval, Guillaume Enchéry, Riad Sanchez and Quang Huy Tran

Abstract Reduced-basis methods (RB) have demonstrated their efficiency for a wide variety of problems, most of which are elliptic PDEs solved by finite element methods. In this work, we attempt to apply the RB philosophy to a simple "real-life" model for two-phase flows in porous media, whose reference scheme is a finite volume method. This model is parameterized by the viscosity of water. Because of the mixed parabolic-elliptic nature of the system, we first propose to restrict the RB approach to the pressure subsystem corresponding to the end time. The resulting parametric dependence is, however, much more intricate than in the classical examples. This difficulty will be discussed and illustrated by numerical results.

Keywords Two-phase flow · Finite volumes · Reduced-basis · Aposteriori error estimate · Empirical interpolation

MSC (2010): 35J50 · 65M08 · 65N15 · 76S05

S. Boyaval
Laboratoire D'hydraulique Saint-Venant, Ecole des Ponts ParisTech–EDF R&D–CEREMA, 6 Quai Watier, BP 49, 78401 Chatou Cedex, France
e-mail: sebastien.boyaval@enpc.fr

G. Enchéry · R. Sanchez (✉) · Q.H. Tran
IFP Energies Nouvelles, 1 Et 4 Avenue de Bois-Préau, 92852 Rueil-malmaison, France
e-mail: riad.sanchez@ifpen.fr

G. Enchéry
e-mail: guillaume.enchery@ifpen.fr

Q.H. Tran
e-mail: quang-huy.tran@ifpen.fr

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_50

1 Model Problem and Parametrization

In reservoir engineering, one simple model for describing the sweeping of oil (o) by water (w) in a porous domain $\Omega \subset \mathbb{R}^d$ over a time interval $(0, T)$ can be derived by assuming that both fluids are incompressible and by neglecting gravity as well as capillarity effects. After expressing the balance laws of the two phases $\alpha \in \{o, w\}$ and invoking Darcy's laws, we obtain

$$\phi \partial_t S_\alpha + \nabla \cdot \mathbf{v}_\alpha = 0 \qquad \text{in } \Omega \times (0, T), \tag{1a}$$

$$\mathbf{v}_\alpha = -\mu_\alpha^{-1} k k_{r_\alpha}(S_\alpha) \nabla P \qquad \text{in } \Omega \times (0, T). \tag{1b}$$

The unknowns are the two saturations S_α, that satisfy the additional equation

$$S_o + S_w = 1, \tag{2}$$

the two velocities $\mathbf{v}_\alpha$ and the common pressure P. The (supposedly known) data are the rock porosity ϕ and permeability k, the two relative permeabilities k_{r_α} and viscosities μ_α. System (1)–(2) is completed by the boundary and initial conditions

$$P = P_{\mathrm{D}} \qquad \text{on } \Gamma_{\mathrm{D}} \times (0, T), \tag{3a}$$

$$\nabla P \cdot \mathbf{n} = 0 \qquad \text{on } \Gamma_{\mathrm{N}} \times (0, T), \tag{3b}$$

$$S_w = 1 \qquad \text{if } \nabla P \cdot \mathbf{n} < 0 \text{ on } \Gamma_{\mathrm{D}} \times (0, T), \tag{3c}$$

$$S_w(\cdot, 0) = 0, \qquad \text{in } \Omega, \tag{3d}$$

where $\Gamma_{\mathrm{D}} \cup \Gamma_{\mathrm{N}} = \partial\Omega$, $\Gamma_{\mathrm{D}} \cap \Gamma_{\mathrm{N}} = \emptyset$, is a prescribed Dirichlet-Neumann decomposition of the boundary $\partial\Omega$.

Setting $S = S_w$, summing the mass balances (1a) over α and introducing the total velocity $\mathbf{v} = \mathbf{v}_o + \mathbf{v}_w$, we obtain the system

$$\mathbf{v} + k\lambda(S)\nabla P = 0 \qquad \text{in } \Omega \times (0, T) \tag{4a}$$

$$\nabla \cdot \mathbf{v} = 0 \qquad \text{in } \Omega \times (0, T), \tag{4b}$$

$$\phi \partial S_t + \nabla \cdot (f(S)\mathbf{v}) = 0 \qquad \text{in } \Omega \times (0, T), \tag{4c}$$

in the unknowns $(S, \mathbf{v}, P)$, the auxiliary functions

$$\lambda(S) = \lambda_w(S) + \lambda_o(1 - S), \quad \lambda_\alpha(S) = \frac{k_{r_\alpha}(S)}{\mu_\alpha}, \quad f(S) = \frac{\lambda_w(S)}{\lambda(S)} \tag{5}$$

being respectively the total mobility, the mobility of phase α and the fractional flux of water. Problem (3)–(4) is the definitive form that we shall be working with.

From a broader perspective, (3)–(4) ought to be regarded as a forward problem within an optimization process whose purpose is to calibrate the petrophysical properties ϕ, k, k_{r_α} and μ_α. These are indeed often poorly known by geoscientists for

various reasons. In this study, we assume that ϕ, k, k_{r_α} and μ_o are well determined but $\mu_w =: \mu$ has to be calibrated. Then, the inverse problem requires a large number of forward problems to be numerically solved, which correspond to a large number of varying trial parameters μ.

Nevertheless, a single forward simulation is already quite expensive. In this respect, the objective of a RB method is to significantly decrease the amount of computational time associated to many forward problems, at the price of some further approximation errors (see [2] for other developments of reduced order modelling on porous media). Before constructing such a RB method, let us review the reference numerical scheme used for a single forward simulation, the solution of which will be called reference solution.

2 Reference Scheme

The two-phase flow problem (4) is discretized in time using the so-called IMPIMS (implicit in pressure, implicit in saturation) scheme, which reads

$$\mathbf{v}^{n+1} + k\lambda(S^n)\,\nabla P^{n+1} = 0, \tag{6a}$$

$$\nabla \cdot \mathbf{v}^{n+1} = 0, \tag{6b}$$

$$\phi \frac{S^{n+1} - S^n}{\Delta t^n} + \nabla \cdot (f(S^{n+1})\mathbf{v}^{n+1}) = 0, \tag{6c}$$

where $\Delta t^n = t^{n+1} - t^n$ denotes the time-step. In reservoir engineering, system (6) is usually discretized in space by a finite volume method [4]. To this end, the domain Ω is assumed polygonal and is partitioned by an admissible mesh $(\mathcal{M}, \mathcal{E})$ in the sense of [4, Definition 3.1]. In each cell $K \in \mathcal{M}$, we define the data (ϕ_K, k_K) and the unknowns (S_K, P_K). If $\sigma \in \mathcal{E}$ is an edge contained in ∂K, the distance from the center of K to σ is designated by $\mathrm{d}_{K,\sigma}$.

The two-point flux approximation (TPFA) of $\nabla \cdot (-k\lambda(S^n)\nabla P^{n+1}) = 0$, which results from (6a)–(6b), proceeds as follows. To each pair (K, σ), where $\sigma \in \mathcal{E}$ is an edge contained in ∂K, we associate the discrete flux

$$F_{K,\sigma} = \begin{cases} \tau_\sigma^n (P_K^{n+1} - P_L^{n+1}) & \text{if } \sigma = K|L, \\ \tau_\sigma^n (P_K^{n+1} - P_{\mathrm{D},\sigma}^{n+1}) & \text{if } \sigma \subset \Gamma_\mathrm{D} \cap \partial K, \\ 0 & \text{if } \sigma \subset \Gamma_\mathrm{N} \cap \partial K, \end{cases} \tag{7}$$

intended to be an approximation of the outgoing flux $\int_\sigma -k\lambda(S^n)\nabla P^{n+1} \cdot \mathbf{n}_{K,\sigma}\,\mathrm{d}s$ from K through σ. The transmissivity τ_σ^n in (7) is given by

$$\tau_\sigma^n = k_\sigma \lambda_\sigma^n \frac{|\sigma|}{\mathrm{d}_\sigma}, \tag{8}$$

where $|\sigma|$ denotes the measure of σ and

$$\mathrm{d}_\sigma = \mathrm{d}_{K,\sigma} + \mathrm{d}_{L,\sigma}, \quad k_\sigma = \frac{k_K k_L \mathrm{d}_\sigma}{k_K \mathrm{d}_{L,\sigma} + k_L \mathrm{d}_{K,\sigma}}, \quad \lambda_\sigma^n = \frac{\lambda(S_K^n) + \lambda(S_L^n)}{2} \tag{9}$$

if $\sigma = K|L$ is an inner edge, whereas

$$\mathrm{d}_\sigma = \mathrm{d}_{K,\sigma}, \quad k_\sigma = k_K, \quad \lambda_\sigma^n = \lambda(S_K^n) \tag{10}$$

if $\sigma \subset \Gamma_\mathrm{D} \cap \partial K$ is a Dirichlet boundary edge. The pressures $\{P_K^{n+1}\}_{K\in\mathcal{M}}$ are then required to solve the linear system

$$\sum_{\sigma\in\mathcal{E},\, \sigma\subset\partial K} F_{K,\sigma} = 0, \qquad K \in \mathcal{M}, \tag{11}$$

which expresses the fluid volume balance over the cells.

The set of Eqs. (11) can be interpreted as the optimality condition for the minimization problem

$$\{P_K^{n+1}\}_{K\in\mathcal{M}} = \arg \min_{q\in Q_\mathcal{N}} \mathscr{E}_\mathcal{N}(q), \tag{12}$$

in which $\mathcal{N}$ denotes the number of cells in the mesh, $Q_\mathcal{N}$ is the $\mathcal{N}$-dimensional space of cellwise-constant real-valued functions and

$$\mathscr{E}_\mathcal{N}(q) = \frac{1}{2} \sum_{\substack{\sigma\in\mathcal{E} \\ \sigma=K|L}} \tau_\sigma^n |q_K - q_L|^2 + \frac{1}{2} \sum_{\substack{\sigma\in\mathcal{E} \\ \sigma\subset\Gamma_\mathrm{D}\cap\partial K}} \tau_\sigma^n |q_K - P_{\mathrm{D},\sigma}^{n+1}|^2 \tag{13}$$

represents an energy functional defined over $Q_\mathcal{N}$. In this light, it is possible to give (11) a "variational" form: find $P^{n+1} =: P_\mathcal{N} \in Q_\mathcal{N}$ such that

$$a_\mathcal{N}(P_\mathcal{N}, q) = b_\mathcal{N}(q) \qquad \text{for all } q \in Q_\mathcal{N}, \tag{14}$$

using the bilinear form on $Q_\mathcal{N} \times Q_\mathcal{N}$

$$a_\mathcal{N}(p, q) = \sum_{\substack{\sigma\in\mathcal{E} \\ \sigma=K|L}} \tau_\sigma^n (p_K - p_L)(q_K - q_L) + \sum_{\substack{\sigma\in\mathcal{E} \\ \sigma\subset\Gamma_\mathrm{D}\cap\partial K}} \tau_\sigma^n p_K q_K \tag{15}$$

and the linear form on $Q_\mathcal{N}$

$$b_\mathcal{N}(q) = \sum_{\substack{\sigma\in\mathcal{E} \\ \sigma\subset\Gamma_\mathrm{D}\cap\partial K}} \tau_\sigma^n P_{\mathrm{D},\sigma}^{n+1} q_K. \tag{16}$$

The energy viewpoint (12)–(16) is most helpful for constructing a RB method.

3 Reduced-Basis Method

As explained in Sect. 1, we have to solve a great many $\mathcal{N} \times \mathcal{N}$ linear systems (14), which depend on a water viscosity μ that ranges in some set of parameters $\mathcal{P}$. Here, the influence of μ on the solution is more subtle than in pure elliptic problems. A change in μ has an impact on λ_w and f, as can be seen from (5). By the transport Eq. (4c), the whole history of S is thus modified. This in turn alters $\lambda(S)$ in the Darcy velocity (4a), as well as the energy (13) and the forms (15)–(16).

To alleviate notations while highlighting the influence of μ, from now on we omit all time superscripts n or $n+1$ but write out μ whenever necessary. The characterization (14) becomes: find $P_\mathcal{N}(\mu) \in Q_\mathcal{N}$ such that

$$a_\mathcal{N}(P_\mathcal{N}(\mu), q; \mu) = b_\mathcal{N}(q; \mu) \qquad \text{for all } q \in Q_\mathcal{N}. \tag{17}$$

Let $N \ll \mathcal{N}$ and consider $\mathcal{P}_N = \{\mu_\ell\}_{\ell=1}^N$ a sample of parameters from $\mathcal{P}$. The N-dimensional space $Q_N = \text{span}\{P_\mathcal{N}(\mu_\ell), \quad \mu_\ell \in \mathcal{P}_N\}$ is plainly a subspace of $Q_\mathcal{N}$, so that the energy functional $\mathscr{E}_\mathcal{N}$ is well-defined on Q_N. Instead of minimizing it over $Q_\mathcal{N}$ as in (12), we content ourselves with a minimization over Q_N to obtain

$$P_N(\mu) = \arg \min_{q \in Q_N} \mathscr{E}_\mathcal{N}(q; \mu) \tag{18}$$

for all $\mu \in \mathcal{P}$. In other words, the *RB solution* $P_N(\mu)$ is a Galerkin approximation of the *reference solution* $P_\mathcal{N}(\mu)$. The optimality condition for (18) gives rise to the variational formulation: find $P_N(\mu) \in Q_N$ such that

$$a_\mathcal{N}(P_N(\mu), q; \mu) = b_\mathcal{N}(q; \mu) \qquad \text{for all } q \in Q_N, \tag{19}$$

with the same forms $a_\mathcal{N}$ and $b_\mathcal{N}$ as above. The RB approximation (19) now leads to a $N \times N$ linear system for each μ. It bears some similarity to that of Haasdonk and Ohlberger [5] for a convection-diffusion problem with a finite volume reference scheme, but in our context the derivation is more straightforward. Note that, anyhow, the RB solution $P_N(\mu)$ does not result from a cellwise mass balance and therefore cannot be associated with edgewise fluxes.

In order to assess the reliability of this RB method, it is essential to be able to work out explicitly computable bounds for the error $e_N(\mu) := P_N(\mu) - P_\mathcal{N}(\mu)$. Following the *a posteriori* approach, we introduce the residue $\mathscr{R}_\mathcal{N}[P_N(\mu)]$ associated to the RB solution as the continuous linear form that sends every $q \in Q_\mathcal{N}$ to the real number

$$\langle \mathscr{R}_\mathcal{N}[P_N(\mu)], q \rangle_{Q'_\mathcal{N} \times Q_\mathcal{N}} = a_\mathcal{N}(P_N(\mu), q; \mu) - b_\mathcal{N}(q; \mu). \tag{20}$$

Although continuity is obvious in finite dimension, for actual computations it is capital to choose a norm well suited to the problem. We recommend to equip the high-resolution space $Q_{\mathcal{N}}$ with the discrete energy norm

$$\|q\|_{1,\mathcal{N},\mu^*} = \{a_{\mathcal{N}}(q,q;\mu^*)\}^{1/2} \tag{21}$$

at a fixed parameter value μ^*. This enables us to introduce the residual dual norm

$$\|\mathscr{R}_{\mathcal{N}}[P_N(\mu)]\|_{-1,\mathcal{N},\mu^*} = \sup_{q\in Q_{\mathcal{N}}\setminus\{0\}} \frac{\langle \mathscr{R}_{\mathcal{N}}[P_N(\mu)], q\rangle}{\|q\|_{1,\mathcal{N},\mu^*}}. \tag{22}$$

Proposition 1 *The residual dual norm is connected to the true error by*

$$\|e_N(\mu)\|_{1,\mathcal{N},\mu^*} \le \frac{1}{\alpha_{\mathcal{N}}(\mu)} \|\mathscr{R}_{\mathcal{N}}[P_N(\mu)]\|_{-1,\mathcal{N},\mu^*} =: \Delta_N(\mu), \tag{23}$$

where

$$\alpha_{\mathcal{N}}(\mu) = \inf_{q\in Q_{\mathcal{N}}\setminus\{0\}} \frac{a_{\mathcal{N}}(q,q;\mu)}{\|q\|^2_{1,\mathcal{N},\mu^*}}. \tag{24}$$

Proof Subtraction of $0 = a_{\mathcal{N}}(P_{\mathcal{N}}(\mu), q;\mu) - b_{\mathcal{N}}(q;\mu)$ from (20) yields

$$a_{\mathcal{N}}(P_N(\mu) - P_{\mathcal{N}}(\mu), q;\mu) = \langle \mathscr{R}_{\mathcal{N}}[P_N(\mu)], q\rangle, \qquad \text{for all } q \in Q_{\mathcal{N}}.$$

Then, inequality (23) follows from the coercivity of $a_{\mathcal{N}}$ and definition (22). □

Thanks to finite dimensionality, the residual dual norm (22) can be computed explicitly by maximizing the linear objective function $q \mapsto \langle \mathscr{R}_{\mathcal{N}}[P_N(\mu)], q\rangle$ under the quadratic constraint $\|q\|^2_{1,\mathcal{N},\mu^*} = 1$ in $\mathbb{R}^{\mathcal{N}}$. The question remains as to: (1) how the sample $\mathcal{P}_N$ should be selected; (2) how to efficiently assemble the $N \times N$ matrix of the RB problem (19) for each μ without having to return into $\mathbb{R}^{\mathcal{N}}$.

4 Offline-Online, Greedy Algorithm and Empirical Interpolation

To address the first issue, we consider a *train* sample $\mathcal{P}_{\mathrm{T}} = \{\mu^{\mathrm{T}}_1, \ldots, \mu^{\mathrm{T}}_{\mathfrak{N}}\}$ consisting of $\mathfrak{N}$ equidistributed points in $\mathcal{P}$ (assumed to be an interval of $\mathbb{R}$). Offline, we perform an incremental optimization using a greedy algorithm: if $\mathcal{P}_\ell = \{\mu_1, \ldots, \mu_\ell\}$ is known, then we enlarge $\mathcal{P}_{\ell+1} = \mathcal{P}_\ell \cup \{\mu_{\ell+1}\}$ by inserting the worst-case parameter

$$\mu_{\ell+1} = \arg\max_{\mu\in\mathcal{P}_{\mathrm{T}}} \|e_\ell(\mu)\|_{1,\mathcal{N},\mu^*}. \tag{25}$$

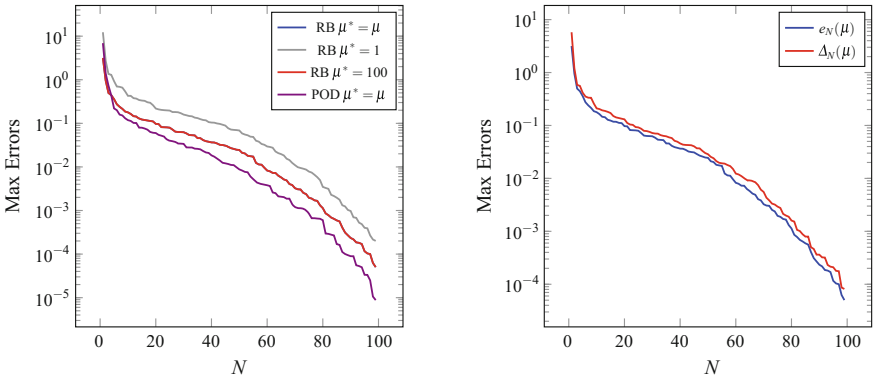


Fig. 1 Convergence of the greedy algorithm. *Left* $\max_{\mu\in\mathcal{P}_T} \|e_N(\mu)\|_{1,\mathcal{N},\mu^*}$ versus N for different values of μ^*. *Right* $\max_{\mu\in\mathcal{P}_T} \Delta_N(\mu)$ versus N for $\mu^* = 100$

In the left panel of Fig. 1, we report the convergence history of this algorithm for different values of μ^*, with $d = 2$, $\mathcal{P} = [1, 100]$, $\mathfrak{N} = 100$ and k coming from the 85th layer of the SPE10 benchmark [3]. A comparison is also carried out with the POD technique, known to be optimal for the L^2 topology. We observe that the RB greedy algorithm performs well here; the reduced bases Q_N obtained yield fast decaying RB approximation errors when N increases. The right panel of Fig. 1 shows that the error estimate $\Delta_N(\mu)$ defined in Proposition 1 is very sharp and hence can be safely used to certify the quality of RB approximations.

To address the second issue, we recall that the efficiency of RB methods crucially relies on the availability of an affine parametric dependence of $a_\mathcal{N}$ and $b_\mathcal{N}$. In our model, this assumption is unfortunately not fulfilled. Following [1], we resort to the empirical interpolation method (EIM) to approximate $\{\lambda_K(\mu)\}_{K\in\mathcal{M}}$ by

$$\lambda_K^M(\mu) = \sum_{m=1}^{M} \Theta_m(\mu)\widetilde{\lambda}_K^m, \qquad K \in \mathcal{M}, \tag{26}$$

where $\Theta_m(\mu) : \mathcal{P} \to \mathbb{R}$ does not depend on K and $\widetilde{\lambda}_K^m$ does not depend on μ. The EIM-RB problem is then: find $P_N^M(\mu) \in Q_N$ such that

$$a_\mathcal{N}^M(P_N^M(\mu), q; \mu) = b_\mathcal{N}^M(q; \mu), \qquad \text{for all } q \in Q_N, \tag{27}$$

where, for $(p, q) \in Q_N \times Q_N$,

$$a_\mathcal{N}^M(p, q; \mu) = \sum_{m=1}^{M} \Theta_m(\mu)\widetilde{a}_\mathcal{N}^m(p, q), \qquad b_\mathcal{N}^M(q; \mu) = \sum_{m=1}^{M} \Theta_m(\mu)\widetilde{b}_\mathcal{N}^m(q). \tag{28}$$

The elementary bilinear form

$$\widetilde{a}^m_{\mathcal{N}}(p,q) = \sum_{\substack{\sigma\in\mathcal{E}\\ \sigma=K|L}} k_\sigma \widetilde{\lambda}^m_\sigma \frac{|\sigma|}{\mathrm{d}_\sigma}(p_K - p_L)(q_K - q_L) + \sum_{\substack{\sigma\in\mathcal{E}\\ \sigma\subset\Gamma_{\mathrm{D}}\cap\partial K}} k_\sigma \widetilde{\lambda}^m_\sigma \frac{|\sigma|}{\mathrm{d}_\sigma} p_K q_K \tag{29}$$

is defined using the elementary edge mobility

$$\widetilde{\lambda}^m_\sigma = \begin{cases} \frac{1}{2}\left(\widetilde{\lambda}^m_K + \widetilde{\lambda}^m_L\right) & \text{if } \sigma = K|L, \\ \widetilde{\lambda}^m_K & \text{if } \sigma \subset \Gamma_{\mathrm{D}} \cap \partial K. \end{cases} \tag{30}$$

The elementary linear form $\widetilde{b}^m_{\mathcal{N}}(\cdot,\cdot)$ is defined similarly. The quantities $\widetilde{a}^m_{\mathcal{N}}(q_j, q_i)$ and $\widetilde{b}^m_{\mathcal{N}}(q_j, q_i)$ in a chosen basis $(q_1, \ldots, q_N)$ of Q_N can be pre-computed offline.

The left panel of Fig. 2 displays the relative cumulated error

$$\epsilon_{N,M,\max,\mathrm{rel}} = \frac{\max_{\mu\in\mathcal{P}_{\mathrm{T}}} \| P_{\mathcal{N}}(\mu) - P^M_N(\mu)\|_{1,\mathcal{N},\mu^*}}{\max_{\mu\in\mathcal{P}_{\mathrm{T}}} \| P_{\mathcal{N}}(\mu)\|_{1,\mathcal{N},\mu^*}} \tag{31}$$

as a function of N and M. We observe that the RB approximation converges rapidly. It is a difficult task to control the cumulated error $P_{\mathcal{N}} - P^M_N$, since one needs to combine adequately an error estimate (23) for the affine part and a suitable indicator for the EI error. To investigate the influence of the EI error on the reliability of RB approximations, we introduce the new residue

$$\langle \mathscr{R}^M_{\mathcal{N}}[P^M_N(\mu)], q\rangle_{Q'_{\mathcal{N}}\times Q_{\mathcal{N}}} = a^M_{\mathcal{N}}(P^M_N(\mu), q; \mu) - b^M_{\mathcal{N}}(q, \mu).$$

In the right panel of Fig. 2, we plot the evolution of

$$\Delta_{N,M,\max,\mathrm{rel}} = \frac{\max_{\mu\in\mathcal{P}_{\mathrm{T}}} \|\mathscr{R}^M_{\mathcal{N}}[P^M_N(\mu)]\|_{-1,\mathcal{N},\mu^*}}{\max_{\mu\in\mathcal{P}_{\mathrm{T}}} \| P_{\mathcal{N}}(\mu)\|_{1,\mathcal{N},\mu^*}}. \tag{32}$$

For small M, the bound (32) underestimates the truth error: the EI error cannot be neglected. Increasing M makes the EIM more accurate and the bound more sharp.

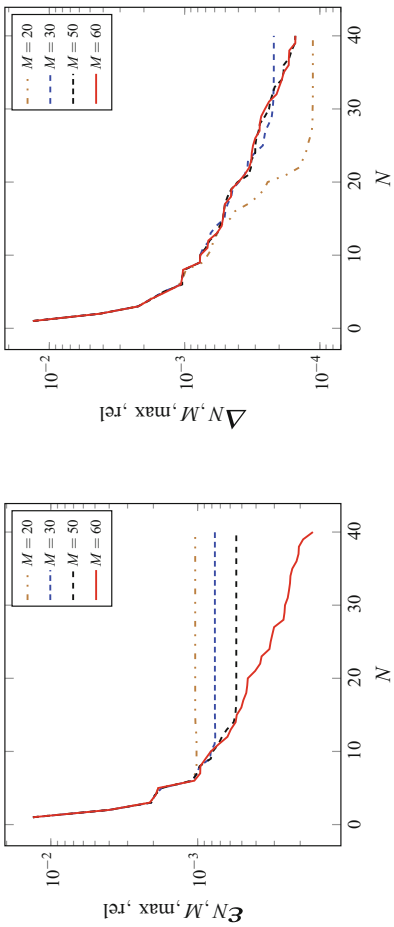


Fig. 2 Convergence of the error at the final time between the EIM-RB and the FV approximation for different values of M. *Left* exact error $\epsilon_{N,M,\max,\mathrm{rel}}$ versus N. *Right* error bound $\Delta_{N,M,\max,\mathrm{rel}}$ versus N

5 Conclusion

Similar results are obtained when considering uncertainty on the permeability. The next development is to build a reduced-basis model for the pressure problem at each time step of the simulation. The RB approximations thus obtained will be used as surrogates to efficiently update the velocity needed for the saturation equation.

References

1. Barrault, M., Maday, Y., Nguyen, N.C., Patera, A.T.: An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equations. C. R. Acad. Sci. Paris Ser. I **339**(9), 667–672 (2004)
2. Chaturantabut, S., Sorensen, D.C.: Application of pod and deim on dimension reduction of non-linear miscible viscous fingering in porous media. MCMDS **17**(4), 337–353 (2011)
3. Christie, M.A., Blunt, M.J.: Tenth SPE comparative solution project: a comparison of upscaling techniques. In: SPE Reservoir Simulation Symposium, 11–14 February, pp. 308–317. Houston, Texas, Society of Petroleum Engineers, (2001). SPE-66599-MS
4. Eymard, R., Gallouët, T., Herbin, R.: Finite volume methods. In: Ciarlet, P.G., Lions, J.L. (eds.) Techniques of Scientific Computing (Part 3), Handbook of Numerical Analysis, vol. VII, pp. 713–1018. North-Holland, Elsevier, Amsterdam (2000)
5. Haasdonk, B., Ohlberger, M.: Reduced basis method for finite volume approximations of parametrized linear evolution equations. ESAIM M2AN **42**(2), 277–302 (2008)

On the Capillary Pressure in Basin Modeling

Laurent Quaglia

Abstract This paper is devoted to the numerical simulation of sedimentary basins using a two phase flow Darcy model. The main objective is to improve the prediction of the position of the oil reservoirs and of the quantity of oil trapped in these reservoirs by modifying the dependency of the capillary pressure on the saturation, and by an adequate discretization.

Keywords Basin modeling · Capillary pressure

1 Introduction

This paper is devoted to the modeling and the numerical simulation of sedimentary basins. The main objective is to obtain a correct prediction of the position of the oil reservoirs and of the quantity of oil trapped in these reservoirs.

A sedimentary basin is a porous medium formed of a set of geological layers composed of sediments accumulated over several million years. In the source rock, the kerogen (which is an organic matter) turns into hydrocarbon. A part of these hydrocarbons stays in the source rock (and in particular yields the now famous shale gas); another part is expelled from the source rock and migrates towards the surface, essentially by buoyancy (and pressure gradient). Along the way, they can be trapped in some zones and this gives rise to the oil reservoirs which are exploited in the so called conventional oil extraction industry. This accumulation of oil can be due to the existence of impermeable rocks (generally clays or evaporites) but also to some capillary effects, known as the capillary barriers, which is the object of the present study.

In the modeling of these phenomena, the fluid is often considered as composed by two phases (or components): oil and water. The saturation of a phase is the ratio

L. Quaglia (✉)
Aix Marseille Univ, I2M, Marseille, France
e-mail: laurent.quaglia13@gmail.com

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_51

of the pore volume occupied by the phase, and each phase is submitted to a pressure. The difference of the pressures of these two phases is the capillary pressure.

Experimental data give some values for this capillary pressure, essentially as a constant value for each rock type. The dependency of this capillary pressure with respect to the saturation is not well known, in particular for very small or very large saturations, and unfortunately, when the stationary state is reached (and this is the state we want to compute) the water saturation is close to 0 under the capillary barrier and close to 1 over the capillary barrier.

Another difficulty is that these experimental data are obtained at a small scale (using rocks in laboratories) and the way to deduce the capillary pressure at the scale of a basin is not clear.

Different models are used for practical simulations. A first quite recent model (see [2]) is an invasion-percolation model; it has the advantage of being cheap from the computational point of view. It does not take into account the permeability of the rock and does not compute the time evolution of the migration of the oil. It essentially takes into account the gravity effects, the pressure field and the capillary effect. It predicts the way used by the hydrocarbons to reach the trapped zone and often gives a good prediction of this trapped zone, that is a prediction which seems close to the observed trapped zones for some known reservoirs or, at least, close to the expected zones given by geological experts. A second model uses a more precise description of the oil through the so called Darcy law for a two phases flow [1, 3]. A comparison of these two models is given, for instance, in the thesis of Sylvie Pegaz-Fiornet [4]. However, when the two phase flow model is used with a capillary pressure taken as piecewise constant for each rock type, it does not yield reasonable results. In fact, it is quite easy to see that with a constant capillary pressure for each rock type, the two phase flow model is not convenient since it does not allow the two phases to cross simultaneously the interface between two domains with a different capillary pressure function (in the model, the pressure of the phase has to be continuous as long as the corresponding relative permeability is positive). Our objective is to find some choice of the capillary pressure as a function of the saturation which is in accordance with the experimental data and gives the expected results in practical simulations of simple cases. A first possibility is to use some power laws for this capillary pressure function, but, in order to obtain reasonably good results, one has to use exponents in the power laws so high that the resulting non linear problem is too difficult to solve. We present other choices of these capillary pressure functions, which are piecewise linear and in agreement with the experimental data (that is, essentially, the constant values known for each type of rock), and which give, coupled with a convenient discretization, satisfactory results.

We now present the two phase Darcy model considered in this work. Each phase (water and oil) is supposed to be incompressible and the phases are immiscible. Then, the mass conservation of each phase reads

$$\partial_t(\phi u) + div(\overrightarrow{v_o}) = 0, \quad \partial_t(\phi(1-u)) + div(\overrightarrow{v_w}) = 0.$$

The quantity ϕ is the porosity of the medium, we will take it as a constant given value. The oil saturation, denoted by u, is the main unknown of the model. It depends of the space variable and of the time variable. The water saturation is equal to $(1-u)$. The quantities $\overrightarrow{v_o}$ and $\overrightarrow{v_w}$ are the filtration velocities. They are given using the Darcy law for a two phase flow,

$$\partial_t(\phi u) - div(\frac{K k_{r,o}(u)}{\mu_o}(\overrightarrow{\nabla} p_o - \rho_o \overrightarrow{g})) = 0, \tag{1}$$

$$-\partial_t(\phi u) - div(\frac{K k_{r,w}(u)}{\mu_w}(\overrightarrow{\nabla} p_w - \rho_w \overrightarrow{g})) = 0. \tag{2}$$

In these equations, K is the permeability tensor, μ_w is the dynamic viscosity of the water, μ_o that of the oil, ρ_w the density of the water and ρ_o the density of the oil, $\overrightarrow{g}$ is the gravity vector. All these quantities are assumed to be constant and given. The quantities $k_{r,o}$ and $k_{r,w}$ are the relative permeabilities. They are given functions of u. The pressure of the two phases are p_w and p_o. The difference of these pressure is the capillary pressure:

$$p_o - p_w = \pi(u), \tag{3}$$

where π is an increasing function representing the capillary pressure. This function depends on the rock type and the choice of this function is the main purpose of this paper.

The unknowns of the model are u, p_w and p_0. Of course, Eqs. (1), (2), (3) must be supplemented with convenient boundary conditions and initial condition. These conditions will be given for the tests below.

2 The Power Law Capillary Pressure

Some numerical codes use a capillary pressure of the form $\pi(u) = \beta + \gamma \left(\frac{u-u_{res}}{u_{irr}-u_{res}}\right)^n$ where the parameters β, γ, u_{res}, u_{irr} and n depend on the rock type. The domain of study consists of various rock types, and therefore it is formed of different subdomains, each with a given set of values of these parameters.

In fact, as we said in the introduction, the justification of this choice is not so clear. Experiments in laboratories essentially give a mean value for the capillary pressure and some values for u_{res} and u_{irr} (note that $k_{r,o}(u) = 0$ for $u < u_{res}$ and $k_{r,w}(u) = 0$ for $u > u_{irr}$, u_{res} is close to 0 and u_{irr} is close to 1). The extrapolation of this quantities at the basin scale is not obvious. A main arbitrary choice is the choice of n, it is done essentially in order to obtain "realistic" trapped zones. This is more or less obtained with very large values of n, but this choice leads to large numerical difficulties. The main problem is that for reasonable values of n (for which the code

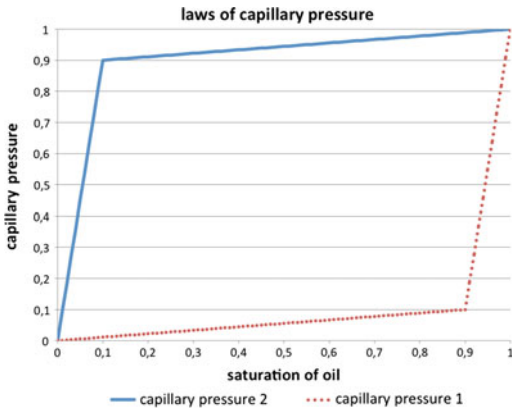


Fig. 1 Capillary pressures as functions of oil saturation, 1st-choice

is able to perform computations) the height of the trapped zone is clearly largely underestimated by the numerical code.

In the next section we present another approach for the choice of the capillary pressure functions.

3 Piecewise Linear Capillary Pressure Functions

We consider in this section a 1D-model and we will take the following values: $\overrightarrow{g} = 1$, $\rho_w = 1$, $\rho_o = 0.8$, $\phi = 1$, $K = 1$, $\mu_w = 1$, $\mu_o = 4$, $k_{r,o}(u) = u^2$ and $k_{r,w}(u) = (1-u)^2$.

The domain of the simulation is composed of two subdomains and each subdomain has a law for the capillary pressure. We present below two possible choices of piecewise linear capillary pressure functions. These capillary pressure curves are in agreement with the experimental data and, in some way, the simplest curves which can reproduce the expected oil trapping effect.

The first choice, denoted "1st-choice" below, is defined as follows (π_i is the law for the subdomain Ω_i), see Fig. 1,

$$\pi_1(u) = \alpha u + \beta_1 \text{ if } u \geq 1 - \varepsilon_c, \quad \pi_1(u) = \delta u + \beta_1' \text{ if } u < 1 - \varepsilon_c,$$
$$\pi_2(u) = \alpha u + \beta_2 \text{ if } u \leq \varepsilon_c, \quad \pi_2(u) = \delta u + \beta_2' \text{ if } u > \varepsilon_c.$$

We take $u_{res} = 0$ and $u_{irr} = 1$ for simplification, but there is no difficulty to take into account realistic values. Experimental data give values for β_i' and more or less for δ. These data give $\beta_2' > \beta_1'$. The choice of the positive number ε_c is arbitrary. It can be viewed as an alternative to the choice of n in the previous section. Once the choice of ε_c is done, the choice of β_i and α is given by the fact that π is a continuous

function, chosen to be piecewise linear and such that $\pi_1(0) = \pi_2(0)$, $\pi_1(1) = \pi_2(1)$. The parameter ε_c is small (so that α is very much larger that δ, say, for instance $\alpha = 1000\delta$).

We now describe the discretization of the model using a Finite Volume Method (FVM) including a modification of the flux function (this modification is called "shift method" below) which allows to obtain the expected trapped zone (at the end of the simulation).

Thanks to this modification, the main result (namely the height of the trapped zone) essentially does not depend on the choice of ε_c provided that ε_c is small enough (say $\varepsilon_c < 0.1$).

The discretization of the equations is made with a classical upwind Finite Volume scheme, with a 2 points discretization of the pressure gradient. It is important to notice that the scheme is fully implicit and some Newton iterations are used at each time step to compute the solution. We only describe the "shift method" which seems useful to obtain the expected height of trapped oil under a capillary barrier.

In the cases described in this section, it is possible to compute the exact solution and this computation gives that the height H of the expected trapped zone is given by the formula $\pi_2(1 - \varepsilon_c) - \pi_1(\varepsilon_c) = g(\rho_w - \rho_o)H$.

A main difficulty encountered with the numerical simulation of this 1D model is that we need to allow the fact that both phases flow through the interface between the two rock types (if not, generally too much oil is trapped). This is not possible if $\varepsilon_c = 0$ and $\beta_2' > \beta_1' + \alpha$ (which is a frequent case) because the capillary pressure has to be continuous at this interface when the two phases flow through the interface. This is the reason of the introduction of $\varepsilon_c > 0$. But, with this modification, the risk is that not enough oil is trapped and, in order to obtain a correct height of the trapped zone, we introduce a modification of the capillary pressure functions. We recall that u is close to 1 under the capillary barrier and u is close to 0 over the capillary barrier (the interface between the two rock type).

Then, the modification of the functions π_i in done for some particular values of u. It reads as follow, where $\bar{u}$ is the oil saturation under the capillary barrier and u_i the oil saturation in the cell i:

Algorithm 0.1 Modification of the function π in the shift method

$$\text{If } \max\{1 - 3\varepsilon_c, \bar{u} - \varepsilon_c\} < u_i < 1 - \varepsilon_c \text{ then } \pi_1(u_i) = \alpha u_i + \beta_1,$$
$$\text{If } \max\{1 - 3\varepsilon_c, \varepsilon_c\} < u_i < 1 + \varepsilon_c - \bar{u} \text{ then } \pi_2(u_i) = \alpha u_i + \beta_2',$$

The test is carried out by injecting oil from below (corresponding to $x = 0$ in Fig. 2) in the domain $0 < x < 1$ initially full of water. Since the phases are incompressible, the total flow is the same at $x = 1$ (than at $x = 0$), then, the flow of each phase is obtained thanks to the upwinding used for the saturation in the FV scheme. In Fig. 2, we give the water saturation obtained at the end of the simulation, that is

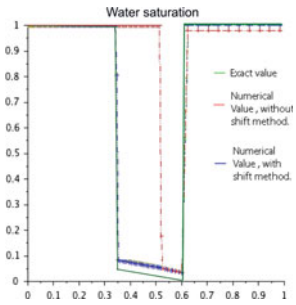


Fig. 2 Water saturation, 1st-choice

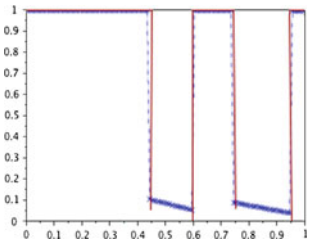


Fig. 3 Water saturation, 1st-choice, two barriers

when the stationary state is reached. The numerical result with the shift method is in blue. The numerical result without the shift method is in red. The exact solution is in green. Without the shift method, the height of the trapped zone (corresponding to the values of the water saturation, $(1-u)$, close to 0) is much smaller than that the exact one. With the shift method, the height is close to the exact one. In Fig. 2, the capillary barrier is at point $x = 0.6$. The simulation uses 90 cells under the barrier, 10 cells over the barrier and 60000 time steps.

In Fig. 3, we give an example with two capillary barriers. Here also, the stationary state is clearly accurately obtained with the shift method. Boundary conditions and initial condition are the same as above. Since there are two barriers, one obtains two trapped zones. Here also, we have a good accordance between exact solution (in red) and numerical solution (in blue). The capillary barriers are at points 0.6 and 0.95. The simulation uses 100 cells and 60000 time steps.

We present now a second choice of the piecewise linear capillary pressure functions, denoted as "2nd-choice" below. It gives the expected height of the capillary, without the shift method. In the case of two rock types, it reads (see Fig. 4)

$$\pi_1(u) = \delta u + \beta_1' \text{ if } u \leq 1-\varepsilon_c, \quad \pi_1(u) = \gamma_1 u + \delta_1 \text{ if } 1-\varepsilon_c < u,$$
$$\pi_2(u) = \delta u + \beta_2' \text{ if } u \leq 1-\varepsilon_c, \quad \pi_2(u) = \gamma_2 u + \delta_2 \text{ if } 1-\varepsilon_c < u,$$
$$\text{with } \gamma_1, \gamma_2, \delta_1, \delta_2 \text{ such as } \pi_1(1) = \pi_2(1) = \pi_2(1-\varepsilon_c) + \frac{1}{\varepsilon_c}$$

This choice was suggested to us by Robert Eymard. The principle of this new choice is also in order to allow the two phases to cross simultaneously the interface between

two domains with a different capillary pressure function. This is done by a convenient choice of the capillary pressure functions near $u = 1$. It is interesting to notice that, with this choice of the capillary pressure functions, the saturations (of oil) are close to 1 under and over the barrier while the two phases flow trough the interface between the two rock types. Then, the saturation of oil tends to 0 over the barrier (as time tends to $+\infty$) when the hydrostatic equilibrium is achieved under the barrier. With the same conditions that in Fig. 2, we obtain, with the new capillary pressure functions, the rights values (Fig. 5).

2D Simulation

In 2D simulations, with a non trivial geometry, a new difficulty occurs with the choice of the capillary pressure functions given in Fig. 1, that is with the 1st-choice. Due to the presence of horizontal flows, the height of the trapped zone is not always the expected one (we obtain too much accumulation of oil). A solution to this new problem is to temporarily disconnect the horizontal flows (in other words, they are not taken into account for some time steps) but this method is not completely satisfactory, because it is difficult to estimate when this disconnection has to be done. With the

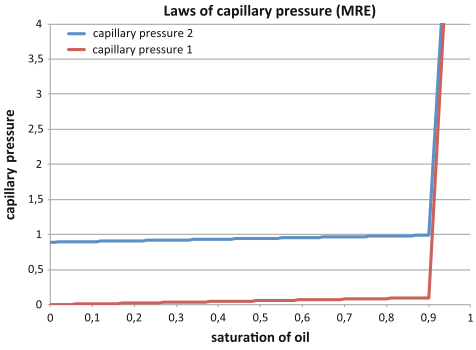


Fig. 4 Capillary pressures functions of oil saturation, 2nd-choice

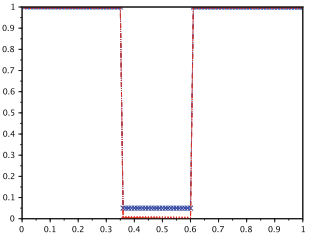


Fig. 5 water saturation, 2nd-choice

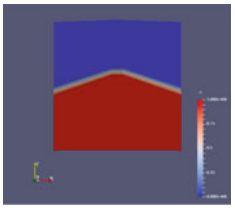


Fig. 6 Geometry

Fig. 7 Numerical solution

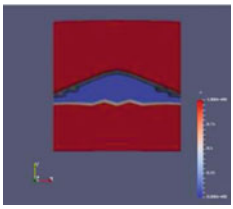
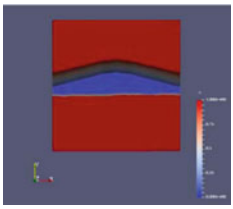


Fig. 8 Exact solution



2nd-choice (Fig. 4), we obtain a very good numerical solution, without using the disconnection of the horizontal flows, at least when the oil is injected at the bottom of the domain during a limited time (this is a realistic case). We present below a result with this 2nd-choice. One has two rock types, see Fig. 6. An injection of oil is made at the bottom of a domain during a short time. There are no fluxes on the lateral sides of the domain. At the top of the domain the fluid is allowed to flow. For the initial condition, we take $u = 1$. As usual in industrial codes for such a model, the discretization of the model is performed with a mesh whose interfaces are following the geometry of the geological layers. In Fig. 7 is the numerical solution obtained with the 2nd-choice and, finally, in Fig. 8 we give the exact solution. In Figs. 7 and 8, the water saturation is in red and the oil saturation in blue. The results are obtained with 400 cells and 60000 time steps.

4 Conclusion

The study of the heights of hydrocarbon trapping zones in a sedimentary basin is still under way and the optimal solution still under study. However, it has been observed that the choice of piecewise linear capillary pressures functions together with a convenient discretization give the expected results in one space dimension (1st-choice with shift method or 2nd-choice). For two dimensional problems, currently, we prefer to use the 2nd-choice which allows us to retrieve the expected results.

References

1. Cancès, C.: Two-phase flows in heterogeneous porous media: modeling and analysis of the flows of the effects involved by the discontinuities of the capillary pressure. Université de Provence - Aix-Marseille I, Theses (2008)
2. Carruthers, D.: Modeling of secondary petroleum migration using invasion percolation techniques. Discovery Series (2003)
3. Faille, I., Thibaut, M., Cacas, M.C., Havé, P., Willien, F., Wolf, S., Agélas, L., Pegaz-Fiornet, S.: Modeling fluid flow in faulted basins. Oil Gas Sci. Technol. Revue d'IFP Energies Nouvelles **69**(4), 529–553 (2014). doi:10.2516/ogst/2013204
4. Pegaz-Fiornet, S.: Study of hydrocarbon migration models for basin simulators. Theses, Université d'Aix-Marseille (2011). https://tel.archives-ouvertes.fr/tel-01451247

A Finite Volume Scheme for Nernst-Planck-Poisson Systems with Ion Size and Solvation Effects

Jürgen Fuhrmann and Clemens Guhlke

Abstract We introduce a recent model of an isothermal, incompressible mixture of ionic species with finite ion size and solvation effects. A two point flux finite volume ansatz on unstructured meshes is chosen to discretize the model. Based on a reformulation of the continuous problem in terms of absolute activities, the Scharfetter-Gummel upwind scheme is generalized to take into account finite ion size and solvation effects in a thermodynamically consistent manner.

Keywords Finite volume scheme · Nernst-Planck equations

MSC (2010): 65N08 78A57

1 A Generalized Nernst-Planck-Poisson System

Regard an isothermal, incompressible mixture of $N + 1$ species characterized by charge numbers z_i, molar densities ("concentrations") c_i, molar chemical potentials μ_i and molar volumes v_i. If the reference component $i = 0$ is electroneutral ($z_0 = 0$) it is regarded as a solvent, and each ion of species $i = 1 \dots N$ is allowed to be surrounded by a solvation shell consisting of κ_i solvent molecules.

The Nernst-Planck-Poisson system defines the motion of charged species due to convection in a barycentric velocity field $\mathbf{v}$ and due to gradients of the electrostatic potential ϕ and the respective chemical potentials μ_i while maintaining a self-consistent electric field [1, 2, 5, 6]:

J. Fuhrmann (✉) · C. Guhlke
Weierstrass Institute, Mohrenstraße 39, 10117 Berlin, Germany
e-mail: juergen.fuhrmann@wias-berlin.de

C. Guhlke
e-mail: clemens.guhlke@wias-berlin.de

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_52

$$-\nabla \cdot (\varepsilon_0 \varepsilon_r \nabla \phi) = F \sum_{i=0}^{N} z_i c_i \tag{1a}$$

$$\partial_t c_i + \nabla \cdot (c_i \mathbf{v} + \mathbf{N}_i) = 0 \qquad i = 1 \dots N \tag{1b}$$

$$\mathbf{N}_i = -(D_i/RT) c_i \left(\nabla \tilde{\mu}_i + \tilde{z}_i F \nabla \phi\right). \qquad i = 1 \dots N \tag{1c}$$

For $i = 1 \dots N$, $\tilde{m}_i := \frac{M_i}{M_0}$ are the molar mass ratios, $\tilde{\mu}_i := \mu_i - \tilde{m}_i \mu_0$ are effective chemical potentials [5] (or entropy variables, see [9]), and $\tilde{z}_i := z_i - \tilde{m}_i z_0$ are effective charge numbers with respect to the reference species. Further notations: D_i: species diffusion coefficients, R: molar gas constant, T: temperature, F: Faraday constant ε_0: vacuum dielectric permittivity, ε_r: relative dielectric permittivity.

The solvent molecules are split into two categories: the free solvent molecules with concentration c_0, and the solvent molecules located in the solvation shells of the ions with concentration $c_0^B = \sum_{i=1}^{N} \kappa_i c_i$ [1].

The molar density of the mixture $\bar{c} = \sum_{i=0}^{N} c_i$ is the sum of the molar densities of its components. The molar volumes ν_i and the solvation numbers κ_i may differ between the species, thus unlike in [2, 5], $\bar{c}$ exhibits spatial variations.

After introducing $\kappa_0 = 0$, for $i = 0 \dots N$, the molar volume of species i including the solvation shells is the effective molar volume $\hat{\nu}_i := \kappa_i \nu_0 + \nu_i$. The molar volume $\bar{\nu}$ of the mixture is the sum of the effective molar volumes of the species weighted by their molar fractions $\frac{c_i}{\bar{c}}$: $\bar{\nu} = \sum_{i=0}^{N} \hat{\nu}_i \frac{c_i}{\bar{c}}$. Incompressibility means $\bar{c}\bar{\nu} = 1$, or $\sum_{i=0}^{N} \hat{\nu}_i c_i = 1$.

The incompressibility constraint immediately leads to the limitation of the concentrations $c_i \le \frac{1}{\hat{\nu}_i}$. Therefore, this model prevents the overcrowding effect which is present in classical Nernst-Planck models which are based on the assumption of zero ion size. It allows to express the concentration c_0 of the reference species and the mixture concentration $\bar{c}$ directly from the concentrations $c_1 \dots c_N$:

$$c_0 = \frac{1}{\nu_0} - \sum_{i=1}^{N} \frac{\hat{\nu}_i}{\nu_0} c_i, \quad \bar{c} = \frac{1}{\nu_0} + \sum_{i=1}^{N} \frac{\nu_0 - \hat{\nu}_i}{\nu_0} c_i.$$

Due to the barycentric velocity setting, as in [8], the $N + 1$ mass diffusion fluxes $M_i \mathbf{N}_i$ ($i = 0, 1 \dots N$) sum up to zero, defining $\mathbf{N}_0 = -\sum_{i=1}^{N} \tilde{m}_i \mathbf{N}_i$.

The density of the mixture ρ can be expressed as

$$\rho = M_0 c_0 + \sum_{i=1}^{N} (\kappa_i M_0 + M_i) c_i = \frac{M_0}{\nu_0} + \sum_{i=1}^{N} \hat{M}_i c_i. \tag{2}$$

where $\hat{M}_i := M_i - M_0 \frac{\nu_i}{\nu_0}$ is a volume related effective molar mass. By comparing (2) to the incompressibility constraint it is easy to see but important to notice that incompressibility is not synonymous with constant density. Using for $i = 1 \dots N$ the volume related effective charge numbers $\hat{z}_i := z_i - z_0 \frac{\hat{\nu}_i}{\nu_0}$, the space charge q of the mixture is expressed as

$$q = F\sum_{i=0}^{N} z_i c_i = F\frac{z_0}{v_0} + F\sum_{i=1}^{N} \hat{z}_i c_i. \tag{3}$$

The evolution of the velocity field is described by the incompressible Navier–Stokes equations for the barycentric velocity $\mathbf{v}$ and the pressure p under a body force exerted by the self-consistent electric field. Throughout this paper, mechanical equilibrium [8] is assumed: $\mathbf{v} = 0$. The Navier–Stokes equations reduce to

$$\nabla p = -q\nabla\phi \tag{4}$$

$$\partial_t \rho = 0 \tag{5}$$

As discussed in [2], (4) describes the balance between the body force exerted by the electric field on the charged molecules in the mixture and the pressure gradient. As in [5], taking the divergence on both sides of (4) gives

$$-\Delta p = \nabla \cdot (q\nabla\phi). \tag{6}$$

It can be assumed that far from an electrode, the pressure p can be set equal to a fixed reference pressure p°. Equipped otherwise with Neumann boundary conditions derived from (4), (6) uniquely defines the pressure and up to a rotational part implies the force balance (4).

Equation (5) would be trivially fulfilled in the case of constant density. However, due to (2), the density depends on the local composition of the mixture and its time evolution. As a consequence, we have to assume either the stationary case $\partial_t c_i = 0$ or the equality of all species molar volumes and molar masses.

In order to close system (1), constitutive relationships between the chemical potentials $\mu_0 \dots \mu_N$ and the other quantities describing the system, in particular the concentrations, are introduced following the approach from [1]:

$$\mu_0 = \mu_0^\circ + v_0(p - p^\circ) + RT \ln \frac{c_0}{\bar{c}} \tag{7a}$$

$$\mu_i = \mu_i^\circ + v_i(p - p^\circ) + RT \ln \frac{c_i}{\bar{c}} - \kappa_i RT \ln \frac{c_0}{\bar{c}}. \qquad (i = 1 \dots N) \tag{7b}$$

Here, μ_i° are constant reference chemical potentials. In accordance to [1], μ_0 is the chemical potential of *all* solvent molecules including those in the solvation shells, and μ_i $(i = 1 \dots N)$ is the chemical potential of the *unsolvated* ions. The resulting effective chemical potential is

$$\tilde{\mu}_i = \mu_i - \frac{M_i}{M_0}\mu_0 = \tilde{\mu}_i^\circ + \tilde{v}_i(p - p^\circ) + RT \ln \frac{c_i}{\bar{c}} - \tilde{\kappa}_i RT \ln \frac{c_0}{\bar{c}} \tag{8}$$

where $\tilde{v}_i = v_i - \tilde{m}_i v_0$, $\tilde{\kappa}_i = \kappa_i + \tilde{m}_i$ and $\tilde{\mu}_i^\circ = \mu_i^\circ - \mu_0^\circ$.

2 Reformulation in Activities

In order to derive a flux expression which is easily handled numerically, we follow the approach introduced in [5] and reformulate the system in terms of (absolute effective) activities $a_i = \exp \frac{\tilde{\mu}_i}{RT}$. Denote by $\beta_i = \beta_i(a_1, a_2, \ldots, a_N, p)$ the (generalized absolute effective) *inverse activity coefficient* of species i characterized by $c_i = \beta_i a_i$.

The Nernst-Planck-Poisson system (1) transforms to

$$-\nabla \cdot \varepsilon_0 \varepsilon_r \nabla \phi = q = F\frac{z_0}{v_0} + F\sum_{i=1}^{N} \hat{z}_i \beta_i a_i \tag{9a}$$

$$\partial_t(\beta_i a_i) + \nabla \cdot (\mathbf{N}_i) = 0 \qquad i = 1 \ldots N \tag{9b}$$

$$\mathbf{N}_i = -D_i \beta_i \left(\nabla a_i + a_i \tilde{z}_i \frac{F}{RT} \nabla \phi \right). \qquad i = 1 \ldots N \tag{9c}$$

together with (4) and (5). As discussed above, we replace (4) by (6). In order to fulfill (5) we assume stationarity or equal molar masses and molar volumes.

One observes that under the time derivative and the divergence operator, expressions in the activities occur which are formally equal to the rather well understood classical Nernst-Planck case described by drift and Fickian diffusion, and which are multiplied by the inverse activity coefficients. This structure provides a rather straightforward way to generalize the Scharfetter-Gummel scheme. We note that a formulation in concentrations would lead to a cross diffusion structure which appears to be more complicated to handle compared to the additional nonlinear equations for β which just can be added to the overall nonlinear system of equations.

Equations (9a)–(9c) are the same as those proposed in [5]. All specifics of the model including differing molar volumes and solvation effects are expressed in the relationship defining β_i ($i = 1 \ldots N$). From (8) one obtains

$$a_i = a_i^\circ \exp \frac{\tilde{v}_i(p - p^\circ)}{RT} \frac{c_i}{\bar{c}} \left(\frac{c_0}{\bar{c}} \right)^{-\tilde{\kappa}_i},$$

where $a_i^\circ = \exp \mu_i^\circ$. For $i = 1 \ldots N$ this leads to

$$\begin{aligned} \exp \frac{\tilde{v}_i(p - p^\circ)}{RT} a_i^\circ \beta_i &= \bar{c} \left(\frac{c_0}{\bar{c}} \right)^{\tilde{\kappa}_i} = \bar{c}^{1-\tilde{\kappa}_i} c_0^{\tilde{\kappa}_i} \\ &= \frac{1}{v_0} \left(1 + \sum_{j=1}^{N} (v_0 - \hat{v}_j) a_j \beta_j \right)^{1-\tilde{\kappa}_i} \left(1 - \sum_{j=1}^{N} \hat{v}_j a_j \beta_j \right)^{\tilde{\kappa}_i}. \end{aligned}$$

As a result, one obtains a nonlinear system of N equations defining $\beta_1 \ldots \beta_N$ through the values of the pressure p and the activities $a_1 \ldots a_N$:

$$\beta_i = \mathscr{B}_i(p, a_1 \dots a_N, \beta_1 \dots \beta_N)$$
$$:= \frac{1}{a_i^\circ v_0} \exp\left(-\frac{\tilde{v}_i(p-p^\circ)}{RT}\right)\left(1+\sum_{j=1}^{N}(v_0-\hat{v}_j)a_j\beta_j\right)^{1-\tilde{\kappa}_i}\left(1-\sum_{j=1}^{N}\hat{v}_j a_j\beta_j\right)^{\tilde{\kappa}_i} \quad i=1\dots N. \tag{9d}$$

For the model regarded in [5] ($z_0 = 0, \kappa_i = 0, v_i = v_0 (i = 1 \dots N)$) it was possible to prove the existence and uniqueness of a solution of system (9d).

Thermodynamic equilibrium. Using the activity based flux expressions it is straightforward to derive expressions for the corresponding modified Poisson–Boltzmann equations describing thermodynamical equilibrium. Assuming zero flux due to thermodynamical equilibrium, one arrives at $\nabla\tilde{\mu}_i = -\tilde{z}_i F\nabla\phi$ $(i = 1 \dots N)$. To fulfill this, we introduce a constant electrochemical potential ψ_i and set $\tilde{\mu}_i = \tilde{z}_i F(\psi_i - \phi)$. The thermodynamical equilibrium then is described by the force balance (6), the algebraic system defining the inverse activity coefficients (9d) and

$$-\nabla\cdot\varepsilon_0\varepsilon_r\nabla\phi = F\frac{z_0}{v_0} + F\sum_{i=1}^{N}\hat{z}_i\beta_i a_i, \quad a_i = \exp\left(\frac{\tilde{z}_i F}{RT}(\psi_i-\phi)\right)(i=1\dots N). \tag{10}$$

Bikerman model. Assume that all species including the solvent have the same molar volume v_0, the same molar mass M_0 and the solvation number 0, and that the solvent is neutral. Then, for $i = 1 \dots N$, $\hat{z}_i = \tilde{z}_i = z_i, \tilde{m}_i = 1, \tilde{\kappa}_i = 1, \tilde{v}_i = 0, \hat{v}_i = v_0$. Then Eq. (9d) yields [5] $\beta_i = \frac{1}{v_0}\frac{1}{a_i^\circ + \sum_{j=1}^{N} a_j}$. Species and potential distributions can be obtained without the pressure which nevertheless is defined by (6).

3 Finite Volume Based Numerical Approach

Two point flux finite volume methods have structural advantages when considering coupled nonlinear problems and ensuring nonnegative and non-oscillatory solutions [3, 7, 12]. The Voronoï control volume method [10] is implemented in the framework pdelib [13] which is used to implement the proposed discretization approach.

The domain Ω is subdivided into a finite number of polygonal control volumes $K \in \mathscr{K}$ around the collocation points $\mathbf{x}_K$. Such a subdivision can be obtained by using a triangular or tetrahedral grid exhibiting the boundary-conforming Delaunay property [12]. The control volumes surrounding each given collocation point are obtained by joining the circumcenters of the simplices adjacent to it. For two neighboring control volumes, the grid edge $\mathbf{x}_L\mathbf{x}_K$ is orthogonal to the face separating the control volumes. This construction fits to the definition of an admissible grid in [4].

Discrete approximations $u_\mathscr{K} = \{u_K\}_{K\in\mathscr{K}}$ of a continuous function u are seen as piecewise constant functions which are constant in each control volume.

The discretization scheme described in the sequel will be based on the following set of degrees of freedom considered in each collocation point $\mathbf{x}_K$: electrostatic potential ϕ_K, pressure p_K, species activities $a_{1,K} \dots a_{N,K}$, inverse activity coefficients $\beta_{1,K} \dots \beta_{N,K}$. The overall number of unknowns in the discrete system is then $|\mathscr{K}|(2N+2)$.

Denote by ∂K the boundary of the control volume K, and by $|\xi|$, the measure (volume, surface, length) of a geometrical object ξ. Let $\sigma_{KL} = |\partial K \cap \partial L|$. Let $\mathscr{N}_K$ be the set of control volumes L such that $\sigma_{KL} > 0$. Let

$$q_K = F\frac{z_0}{\nu_0} + F\sum_{i=1}^{N} \hat{z}_i \beta_{i,K} a_{i,K} \qquad (K \in \mathscr{K}) \tag{11}$$

and $\tilde{Z}_i = \tilde{z}_i \frac{F}{RT}$. As in [5], one arrives at the finite volume discretization of the Poisson equation coupled to momentum balance and the discrete equation for the inverse activity coefficients:

$$\varepsilon_r \varepsilon_0 \sum_{L\in\mathscr{N}_K} \sigma_{KL} \frac{\phi_K - \phi_L}{|\mathbf{x}_K - \mathbf{x}_L|} = |K| q_K \quad (K \in \mathscr{K}) \tag{12a}$$

$$\sum_{L\in\mathscr{N}_K} \sigma_{KL} \frac{p_K - p_L}{|\mathbf{x}_K - \mathbf{x}_L|} = -\sum_{L\in\mathscr{N}_K} \sigma_{KL} \operatorname{avg}(q_K, q_L) \frac{\phi_K - \phi_L}{|\mathbf{x}_K - \mathbf{x}_L|} \quad (K \in \mathscr{K}) \tag{12b}$$

$$\beta_{i,K} = \mathscr{B}_i(p_K, a_{1,K} \dots a_{N,K}, \beta_{1,K} \dots \beta_{N,K}) \quad (i = 1 \dots N,\ K \in \mathscr{K}) \tag{12c}$$

In thermodynamic equilibrium, one assumes

$$a_{i,K} = \exp\left(\tilde{Z}_i(\psi_i - \phi_K)\right) \qquad (i = 1 \dots N,\ K \in \mathscr{K}). \tag{12d}$$

Here, $\operatorname{avg}(\cdot,\cdot)$ is some average expression, e.g. arithmetic average.

To discretize the nonstationary Poisson-Nernst-Planck system, abbreviate the subsystem (12a)–(12c) of system (12) by

$$\mathscr{P}(\phi_{\mathscr{K}}, p_{\mathscr{K}}, a_{1,\mathscr{K}} \dots a_{N,\mathscr{K}}, \beta_{1,\mathscr{K}} \dots \beta_{N,\mathscr{K}}) = 0. \tag{13}$$

Assume a subdivision of the time axis $0 = t^0 < t^1 < \dots < t^n < \dots$. For a continuous function u of space and time, let $u^n_{\mathscr{K}} = \{u^n_K\}_{K\in\mathscr{K}}$ denote the space-time approximation at t^n. As in [5], one chooses the unconditionally stable backward Euler method to discretize the continuity equation. For each timestep $n > 0$, this leads to a nonlinear system of equations defining the unknowns at the new time layer n from those on the old time layer $n-1$. The part concerning potential, pressure and inverse activity coefficients is described by (13). To establish the discrete extended Poisson-Nernst-Planck system, it is combined with a discrete analogue of the con-

tinuity equation and a relation defining species fluxes $N^n_{i,KL}$ between neighboring control volumes K and L replacing the equilibrium expression (12d):

$$|K|\frac{\beta^n_{i,K}a^n_{i,K} - \beta^{n-1}_{i,K}a^{n-1}_{i,K}}{t^n - t^{n-1}} - \sum_{L\in\mathcal{N}_K} \sigma_{KL}N^n_{i,KL} = 0 \quad (i = 1\ldots N, K \in \mathscr{K}) \tag{14a}$$

$$N^n_{i,KL} = D_i \,\mathrm{avg}(\beta^n_{i,K}, \beta^n_{i,L})\left(B\left(\tilde{Z}_i(\phi^n_L - \phi^n_K)\right)a^n_{i,K} - B\left(\tilde{Z}_i(\phi_K - \phi_L)\right)a^n_{i,L}\right) \tag{14b}$$

$$(i = 1\ldots N;\ K, L \in \mathscr{K} \text{ with } \sigma_{KL} > 0)$$

Here, $B(\xi) = \frac{\xi}{\exp(\xi)-1}$ is the Bernoulli function. The expression (14b) generalizes the Scharfetter-Gummel scheme [11]. The flux expression (14b) is consistent to the thermodynamic equilibrium, i.e. independent of the choice of the function $\mathrm{avg}(\cdot,\cdot)$, the solution of the discretized nonlinear Poisson system (12) is a solution of the discrete stationary Poisson-Nernst-Planck system (14) with zero fluxes [5].

4 Ionic Current Rectification in a Nanopore

In order to illustrate the capabilities of the finite volume method, we present simulation results for the dependency of the ionic current through a cylindrically symmetric nanopore with charged walls filled with a binary electrolyte of length 30 nm, bottom radius 1 nm, top radius 10 nm. The data have been inspired by [14] but modified in order to include the effect of finite ion sizes and solvation (with solvation number $\kappa = 15$). To resolve the boundary layer, a rectangular grid with graded coordinates is transformed to the desired trapezoidal shape of the pore using a numerical conformal mapping approach [15]. The resulting discretization grid is depicted in Fig. 1 left. Dirichlet boundary conditions for the concentrations at top and bottom are set to fixed values assuring a certain given molarity of the ionic reservoir.

At the top boundary, the electrostatic potential is fixed at 0V. Inhomogeneous Neumann boundary $\varepsilon\varepsilon_0\nabla\phi\cdot\mathbf{n} = \sigma$ together with homogeneous Neumann boundary conditions for the ionic species describe the charged wall with surface charge density σ. The potential at the bottom boundary is varied between -1 and 1 V, and stationary solutions are obtained resulting the corresponding ionic current plotted in Fig. 2. For large reservoir concentrations, the overcrowding of the boundary layer under the assumption of zero ion diameters and absence of solvation leads to an overestimation of the ionic current compared to the improved model presented in this paper. In-deep comparisons with other simulation approaches, and more thorough parameter studies are subject to forthcoming publications.

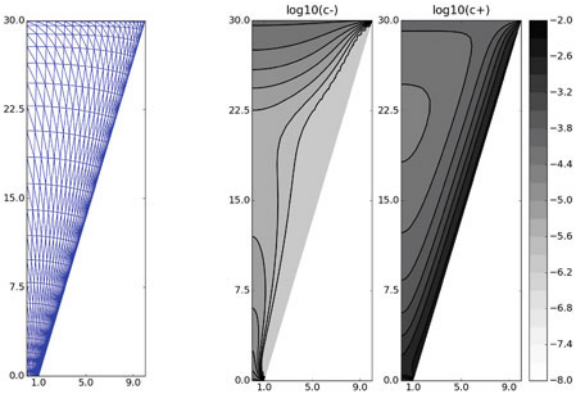


Fig. 1 Results for the example described in Sect. 4. *Left*: discretization grid. The grid has been refined close to the charged wall boundary in order to resolve the concentration boundary layers. *Right*: isolines of the logarithms of cation and anion concentrations for a voltage difference of 1 v between *top* and *bottom*. Positively charged anions are attracted by the charged wall, while the negatively charged cations are repelled

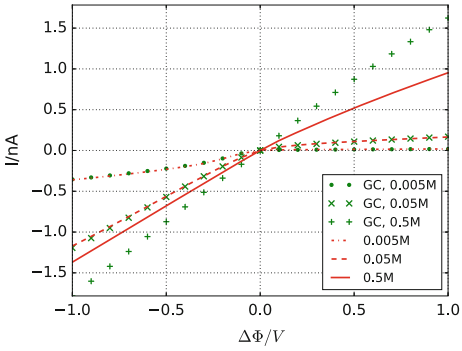


Fig. 2 IV curves for the presented model compared to those obtained with the Gouy-Chapman like model ("GC") assuming zero size ions for different molarities of the electrolytic solution

Acknowledgements This work was carried out in the framework of the project "Macroscopic Modeling of Transport and Reaction Processes in Magnesium-Air-Batteries" (Grant 03EK3027D) under the research initiative "Energy storage" of the German Federal government.

References

1. Dreyer, W., Guhlke, C., Landstorfer, M.: A mixture theory of electrolytes containing solvation effects. Electrochem. Comm. **43**, 7578 (2014)
2. Dreyer, W., Guhlke, C., Müller, R.: Overcoming the shortcomings of the Nernst-Planck model. Phys. Chem. Chem. Phys. **15**, 7075–7086 (2013)
3. Droniou, J.: Finite volume schemes for diffusion equations: Introduction to and review of modern methods. Math. Mod. Meth. Appl. Sci. **24**(08), 1575–1619 (2014)
4. Eymard, R., Gallouët, T., Herbin, R.: Finite volume methods. In: Handbook of Numerical Analysis, vol. VII, pp. 713–1020. North-Holland (2000)
5. Fuhrmann, J.: Comparison and numerical treatment of generalised NernstPlanck models. Comput. Phys. Commun. **196**, 166178 (2015)
6. Fuhrmann, J.: A numerical strategy for Nernst-Planck systems with solvation effect. Fuel cells, pp. 704 – 714 (2016)

7. Glitzky, A., Gärtner, K.: Energy estimates for continuous and discretized electro-reaction-diffusion systems. Nonlinear Anal. **70**(2), 788–805 (2009)
8. de Groot, S.R., Mazur, P.O.: Non-Equilibrium Thermodynamics. Dover Publications, New York (1962)
9. Jüngel, A.: The boundedness-by-entropy method for cross-diffusion systems. Nonlinearity **28**(6), 1963 (2015)
10. Macneal, R.H.: An asymmetrical finite difference network. Quart. Math. Appl. **11**, 295–310 (1953)
11. Scharfetter, D.L., Gummel, H.K.: Large signal analysis of a silicon read diode. IEEE Trans. Electron. Dev. **16**, 64–77 (1969)
12. Si, H., Gärtner, K., Fuhrmann, J.: Boundary conforming delaunay mesh generation. Comput. Math. Math. Phys. **50**, 38–53 (2010)
13. Streckenbach, T., Fuhrmann, J., et al.: Pdelib- a software toolbox for numerical computations (2017). http://www.pdelib.org. Accessed 01 Jan 2017
14. Th.Wolfram, M.: Forward and inverse solvers for electrodiffusion problems. Ph.D. thesis, Johannes-Kepler University, Linz (2008)
15. Trefethen, L.N.: Numerical computation of the Schwarz-Christoffel transformation. SIAM J. Sci. Stat. Comput. **1**(1), 82–102 (1980)

A Nonlinear Correction FV Scheme for Near-Well Regions

Vasiliy Kramarenko, Kirill Nikitin and Yuri Vassilevski

Abstract We present a finite volume method with improved well modelling for the subsurface flow simulation. The method is based on the nonlinear monotone finite volume scheme developed for diffusion, advection-diffusion and multiphase flow model equations with full anisotropic discontinuous permeability tensors on conformal polyhedral meshes. The new method uses the nonlinear (e.g. logarithmic) correction for the flux approximation in the near-well regions to utilize the singularity of the well-driven flow solution and improve accuracy of the pressure and the flux calculation. The method is applicable for anisotropic media, polyhedral grids, and different well cases including slanted, partially perforated or shifted from the grid cell center. Numerical experiments show the significant reduction of numerical errors compared to the original monotone nonlinear FV scheme with the conventional Peaceman well model or with the given analytical well rate.

Keywords Finite volume method · Improved well modelling · Nonlinear correction

MSC (2010): 35Q86 · 65M08 · 65N08

V. Kramarenko
Moscow Institute of Physics and Technology, 9 Institutskiy Per.,
Dolgoprudny, Moscow Region 141701, Russia
e-mail: kramarenko.vasiliy@gmail.com

K. Nikitin (✉) · Y. Vassilevski
Institute of Numerical Mathematics of Russian Academy of Sciences, 8 Gubkina Str.,
Moscow, Russia
e-mail: nikitin.kira@gmail.com

Y. Vassilevski
e-mail: yuri.vassilevski@gmail.com

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_53

1 Introduction

Cell-centered finite volume methods with nonlinear flux discretization on cell faces have proven to be an effective instrument for multiphase flow modelling and attract growing attention [5]. A monotone second order method with nonlinear two-point discretization of the diffusion and convection fluxes that preserves the non-negativity of the discrete solution was presented in [2]. The method was implemented for the two- and three-phase black oil models [11] on conformal hexahedral meshes, polyhedral meshes based on dynamic octrees [14] or dynamic octrees with cut cells. The scheme was later modified [1, 9] to a nonlinear multi-point scheme which satisfies the Discrete Maximum Principle (DMP). Benefits of using the DMP scheme for two-phase flows were discussed in [10].

The latest enhancement of the nonlinear method aims to incorporate well modelling into the finite volume framework. The well model is the sensitive part of the black-oil simulator and has the largest impact on all calculated well rates and breakthrough times. The solution in the near-well region is highly influenced by the singularity (e.g. logarithmic) of the well. The idea to use the solution singularity in the FV schemes was suggested in [3]. Later this approach was combined with the nonlinear FV method for the well-oriented prismatic grids with isotropic homogeneous and heterogeneous media [4]. Our new method generalizes these ideas for anisotropic media, arbitrary polyhedral grids and arbitrary wells adjusted neither with cells centers nor with edges [8].

The central idea of the method is to use a nonlinear correction for the reconstructed solution inside the nonlinear flux discretization scheme in the near-well region. For the isotropic case the linear-logarithmic reconstruction is used. The resulting method is exact on both linear and logarithmic solutions by construction and is generalized for the anisotropic case and for slanted wells. Numerical experiments show the significant reduction of the numerical errors compared to the original nonlinear FV scheme with the conventional Peaceman well model [12] or with the given analytical well rate.

2 Original FV Method

First we consider the stationary diffusion equation in order to introduce the numerical scheme and remind the basic ideas of the FV schemes construction.

Let Ω be a three-dimensional polyhedral domain with the Lipschitz boundary $\Gamma = \Gamma_N \cup \Gamma_D$. The diffusion equation for unknown pressure p with the Dirichlet or Neumann boundary conditions is written in the mixed form:

$$\begin{aligned} \mathbf{q} = -\mathbb{K}\nabla p, \qquad & \operatorname{div} \mathbf{q} = g \quad \text{in } \Omega, \\ & p = g_D \quad \text{on } \Gamma_D \\ & \mathbf{q} \cdot \mathbf{n} = 0 \quad \text{on } \Gamma_N. \end{aligned} \tag{1}$$

Here $\mathbb{K}(\mathbf{x})$ is a symmetric positive definite (possibly anisotropic) diffusion tensor, $g(\mathbf{x})$ is a source term, $g_D(\mathbf{x})$ is a given value on the Dirichlet part of the boundary Γ_D.

The cell-centered FV scheme uses one degree of freedom per cell T, p_T, collocated at cell barycenter $\mathbf{x}_T$. Integrating the mass balance Eq. (1) over T and using the divergence theorem, we obtain:

$$\sum_{f \in \partial T} \sigma_{T,f}\, q_f |f| = \int_T g \,\mathrm{d}x, \qquad q_f = \frac{1}{|f|} \int_f \mathbf{q} \cdot \mathbf{n}_f \,\mathrm{d}s, \tag{2}$$

where $q_f|f|$ is the normal flux across the face, $|f|$ is the area of face f, and $\sigma_{T,f}$ is either 1 or -1 depending on the mutual orientation of the unit normal vectors $\mathbf{n}_f$ and $\mathbf{n}_T$ ($\mathbf{n}_T$ denotes the outward normal vector for T).

Possible approaches for the flux (2) discretization include the nonlinear monotone two-point scheme [2] and the nonlinear DMP preserving compact multi-point scheme [1, 9]. In the next chapter we present a multi-point scheme designed for the near-well regions.

3 Near-Well Correction Scheme

Consider an isolated well which generates pressure singularity (see Fig. 1). The central idea of the nonlinear correction finite volume (NCFV) method is to select some region around the well and modify the FV scheme (following [3, 4]) to utilize the singularity and take into account the nonlinear component of the solution. In contrast to [4], our method is designed for anisotropic media, arbitrary polyhedral cells and arbitrary well location.

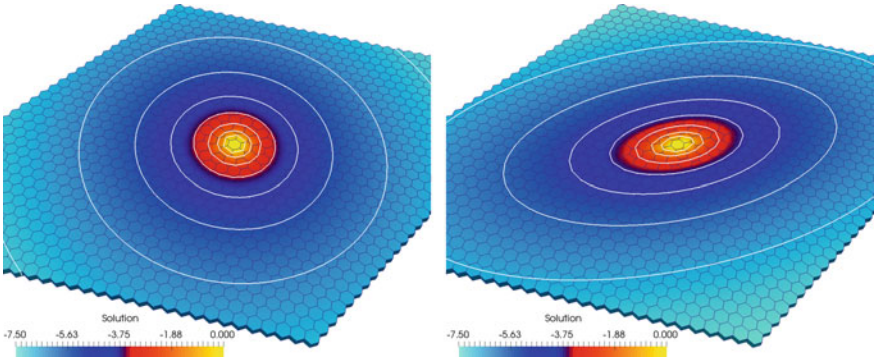


Fig. 1 Example of singularity in the near-well region: isotropic (*left*) and anisotropic (*right*) media

The original nonlinear FV method uses the piecewise linear reconstruction of the unknown field for flux calculation. The NCFV method takes into account the nonlinear component of the solution near the specific objects such as wells or large fractures.

We consider the pressure field to be the sum of the linear and nonlinear functions for each cell in a near-well region:

$$p_T = \underbrace{a\,x + b\,y + c\,z + d}_{p_{lin}} + \underbrace{e\,F(x,y,z)}_{p_F}, \tag{3}$$

where $F(x, y, z)$ is a function representing the singularity.

The finite volume discretization requires the mean value of the normal component of the flux $\mathbf{q} = -\mathbb{K}\nabla p$ to be calculated for each face f of T:

$$\int_f \mathbf{q}\cdot\mathbf{n}_f dS = -\int_f (\mathbb{K}\nabla p_T)\cdot\mathbf{n}_f dS = -\int_f (\mathbb{K}\nabla p_{lin})\cdot\mathbf{n}_f dS - \int_f (\mathbb{K}\nabla p_F)\cdot\mathbf{n}_f dS. \tag{4}$$

Since the method is derived for arbitrary grid cells and well direction, we consider the diagonal permeability tensor $\mathbb{K} = diag\left(k_x, k_y, k_z\right)$ for clarity. Using (3) for the integral (4) gives:

$$\begin{aligned} q_f = -\frac{1}{|f|}\int_f \mathbf{q}\cdot\mathbf{n}_f dS = ak_xS_{fx} + bk_yS_{fy} + ck_zS_{fz} + e\int_f (\mathbb{K}\nabla F(x,y,z))\cdot\mathbf{n}_f dS \\ = a\boldsymbol{\ell}_1 + b\boldsymbol{\ell}_2 + c\boldsymbol{\ell}_3 + e\boldsymbol{\ell}_4. \end{aligned} \tag{5}$$

The coefficients $\boldsymbol{\ell}_i$ depend solely on the mesh and problem data and are calculated explicitly, while the coefficients (a, b, c, e) are recovered from the solution in a set of neighboring cells.

Let T_+ and T_- be neighboring cells sharing a face f, and $\mathbf{x}_+, \mathbf{x}_-$ denote the centers of these cells. We take four points $\mathbf{x}_i$ ($\mathbf{x}_i \neq \mathbf{x}_+$) that denote centers of the neighboring cells or faces of T_+ and call four vectors $\mathbf{t}_i = \mathbf{x}_i - \mathbf{x}_+$ a *quadruplet*. The points are chosen as described below.

Considering the same representation (3) for vectors of quadruplet gives us:

$$\begin{pmatrix} p_1 - p_+ \\ p_2 - p_+ \\ p_3 - p_+ \\ p_4 - p_+ \end{pmatrix} = \begin{pmatrix} x_1 - x_+ & y_1 - y_+ & z_1 - z_+ & F_1 - F_+ \\ x_2 - x_+ & y_2 - y_+ & z_2 - z_+ & F_2 - F_+ \\ x_3 - x_+ & y_3 - y_+ & z_3 - z_+ & F_3 - F_+ \\ x_4 - x_+ & y_4 - y_+ & z_4 - z_+ & F_4 - F_+ \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ e \end{pmatrix}, \tag{6}$$

where $p_i = p(\mathbf{x}_i)$, $p_+ = p(\mathbf{x}_+)$ and $F_i = F(x_1, y_1, z_1)$.

From the set of admissible quadruplets we choose the one with the largest matrix (6) determinant. Solving it provides the coefficients a_+, b_+, c_+, e_+ for the cell T_+:

$$a_+ = \sum_j (p_j - p_+)\, m_{1,j}, \qquad b_+ = \sum_j (p_j - p_+)\, m_{2,j},$$
$$c_+ = \sum_j (p_j - p_+)\, m_{3,j}, \qquad e_+ = \sum_j (p_j - p_+)\, m_{4,j}, \tag{7}$$

where $m_{i,j}$ are the elements of the inverse matrix from (6). Taking T_- instead of T_+ and considering $-\mathbf{q}\cdot\mathbf{n}_f$ provides us the second flux approximation.

Applying (7) to Eq. (5) gives us:

$$q_+ = -\int_f \mathbf{q}\cdot\mathbf{n}_f dS = \Big[\ell_1 \sum_j (p_j - p_+)\, m^+_{1,j} + \ell_2 \sum_j (p_j - p_+)\, m^+_{2,j} + \ell_3 \sum_j (p_j - p_+)\, m^+_{3,j} + \ell_4 \sum_j (p_j - p_+)\, m^+_{4,j}\Big] = \tag{8}$$
$$\Big[\sum_j p_j \underbrace{\sum_i \ell_i\, m^+_{i,j}}_{k^+_j} - p_+ \sum_j \underbrace{\sum_i \ell_i\, m^+_{i,j}}_{k^+_j}\Big] = \Big(\sum_j k^+_j\,(p_j - p_+)\Big).$$

in similar way we get

$$q_- = -\Big(\sum_j k^+_j\,(p_j - p_-)\Big). \tag{9}$$

The resulting flux approximation is obtained as the weighted sum of q_+ and q_- with coefficients $\mu_+ + \mu_- = 1$. The weights can be chosen to ensure specific features of the solution. In our numerical experiments we considered $\mu_+ = \mu_- = 1/2$ which resulted in the *linear* multi-point flux discretization:

$$q_f = \mu_+ \Big(\sum_j k^+_j\,(p_j - p_+)\Big) + \mu_- \Big(\sum_{j'} k^-_{j'}\cdot(p_{j'} - p_-)\Big). \tag{10}$$

Note: Different cases of anisotropic media and non-trivial wells including slanted or partially perforated are handled by choosing an appropriate singularity function $F(x, y, z)$. For the anisotropic case a special F from [13] can be used, while for more complex cases one can implement techniques presented in [7]. For the wells not passing through the grid cell center we use two collocation points for the well cell (the one in the cell center and an additional point on the well), which provides one additional equation and allows to avoid using the conventional Peaceman formula for the well flux (see [8] for more details).

4 Numerical Experiments

Here we consider three numerical experiments for the near-well nonlinear correction scheme (NCFV) compared with the original monotone nonlinear FV scheme (NFV) with conventional Peaceman well model [12] or direct analytical flux to the well cell. For tests 1 and 3 the permeability tensor is scalar $\mathbb{K} = \mathbb{I}$ and test 2 deals with the anisotropic media. More general cases are presented in [8].

Defining the well pressure and flux gives us the analytical solution in the domain. For our experiments we put the Dirichlet conditions on the domain boundaries and use either given well pressure or given well flux from the analytical solution.

If the analytical rate for the well cell is given, we can compare the NFV scheme and the NCFV scheme without the influence of the well cell model. In this case we compute relative L^2-norms for the numerical pressure field errors of the NFV and NCFV schemes compared to known analytical solution: $err(p)_{NFV,anl}$ and $err(p)_{NCFV,anl}$, respectively.

If the well pressure is given, we use the numerical model for the well cell. Peaceman formula is applicable only for the cubic grids and is used with the NFV scheme, while the NCFV scheme is used for all experiments. In this case we compute relative L^2-norm for the pressure error for the NFV scheme + Peaceman ($err(p)_{NFV,pcm}$) and for the NCFV scheme ($err(p)_{NCFV}$), and the errors between the numerical well rate of the NFV and NCFV schemes and the analytical rate ($err(q)_{NFV}$ and $err(q)_{NCFV}$, respectively).

4.1 Test 1: Single Shifted Well, Hexagonal Prismatic Grid

For the first experiment we use one layer of the regular hexagonal prismatic grid. The well is shifted from the well cell centroid along the vector $v = (1, 1, 0)$ by the value $\alpha \cdot d/2$, where d is the cell diagonal length.

Table 1 shows the relative L^2-norms of pressure error for the NFV scheme with the analytical well cell rate and for the NCFV scheme.

Table 1 Solution relative errors for the NFV scheme and the near-well correction method for shifted well on hexagonal prismatic grid

α	$err(p)_{NFV,anl}$	$err(p)_{NCFV}$
0	1.1e-4	8.2e-11
0.1	1.4e-3	2.0e-11
0.3	4.2e-3	1.0e-11
0.5	7.1e-3	1.1e-11

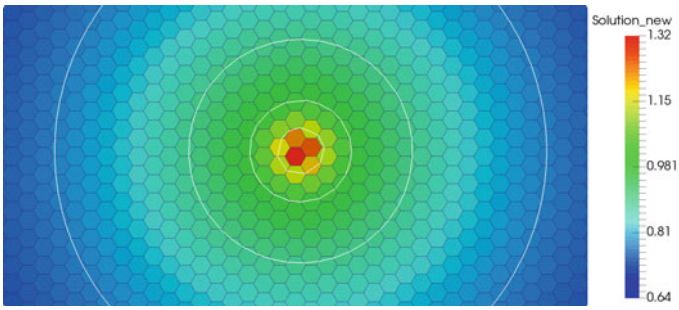


Fig. 2 Solution for the NCFV scheme for shifted well on hexagonal prismatic grid, $\alpha = 0.5$

Table 2 Solution error for the NFV and the NCFV, and the flux error for the NCFV scheme. 3D anisotropic case with the 60° slanted well

$err(p)_{NFV,anl}$	$err(p)_{NCFV}$	$err2(p)_{NFV,anl}$	$err2(p)_{NCFV}$	$err(q)_{NCFV}$
3.8e-6	2.5e-10	1.9e-2	1.2e-6	2.9e-5

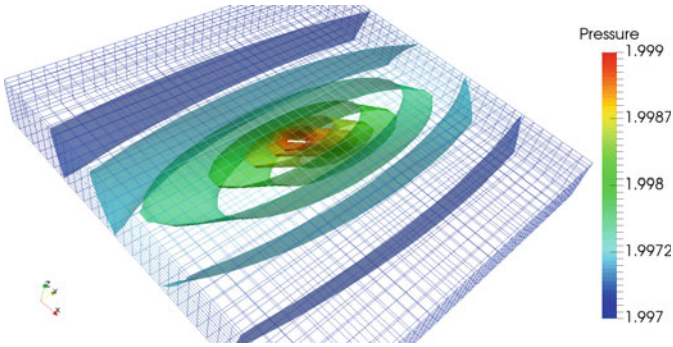


Fig. 3 Analytical solution for 3D anisotropic case with the 60° slanted well

Any well index based method incorporating the well within a single cell, will not provide the non-symmetric solution by construction. In contrast, the NCFV scheme can reproduce a non-symmetric solution (see Fig. 2).

4.2 Test 2: Slanted Well in 3D Anisotropic Media

Now we consider the slanted well in 3D rotated by 60° from the vertical. The tensor is diagonal anisotropic: $\mathbb{K} = diag\{10,\ 100,\ 1\}$. The orthogonal grid has 10 layers and Dirichlet boundary conditions are given for all boundaries (Fig. 3) .

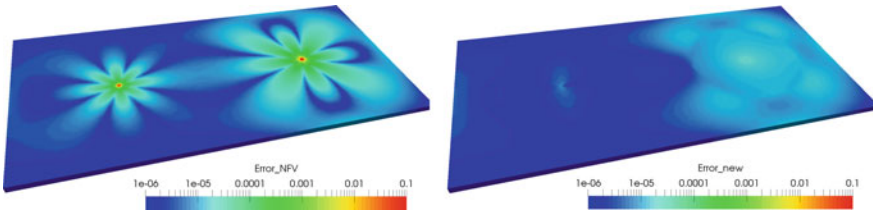


Fig. 4 Relative error for the NFV scheme with Peaceman well model (*left*) and the NCFV scheme (*right*) in the log-scale. Cubic grid $134 \times 67 \times 1$

The pressure and flux errors for this case are presented in Table 2. Due to the high anisotropy the solution variation is very small: $p \in [1.997,\ 1.999]$. To capture the error compared to this variation, we introduce $err2(p)_*$, which is the relative error normalized by $||p_{anl} - p_{anl,min}||$.

4.3 Test 3: Two Vertical Wells, Cubic Grid

The second experiment deals with two wells in the box domain with a cubic grid. The well rates are $q_1 = 1$, $q_2 = 4$ and the analytical solution suggested in [6] is defined by fixing the pressure in the middle point between two wells.

Table 3 shows the relative errors for the NFV and the NCFV scheme for the analytical well rates, relative errors for pressure and well rates (the first and the second well) for the numerical well models: NFV + Peaceman and the NCFV scheme.

Figure 4 presents the error fields for two methods in the log-scale. The NFV scheme reduces to the standard TPFA for this case and the cubic grid is ideal for the Peaceman method. The largest error of the NFV scheme is concentrated in regions around the wells that are covered by the near-well regions of the new method. The NCFV scheme gives considerably smaller errors than the conventional method.

5 Conclusion

We present the near-well nonlinear correction FV scheme applicable for the general case of anisotropic media, polyhedral grids and arbitrarily oriented wells including slanted, shifted and partially perforated cases.

Numerical experiments show the significant reduction of the numerical errors compared to the original monotone nonlinear FV scheme with the conventional Peaceman well model or with the given analytical well rate.

Acknowledgements This work has been supported in part by RFBR grants 15-35-20991, 17-01-00886, Russian President Grant MK-2951.2017.1, and ExxonMobil Upstream Research Company.

Table 3 Solution relative errors and flux errors for q_1 and q_2 for the problem with two wells

$100/h$	$err(p)_{NFV,anl}$	$err(p)_{NCFV,anl}$	$err(p)_{NFV,pcm}$	$err(p)_{NCFV}$	$err(q)_{NFV}$		$err(q)_{NCFV}$	
33	1.2e-2	2.8e-5	1.2e-2	2.8e-5	4.6e-3	1.9e-2	2.1e-5	4.1e-5
67	5.1e-3	7.0e-6	5.2e-3	7.6e-6	4.6e-3	1.9e-2	2.3e-5	5.4e-5
99	3.1e-3	3.2e-6	3.1e-3	4.1e-6	4.6e-3	1.8e-2	2.0e-5	7.0e-5

References

1. Chernyshenko, A., Vassilevski, Y.: A finite volume scheme with the discrete maximum principle for diffusion equations on polyhedral meshes. In: FVCA VII, pp. 197–205 (2014)
2. Danilov, A., Vassilevski, Y.: A monotone nonlinear finite volume method for diffusion equations on conformal polyhedral meshes. RJNAMM **24**(3), 207–227 (2009)
3. Ding, Y., Jeannin, L.: A new methodology for singularity modelling in flow simulations in reservoir engineering. Comput. Geosci. **5**, 93–119 (2001)
4. Dotlić, M., Vidović, D., Pokorni, B., Pušić, M., Dimkić, M.: Second-order accurate finite volume method for well-driven flows. J. Comp. Phys. **307**, 460–475 (2016)
5. Droniou, J.: Finite volume schemes for diffusion equations: introduction to and review of modern methods. Math. Models Methods Appl. Sci. **24**(8), 1575–1619 (2014)
6. Haitjema, H.M.: Analytic Element Modeling of Groundwater Flow. ClassPak Publishing (2005)
7. Korneev, A., Novikov, A., Posvyanskii, D., Posvyanskii, V.: An application of green's function technique for computing well inflow without radial flow assumption. In: ECMOR XV (2016)
8. Kramarenko, V., Nikitin, K., Vassilevski, Y.: Enhanced nonlinear finite volume scheme for multiphase flows. In: ECMOR XV (2016)
9. Lipnikov, K., Svyatskiy, D., Vassilevski, Y.: Minimal stencil finite volume scheme with the discrete maximum principle. Russ. J. Numer. Anal. Math. Modelling **27**(4), 369–385 (2012)
10. Nikitin, K., Novikov, K., Vassilevski, Y.: Nonlinear finite volume method with discrete maximum principle for the two-phase flow model. Lobachevskii J. Math. **37**(4) (2016)
11. Nikitin, K., Terekhov, K., Vassilevski, Y.: A monotone nonlinear finite volume method for diffusion equations and multiphase flows. Comp. Geosci. **18**(3), 311–324 (2014)
12. Peaceman, D.W.: Interpretation of well-block pressures in numerical reservoir simulation. SPEJ **18**(3), 183–194 (1978)
13. Peaceman, D.W.: Interpretation of well-block pressures in numerical reservoir simulation with non-square grid blocks and anisotropic permeability. SPEJ **23**(3), 531–543 (1983)
14. Terekhov, K., Vassilevski, Y.: Two-phase water flooding simulations on dynamic adaptive octree grids with two-point nonlinear fluxes. RJNAMM **28**(3), 267–288 (2013)

A Hybrid High-Order Method for the Convective Cahn–Hilliard Problem in Mixed Form

Florent Chave, Daniele A. Di Pietro and Fabien Marche

Abstract We propose a novel Hybrid High-Order method for the Cahn–Hilliard problem with convection. The proposed method is valid in two and three space dimensions, and it supports arbitrary approximation orders on general meshes containing polyhedral elements and nonmatching interfaces. An extensive numerical validation is presented, which shows robustness with respect to the Péclet number.

Keywords Hybrid high-order · Cahn–Hilliard equation · Phase separation · Mixed formulation · Polyhedral meshes · Arbitrary order

MSC (2010): 65N08 · 65N30 · 65N12

1 Cahn–Hilliard Equation

Let $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, denote a bounded connected convex polyhedral domain with Lipschitz boundary $\partial\Omega$ and outward normal $\mathbf{n}$, and let $t_F > 0$. The convective Cahn–Hilliard problem consists in finding the order-parameter $c : \Omega \times (0, t_F] \to \mathbb{R}$ and the chemical potential $w : \Omega \times (0, t_F] \to \mathbb{R}$ such that

The original version of the book was revised: Missed out corrections have been updated. The erratum to the book is available at https://doi.org/10.1007/978-3-319-57394-6_58

F. Chave (✉) · D.A. Di Pietro · F. Marche
University of Montpellier, Institut Montpelliérain Alexander Grothendieck,
34095 Montpellier, France
e-mail: florent.chave@umontpellier.fr

D.A. Di Pietro
e-mail: daniele.di-pietro@umontpellier.fr

F. Marche
e-mail: fabien.marche@umontpellier.fr

C. Cancés and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_54

$$\mathrm{d}_t c - \frac{1}{\mathrm{Pe}}\Delta w + \nabla\cdot(\mathbf{u}c) = 0 \qquad \text{in } \Omega\times(0,t_F] \tag{1a}$$

$$w = \Phi'(c) - \gamma^2\Delta c \qquad \text{in } \Omega\times(0,t_F] \tag{1b}$$

$$c(0) = c_0 \qquad \text{in } \Omega \tag{1c}$$

$$\partial_{\mathbf{n}} c = \partial_{\mathbf{n}} w = 0 \qquad \text{on } \partial\Omega\times(0,t_F] \tag{1d}$$

where $\gamma > 0$ is the interface parameter (usually taking small values), Pe>0 is the Péclet number, $\mathbf{u}$ the velocity field such that $\nabla\cdot\mathbf{u} = 0$ in Ω and Φ the free-energy such that $\Phi(c) := \frac{1}{4}(1-c^2)^2$. This formulation is an extension of the Cahn–Hilliard model originally introduced in [1, 2] and a first step towards coupling with the Navier–Stokes equations.

In this work we extend the HHO method of [3] to incorporate the convective term in (1a). Therein, a full stability and convergence analysis was carried out for the non-convective case, leading to optimal estimates in $(h^{k+1}+\tau)$ (with h denoting the meshsize and τ the time step) for the the $C^0(H^1)$-error on the order-parameter and $L^2(H^1)$-error on the chemical potential. The convective term is treated in the spirit of [4], where a HHO method fully robust with respect to the Péclet number was presented for a locally degenerate diffusion-advection-reaction problem.

The proposed method offers various assets: (i) fairly general meshes are supported including polyhedral elements and nonmatching interfaces; (ii) arbitrary polynomial orders, including the case $k = 0$, can be considered; (iii) when using a first-order (Newton-like) algorithm to solve the resulting system of nonlinear algebraic equations, element-based unknowns can be statically condensed at each iteration.

The rest of this paper is organized as follows: in Sect. 2, we recall discrete setting including notations and assumptions on meshes, define localy discrete operators and state the discrete formulation of (1). In Sect. 3, we provide an extensive numerical validation.

2 The Hybrid High-Order Method

In this section we recall some assumptions on the mesh, introduce the notation, and state the HHO discretization.

2.1 Discrete Setting

We consider sequences of refined meshes that are regular in the sense of [5, Chap. 1]. Each mesh $\mathcal{T}_h$ in the sequence is a finite collection $\{T\}$ of nonempty, disjoint, polyhedral elements such that $\overline{\Omega} = \bigcup_{T\in\mathcal{T}_h}\overline{T}$ and $h = \max_{T\in\mathcal{T}_h} h_T$ (with h_T the diameter

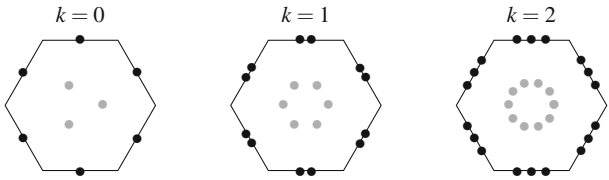


Fig. 1 Local DOF space for $k = 0, 1, 2$. Internal DOFs (in *gray*) can be statically condensed at each Newton iteration

of T). For all $T \in \mathcal{T}_h$, the boundary of T is decomposed into planar faces collected in the set $\mathcal{F}_T$. For admissible mesh sequences, $\mathrm{card}(\mathcal{F}_T)$ is bounded uniformly in h. Interfaces are collected in the set $\mathcal{F}_h^{\mathrm{i}}$, boundary faces in $\mathcal{F}_h^{\mathrm{b}}$ and we define $\mathcal{F}_h := \mathcal{F}_h^{\mathrm{i}} \cup \mathcal{F}_h^{\mathrm{b}}$. For all $T \in \mathcal{T}_h$ and all $F \in \mathcal{F}_T$, the diameter of F is denoted by h_F and the unit normal to F pointing out of T is denoted by $\mathbf{n}_{TF}$.

To discretize in time, we consider for sake of simplicity a uniform partition $(t^n)_{0\le n\le N}$ of the time interval $[0, t_F]$ with $t^0 = 0$, $t^N = t_F$ and $t^n - t^{n-1} = \tau$ for all $1 \le n \le N$. For any sufficiently regular function of time φ taking values in a vector space V, we denote by $\varphi^n \in V$ its value at discrete time t^n, and we introduce the backward differencing operator δ_t such that, for all $1 \le n \le N$,

$$\delta_t \varphi^n := \frac{\varphi^n - \varphi^{n-1}}{\tau} \in V.$$

2.2 Local Space of Degrees of Freedom

For any integer $l \ge 0$ and X a mesh element or face, we denote by $\mathbb{P}^l(X)$ the space spanned by the restrictions to X of d-variate polynomials of order l. Let

$$\underline{U}_h^k := \left(\times_{T\in\mathcal{T}_h} \mathbb{P}^{k+1}(T) \right) \times \left(\times_{F\in\mathcal{F}_h} \mathbb{P}^k(F) \right)$$

be the global degrees of freedoms (DOFs) space with single-valued interface unknowns. We denote by $\underline{v}_h = ((v_T)_{T\in\mathcal{T}_h}, (v_F)_{F\in\mathcal{F}_h})$ a generic element of $\underline{U}_h^k$ and by v_h the piecewise polynomial function such that $v_h|_T = v_T$ for all $T \in \mathcal{T}_h$. For any $T \in \mathcal{T}_h$, we denote by $\underline{U}_T^k$ and $\underline{v}_T = (v_T, (v_F)_{F\in\mathcal{F}_T})$ the restrictions to T of $\underline{U}_h^k$ and $\underline{v}_h$, respectively (Fig. 1) .

2.3 Local Diffusive Contribution

Consider a mesh element $T \in \mathcal{T}_h$. We define the local potential reconstruction $\mathrm{p}_T^{k+1} : \underline{U}_T^k \to \mathbb{P}^{k+1}(T)$ such that, for all $\underline{v}_T := (v_T, (v_F)_{F\in\mathcal{F}_T}) \in \underline{U}_T^k$ and all $z \in \mathbb{P}_T^{k+1}$,

$$(\nabla \mathrm{p}_T^{k+1}\underline{v}_T, \nabla z)_T = -(v_T, \Delta z)_T + \sum_{F\in\mathcal{F}_T} (v_F, \nabla z \cdot \mathbf{n}_{TF})_F,$$

with closure condition $\int_T (\mathrm{p}_T^{k+1}\underline{v}_T - v_T) = 0$. We introduce the local diffusive bilinear form a_T on $\underline{U}_T^k \times \underline{U}_T^k$ such that, for all $(\underline{u}_T, \underline{v}_T) \in \underline{U}_T^k \times \underline{U}_T^k$

$$a_T(\underline{u}_T, \underline{v}_T) := (\nabla \mathrm{p}_T^{k+1}\underline{u}_T, \nabla \mathrm{p}_T^{k+1}\underline{v}_T)_T + s_T(\underline{u}_T, \underline{v}_T),$$

with stabilization bilinear form $s_T : \underline{U}_T^k \times \underline{U}_T^k \to \mathbb{R}$ such that

$$s_T(\underline{u}_T, \underline{v}_T) := \sum_{F\in\mathcal{F}_T} h_F^{-1}(\pi_F^k(u_F - u_T), \pi_F^k(v_F - v_T))_F,$$

where, for all $F \in \mathcal{F}_h$, $\pi_F^k : L^1(F) \to \mathbb{P}^k(F)$ denotes the L^2-orthogonal projector onto $\mathbb{P}^k(F)$.

2.4 Local Convective Contribution

For any mesh element $T \in \mathcal{T}_h$, we define the local convective derivative reconstruction $\mathrm{G}_{\mathbf{u},T}^{k+1} : \underline{U}_T^k \to \mathbb{P}^{k+1}(T)$ such that, for all $\underline{v}_T := (v_T, (v_F)_{F\in\mathcal{F}_T}) \in \underline{U}_T^k$ and all $w \in \mathbb{P}^{k+1}(T)$,

$$(\mathrm{G}_{\mathbf{u},T}^{k+1}\underline{v}_T, w)_T = -(v_T, \mathbf{u} \cdot \nabla w)_T + \sum_{F\in\mathcal{F}_T} (v_F, (\mathbf{u} \cdot \mathbf{n}_{TF})w)_F.$$

The local convective contribution $b_{\mathbf{u},T}$ on $\underline{U}_T^k \times \underline{U}_T^k$ is such that, for all $(\underline{u}_T, \underline{v}_T) \in \underline{U}_T^k \times \underline{U}_T^k$

$$b_{\mathbf{u},T}(\underline{u}_T, \underline{v}_T) := -(u_T, \mathrm{G}_{\mathbf{u},T}^{k+1}\underline{v}_T)_T + s_{\mathbf{u},T}(\underline{u}_T, \underline{v}_T).$$

with local upwind stabilization bilinear form $s_{\mathbf{u},T} : \underline{U}_T^k \times \underline{U}_T^k \to \mathbb{R}$ such that

$$s_{\mathbf{u},T}(\underline{u}_T, \underline{v}_T) := \sum_{F\in\mathcal{F}_T} (\frac{|\mathbf{u} \cdot \mathbf{n}_{TF}| - \mathbf{u} \cdot \mathbf{n}_{TF}}{2}(u_F - u_T), v_F - v_T)_F.$$

Notice that the actual computation of $\mathrm{G}_{\mathbf{u},T}^{k+1}$ is not required, as one can simply use its definition to expand the cell-based term in the bilinear form $b_{\mathbf{u},T}$.

2.5 Discrete Problem

Denote by $\underline{U}_{h,0}^k := \{\underline{v}_h = ((v_T)_{T\in\mathcal{T}_h}, (v_F)_{F\in\mathcal{F}_h}) \in \underline{U}_h^k | \int_\Omega v_h = 0\}$ the zero-average DOFs subspace of $\underline{U}_h^k$. We define the global bilinear forms a_h and $b_{\mathbf{u},h}$ on $\underline{U}_h^k \times \underline{U}_h^k$ such that, for all $(\underline{u}_h, \underline{v}_h) \in \underline{U}_h^k \times \underline{U}_h^k$

$$a_h(\underline{u}_h, \underline{v}_h) := \sum_{T\in\mathcal{T}_h} a_T(\underline{u}_T, \underline{v}_T), \qquad b_{\mathbf{u},h}(\underline{u}_h, \underline{v}_h) := \sum_{T\in\mathcal{T}_h} b_{\mathbf{u},T}(\underline{u}_T, \underline{v}_T).$$

The discrete problem reads: For all $1 \le n \le N$, find $(\underline{c}_h^n, \underline{w}_h^n) \in \underline{U}_{h,0}^k \times \underline{U}_h^k$ such that

$$\begin{aligned}
&(\delta_t c_h^n, \varphi_h) + \frac{1}{\mathrm{Pe}} a_h(\underline{w}_h^n, \underline{\varphi}_h) + b_{\mathbf{u},h}(\underline{c}_h^n, \underline{\varphi}_h) = 0 && \forall \underline{\varphi}_h \in \underline{U}_h^k \\
&(w_h^n, \psi_h) = (\Phi'(c_h^n), \psi_h) + \gamma^2 a_h(\underline{c}_h^n, \underline{\psi}_h) && \forall \underline{\psi}_h \in \underline{U}_h^k
\end{aligned}$$

where $\underline{c}_h^0 \in \underline{U}_{h,0}^k$ solves $a_h(\underline{c}_h^0, \underline{\varphi}_h) = -(\Delta c_0, \varphi_h)$ for all $\underline{\varphi}_h \in \underline{U}_h^k$.

3 Numerical Test Cases

In this section, we numerically validate the HHO method.

3.1 Disturbance of the Steady Solution

For the first test case, we use a piecewise constant approximation ($k = 0$), discretize the domain $\Omega = (0, 1)^2$ by a triangular mesh ($h = 1.92 \cdot 10^{-3}$) with $\gamma = 5 \cdot 10^{-2}$, $\tau = \gamma^2$ and $\mathrm{Pe} = 1$. The initial condition for the order-parameter and the velocity field are given by

$$c_0(\mathbf{x}) := \tanh\left(\frac{2x_1 - 1}{2\sqrt{2}\gamma^2}\right), \quad \mathbf{u}(\mathbf{x}) := 20 \cdot \begin{pmatrix} x_1(x_1 - 1)(2x_2 - 1) \\ -x_2(x_2 - 1)(2x_1 - 1) \end{pmatrix}, \quad \forall \mathbf{x} \in \Omega.$$

The result is depicted in Fig. 2 and shows that the method is well-suited to capture the interface dynamics subject to a strong velocity fields.

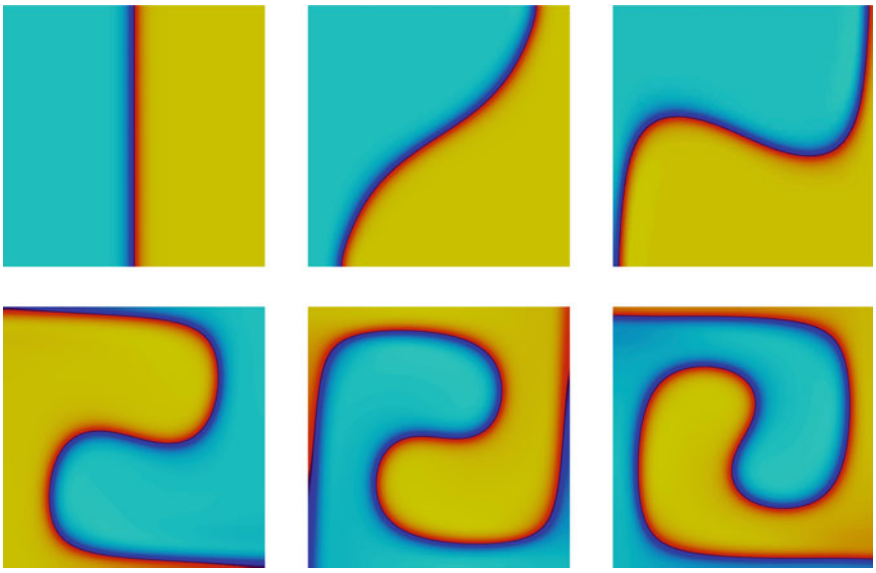


Fig. 2 Steady solution perturbed by a circular velocity field (*left* to *right*, *top* to *bottom*)

3.2 Thin Interface Between Phases

For the second example, we also use a piecewise constant approximation ($k = 0$) with a Cartesian discretization of the domain $\Omega = (0, 1)^2$, where $h = 1.95 \cdot 10^{-3}$. The interface parameter is taken to be very small $\gamma = 5 \cdot 10^{-3}$, the time step is $\tau = 1 \cdot 10^{-5}$ and $\mathrm{Pe} = 50$. The initial condition for the order-parameter is taken to be a random value between -1 and 1 inside a circular partition of the Cartesian mesh and -1 outside. The velocity field is given by

$$\mathbf{u}(\mathbf{x}) := \frac{1}{2}\left(1 + \tanh(80 - 200\|(x_1 - 0.5, x_2 - 0.5)\|_2)\right) \cdot \begin{pmatrix} 2x_2 - 1 \\ 1 - 2x_1 \end{pmatrix}, \quad \forall \mathbf{x} \in \Omega.$$

See Fig. 3 for the numerical result. The method is robust with respect to γ and is also well-suited to approach the thin high-gradient area of the order-parameter.

3.3 Effect of the Péclet Number

The Péclet number is the ratio of the contributions to mass transport by convection to those by diffusion: when Pe is greater than one, the effects of convection exceed

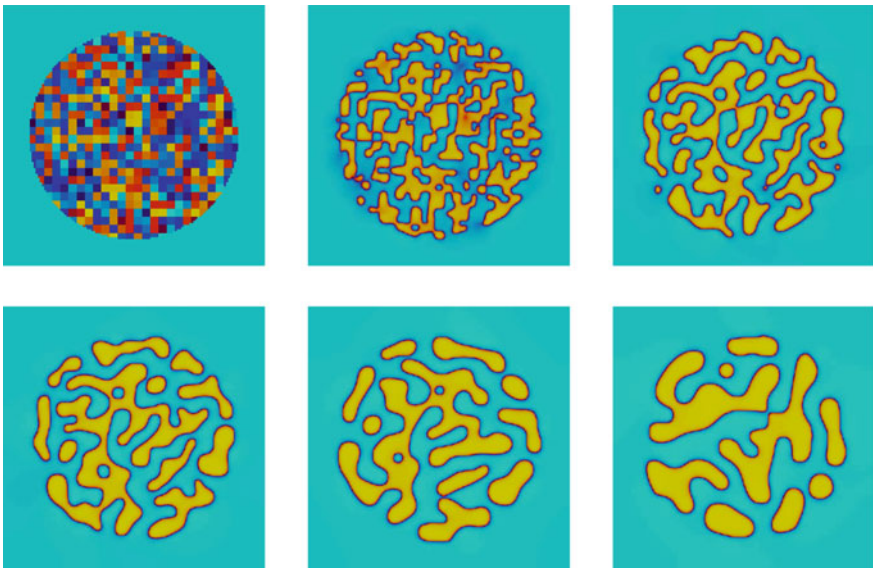


Fig. 3 Evolution of spinodal decomposition with thin interface (*left* to *right*, *top* to *bottom*)

those of diffusion in determining the overall mass flux. In the last test case, we compare several time evolutions obtained with different values of the Péclet number ($\mathrm{Pe} \in \{1, 50, 200\}$), starting from the same initial condition. We use a Voronoi discretization of the domain $\Omega = (0, 1)^2$, where $h = 9.09 \cdot 10^{-3}$, and use piecewise linear approximation ($k = 1$). We choose $\gamma = 1 \cdot 10^{-2}$, $\tau = 1 \cdot 10^{-4}$ and $t_F = 1$. The initial condition is given by a random value between -1 and 1 inside a circular domain of the Voronoi mesh and -1 outside. The convective term is given by

$$\mathbf{u}(\mathbf{x}) := \begin{pmatrix} \sin(\pi x_1)\cos(\pi x_2) \\ -\cos(\pi x_1)\sin(\pi x_2) \end{pmatrix}, \quad \forall \mathbf{x} \in \Omega.$$

Snapshots of the order parameter at several times are shown on Fig. 4 for each value of the Péclet number. For each case, the method takes into account the value of Pe and appropriately models the evolution of the order parameter by prevailing advection to diffusion when $\mathrm{Pe} \gg 1$.

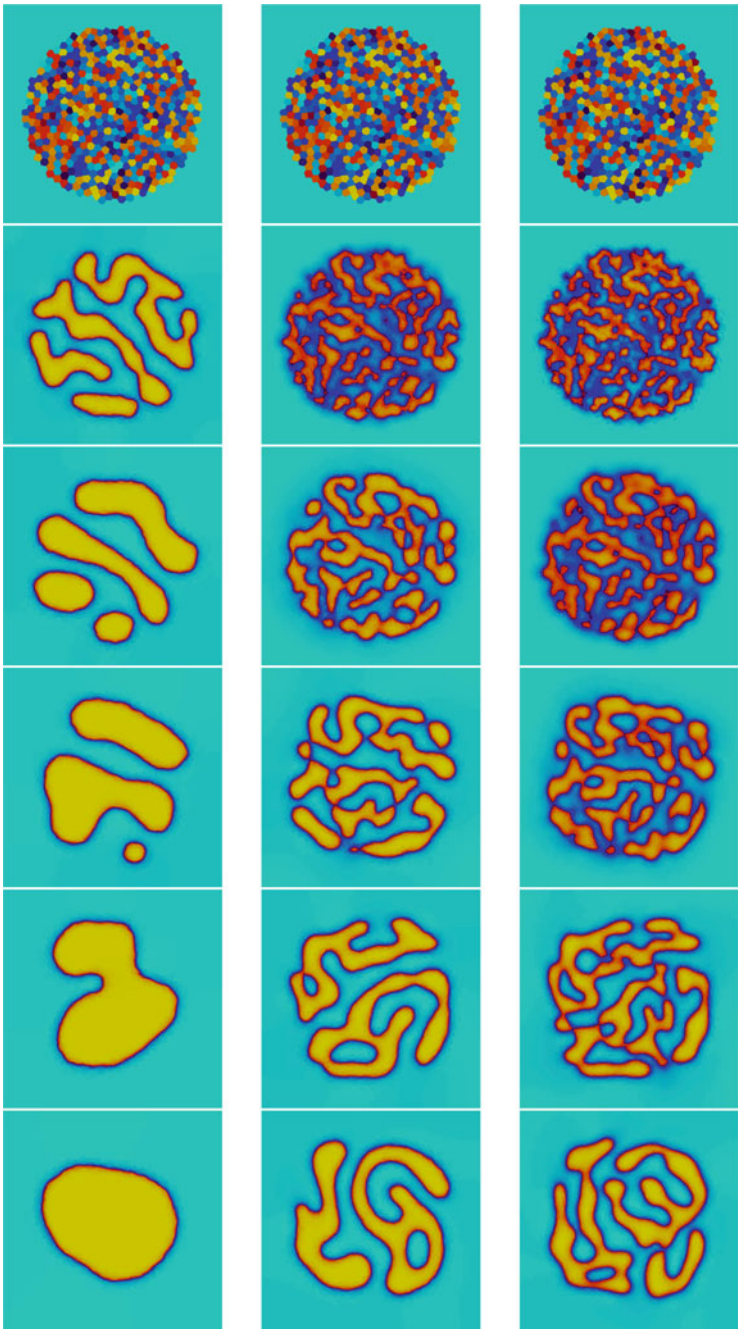


Fig. 4 Comparison at the same time between evolution of solutions with different Péclet number (*top* to *bottom*). *Left*: Pe = 1, *middle*: Pe = 50, *right*: Pe = 200. Displayed times are $t = 0, 1 \cdot 10^{-2}, 6 \cdot 10^{-2}, 2 \cdot 10^{-1}, 5 \cdot 10^{-1}, 1$

Acknowledgements The work of D. A. Di Pietro and F. Marche was partially supported by *Agence Nationale de la Recherche* grant HHOMM (ref. ANR-15-CE40-0005).

References

1. Cahn, J.W.: On spinoidal decomposition. Acta Metall. Mater. **9**, 795–801 (1961)
2. Cahn, J.W., Hilliard, J.E.: Free energy of a nonuniform system, I, interfacial free energy. J. Chem. Phys. **28**, 258–267 (1958)
3. Chave, F., Di Pietro, D.A., Marche, F., Pigeonneau, F.: A hybrid high-order method for the Cahn-Hilliard problem in mixed form. SIAM J. Numer. Anal. **54**(3), 1873–1898 (2016). doi:10.1137/15M1041055
4. Di Pietro, D.A., Droniou, J., Ern, A.: A discontinuous-skeletal method for advection-diffusion-reaction on general meshes. SIAM J. Numer. Anal. **53**(5), 2135–2157 (2015). doi:10.1137/140993971
5. Di Pietro, D.A., Ern, A.: Mathematical aspects of discontinuous Galerkin methods, Mathmatiques & Applications, vol. 69. Springer, Berlin (2012)

A Hybrid Finite Volume—Finite Element Method for Modeling Flows in Fractured Media

Alexey Chernyshenko, Maxim Olshahskii and Yuri Vassilevski

Abstract This work is devoted to the new hybrid method for solving a coupled system of advection–diffusion equations posed in a bulk domain and on an embedded surface. Systems of this kind arise in many engineering and natural science applications, but we consider the modeling of contaminant transport in fractured porous media as an example of an application. Fractures in a porous medium are considered as sharp interfaces between the surrounding bulk subdomains. The method is based on a monotone nonlinear finite volume scheme for equations posed in the bulk and a trace finite element method for equations posed on the surface. The surface is not fitted by the mesh and can cut through the background mesh in an arbitrary way. The background mesh is an octree grid with cubic cells. The surface intersects an octree grid and we get a polyhedral octree mesh with cut-cells. The numerical properties of the hybrid approach are illustrated in a series of numerical experiments with different embedded geometries. The method demonstrates great flexibility in handling curvilinear or branching embedded structures.

Keywords Finite volume method · TraceFEM · Bulk–surface coupled problems · Fractured porous media · Octree grid

MSC (2010): 65M60 · 65N08 · 65Q60

A. Chernyshenko (✉) · Y. Vassilevski
Institute of Numerical Mathematics, Gubkina 8, Moscow 119333, Russia
e-mail: chernyshenko.a@gmail.com

Y. Vassilevski
e-mail: yuri.vassilevski@gmail.com

M. Olshahskii
Department of Mathematics, University of Houston, Houston, TX 77204-3008, USA
e-mail: molshan@math.uh.edu

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_55

1 Introduction

At a recent time, there has been a growing interest in developing methods for the numerical treatment of systems of coupled bulk–surface PDEs. Different approaches can be distinguished depending on how the surface is recovered and equations are treated. If a tetrahedral tessellation of the volume is available that fits the surface, then it is natural to introduce finite element spaces in the volume and on the induced triangulation of the surface. Unfitted finite element methods allow the surface to cut through the background tetrahedral mesh. In the class of finite element methods also known as cutFEM, Nitsche-XFEM or TraceFEM, standard background finite element spaces are employed, while the integration is performed over cut domains and over the embedded surface [2]. The benefits of the unfitted approach are the efficiency in handling implicitly defined surfaces, complex geometries, and the flexibility in dealing with evolving domains. The hybrid method described in this paper belongs to the general class of unfitted methods.

If the finite element method is used for the bulk problem, then it is natural to consider a finite element method for surface PDE as well. However, other discretizations such as finite volume or finite difference methods can be preferred for the PDE posed in the volume.

This paper develops a numerical method based on the sharp-interface representation, which uses a FV-method to discretize the bulk PDE. Our goal is (i) to allow the surface to overlap with the background mesh in an arbitrary way, (ii) to avoid regular triangulating the surface, (iii) to avoid any extension of the surface PDE to the bulk domain. To achieve these goals, we combine the monotone (i.e. satisfying the discrete maximum principle) finite volume method on general meshes [4, 6] with the trace finite element method on octree meshes from [5]. In the octree TraceFEM one considers the bulk finite element space of piecewise trilinear continuous functions and further uses the restrictions (traces) of these functions to the surface. These traces are further used in a variational formulation of the surface PDE. Effectively, this results in the integration of the standard polynomial functions over the (reconstructed) surface. Only degrees of freedom from the cubic cells cut by the surface are active for the surface problem. Surface parametrization is not required, no surface mesh is built, no PDE extension of the surface is needed. The resulting hybrid FV–FE method is very robust with respect to the position of surfaces against the background mesh and is well suited for handling non-smooth surfaces and surfaces given implicitly.

While the present technique can be applied for tetrahedral or more general polyhedral tessellations of the bulk domain, we use octree grid with cubic cells here. The Cartesian structure and built-in hierarchy of octree grids makes mesh adaptation, reconstruction and data access fast and easy. However, an octree grid provides only the first order (staircase) approximation of a general geometry. Allowing the surface to cut through the octree grid in an arbitrary way overcomes this issue, but challenges us with the problem of building efficient bulk–surface discretizations.

We demonstrate that the hybrid TraceFEM–non-linear FV method complements the advantages of using octree grids by delivering the higher order accuracy for both bulk and surface numerical solutions.

2 Mathematical Model

Consider the bulk domain $\Omega \subset \mathbb{R}^3$ and a piecewise smooth surface $\Gamma \subset \Omega$. The surface Γ may have several connected components. If Γ has a boundary, for simplicity we assume that $\partial\Gamma \subset \partial\Omega$, but the model can be extended to immersed surfaces. Thus, we have the subdivision $\overline{\Omega} = \cup_{i=1,\dots,N}\overline{\Omega}_i$ into simply connected subdomains Ω_i such that $\overline{\Omega}_i \cap \overline{\Omega}_j \subset \Gamma$, $i \neq j$.

In each Ω_i, we assume a given Darcy velocity field of the fluid $\mathbf{w}_i(\mathbf{x})$, $\mathbf{x} \in \Omega_i$. By $\mathbf{w}_\Gamma(\mathbf{x})$, $\mathbf{x} \in \Omega_\Gamma$, we denote the velocity field tangential to Γ having the physical meaning of the flow rate through the cross-section of the fracture. Consider an agent that is soluble in the fluid and transported by the flow in the bulk and along the fractures. The fractures are modeled by the surface Γ. The solute volume concentration is denoted by u, $u_i = u|_{\Omega_i}$. The solute surface concentration along Γ is denoted by v. Change of the concentration happens due to convection by the velocity fields $\mathbf{w}_i$ and $\mathbf{w}_\Gamma$, diffusive fluxes in Ω_i, diffusive flux on Γ, as well as the fluid exchange and diffusion flux between the fractures and the porous matrix. These coupled processes can be modeled by the following system of equations [1], in subdomains,

$$\begin{cases} \phi_i \dfrac{\partial u_i}{\partial t} + \operatorname{div}(\mathbf{w}_i u_i - D_i \nabla u_i) = f_i \quad \text{in } \Omega_i, \\ \qquad\qquad\qquad\qquad\qquad u_i = v \quad \text{on } \partial\Omega_i \cap \Gamma, \end{cases} \tag{1}$$

and on the surface,

$$\phi_\Gamma \frac{\partial v}{\partial t} + \operatorname{div}_\Gamma(\mathbf{w}_\Gamma v - d D_\Gamma \nabla_\Gamma v) = F_\Gamma(u) + f_\Gamma \quad \text{on } \Gamma, \tag{2}$$

where we employ the following notations: ∇_Γ, $\operatorname{div}_\Gamma$ denote the surface tangential gradient and divergence operators; $F_\Gamma(u)$ stands for the net flux of the solute per surface area due to fluid leakage and hydrodynamic dispersion; f_i and f_Γ are given source terms in the subdomains and in the fracture; D_i denotes the diffusion tensor in the porous matrix; the surface diffusion tensor is D_Γ. Both D_i, $i = 1, \dots, N$, and D_Γ are symmetric and positive definite; $d > 0$ is the fracture width coefficient; $\phi_i > 0$ and $\phi_\Gamma > 0$ are the constant porosity coefficients for the bulk and the fracture.

The total surface flux $F_\Gamma(u)$ represents the contribution of the bulk to the solute transport in the fracture. The mass balance at Γ leads to the equation

$$F_\Gamma(u) = [-D\mathbf{n} \cdot \nabla u + (\mathbf{n} \cdot \mathbf{w})u]_\Gamma, \tag{3}$$

where $\mathbf{n}$ is a unit normal vector at Γ, $[w(\mathbf{x})]_\Gamma$ denotes the jump of w across Γ in the direction of $\mathbf{n}$.

If Γ is piecewise smooth, we need additional conditions on the edges, assuming the continuity of concentration, conservation of fluid mass and solute flux. Also we add Dirichlet's boundary conditions for the concentration u and v on $\partial\Omega_D$ and $\partial\Gamma_D$ and homogeneous Neumann's boundary conditions on $\partial\Omega_N$ and $\partial\Gamma_N$, respectively. Initial conditions are given by the known concentration u_0 and v_0 at $t = 0$.

3 Hybrid Finite Volume–Finite Element Method

To produce a grid with an octree hierarchical structure we assume a Cartesian background mesh with cubic cells and allow local refinement of the mesh by sequential division of any cubic cell into 8 cubic subcells. This mesh gives the tessellation $\mathscr{T}_h$ of the computational domain Ω, $\overline{\Omega} = \cup_{T\in\mathscr{T}_h}\overline{T}$. The surface $\Gamma \subset \Omega$ cuts through the mesh in an arbitrary way. For the purpose of numerical integration, instead of Γ we consider Γ_h, a given polygonal approximation of Γ. We assume that similar to Γ, the reconstructed surface Γ_h divides Ω into N subdomains $\Omega_{i,h}$, and $\partial\Gamma_h \subset \partial\Omega$. We do not imply any restrictions on how Γ_h intersects the background mesh. The reconstructed surface Γ_h is a $C^{0,1}$ surface that can be partitioned in planar triangular elements:

$$\Gamma_h = \bigcup_{K\in\mathscr{F}_h} K, \tag{4}$$

where $\mathscr{F}_h$ is the set of all triangular segments K. In practice, we construct Γ_h using Multi-material cubical marching squares algorithm [3].

The induced tessellation of $\Omega_{i,h}$ can be considered as a subdivision of the volume into general polyhedra. Let $\mathscr{T}_{i,h}$ be the tessellation of $\Omega_{i,h}$ into non-intersected polyhedra. For the transport and diffusion in the matrix we apply a non-linear FV method devised on general polyhedral meshes in [4], which is monotone and has compact stencil. The trace of the background mesh on Γ_h induces a 'triangulation' of the fracture, which is very irregular, and so we do not use it do build a discretization method. To handle transport and diffusion along the fracture, we first consider finite element space of piecewise trilinear functions for the volume octree mesh $\mathscr{T}_h$. We further, formally, consider the restrictions (traces) of these background functions on Γ_h and use them in a finite element integral form over Γ_h. Thus the irregular triangulation of Γ_h is used for numerical integration only, while the trial and test functions are tailored to the background regular mesh. It appears that the properties of this trace finite element method are driven by the properties of the background mesh, and they are independent on how Γ_h intersects $\mathscr{T}_h$. The TraceFEM was devised and first analysed in [7] and extended for the octree meshes in [5]. A natural way to couple two approaches is to use the restriction of the background FE solution on Γ_h as the boundary data for the FV method and to compute the FV two-side fluxes on Γ_h to

provide the source terms for the surface discrete equation. Further we provide details of the coupling between discrete bulk and surface equations.

The equations in the bulk and on the surface are coupled through the boundary condition $u_i = v$ on $\partial\Omega_{i,h} \cap \Gamma_h$ (second equation in (1)) and the net flux $F_{\Gamma_h}(u)$ on Γ_h, which stands as the source term in the surface Eq. (2). On Γ_h the solution v_h is defined as a trace of the background finite element piecewise trilinear function. The averaged value of v_h is computed on each surface triangle $K \in \mathscr{F}_h$ using a standard quadrature rule. These values assigned to the barycenters of K from $\mathscr{F}_h$ serve as the Dirichlet boundary data for the FV method on Γ_h. The discrete diffusive and convective fluxes are assigned to barycenters of all faces on $\mathscr{T}_{i,h}, i = 1, \ldots, N$. Since each triangle $K \in \mathscr{F}_h$ is a face for two neighbouring cells $T_i \in \mathscr{T}_{i,h}$ and $T_j \in \mathscr{T}_{j,h}$, $i \neq j$, the diffusive and convective fluxes are assigned to K from both sides of Γ_h. The discrete net flux $F_{\Gamma_h}(u_h)$ at the barycenter of K is computed as the jump of the fluxes over K. In the TraceFEM this value is assigned to all $\mathbf{x} \in K$, and numerical integration is done over all surface elements $K \in \mathscr{F}_h$ to compute the right-hand side of the algebraic system.

To satisfy all discretized equations and boundary conditions we iterate between the bulk FV and surface FE solvers on each time step. We assume an implicit time stepping method (in experiments we use backward Euler). This results in the following system on each time step:

$$\begin{cases} \mathscr{L}u := \tilde{\phi}u + \operatorname{div}(\mathbf{w}u - D\nabla)u = \hat{f} \quad \text{in } \Omega \setminus \Gamma, \\ u = v \quad \text{on } \Gamma, \\ \mathbf{n}_{\partial\Omega} \cdot \nabla u = 0 \quad \text{on } \partial\Omega_N, \quad u = u_D \quad \text{on } \partial\Omega_D, \\ \mathscr{L}_\Gamma v := \tilde{\phi}_\Gamma v + \operatorname{div}_\Gamma(\mathbf{w}_\Gamma v - d D_\Gamma \nabla_\Gamma v) = F_\Gamma(u) + \hat{f}_\Gamma \quad \text{on } \Gamma, \\ \mathbf{n}_{\partial\Gamma} \cdot \nabla v = 0 \quad \text{on } \partial\Gamma_N, \quad v = v_D \quad \text{on } \partial\Gamma_D, \end{cases} \tag{5}$$

the right hand sides $\hat{f}$, $\hat{f}_\Gamma$ account for the solution values at the previous time step.

For the sake of brevity we will not describe the iterative process in this paper and continue with numerical experiments.

4 Numerical Results

This section shows several numerical examples, which demonstrate the accuracy and capability of the hybrid method. Here we confine only steady problems. For unsteady problem with given reference solution we observe that the computed solution well approximates the reference one; the computed front has the correct position and not too much smeared, we do not observe overshoots or undershoots in v_h.

4.1 An Example with a Smooth Curved Surface

The first experiment deals with the case when Γ is a smooth surface embedded in a bulk domain Ω. Consider Γ – the unit sphere centered at the origin and $\Omega = [-1, 1]^3$. By Ω_1 we denote the interior of Γ, Ω_2 denotes the exterior part of Ω. Let $\mathbf{v}(\mathbf{x}) = (-y\sqrt{1-z^2}, x\sqrt{1-z^2}, 0)^T$. The transport velocity field is set to be $\mathbf{w}_\Gamma(\mathbf{x}) = \mathbf{v}(\mathbf{x})$ for $\mathbf{x} \in \Gamma$, $\mathbf{w}_i(\mathbf{x}) = \mathbf{v}(\mathbf{x}) + 0.1\mathbf{s}_i$, $\mathbf{s}_1 = (1, 1, 0)^T$, $\mathbf{s}_2 = (2, 1, 0)^T$. Other parameters in (1), (2) are set to be $D_1 = D_2 = I$, $D_\Gamma = 10I$, $I \in \mathbb{R}^{3\times 3}$ is the identity tensor, $d = 0.1$. In this test we solve for a steady-state solution, so we set $\phi_1 = \phi_2 = \phi_\Gamma = 0$.

For the exact solution on the surface we take $v(\mathbf{x}) = xy \arctan(2z)$ on Γ.

In Ω_2 the bulk concentration u_2 is defined by the same formula as v, and in Ω_1 the bulk concentration is defined by the equality

$$u_1(\mathbf{x}) = xy \arctan(2z) \cdot \exp(1 - |\mathbf{x}|^2) \quad \text{in } \Omega_1.$$

The concentration is continuous across Γ, i.e. the second equation in (1) is satisfied. However, the diffusive flux in (3) is discontinuous across Γ.

We prescribe Dirichlet boundary conditions on $\partial\Omega$. The source terms f_i and f_Γ are computed such that the triple $\{v, u_1, u_2\}$ solves the stationary Eqs. (1)–(2).

Next we apply non-uniform refinement of the bulk mesh, starting with a uniform grid and $h = \frac{1}{4}$. On each refinement step the cells intersected by the surface are refined four times, and the mesh in the bulk is refined one time. The mesh is gradely refined between the surface and bulk cells, see Fig. 1 (right). Table 1 shows the convergence results for the method on the sequence of locally refined grids. The computed solution after one refinement step is demonstrated in Fig. 1. The convergence rates for the fracture solution varies because the refinement is not uniform, but asymptotically the second order can be observed.

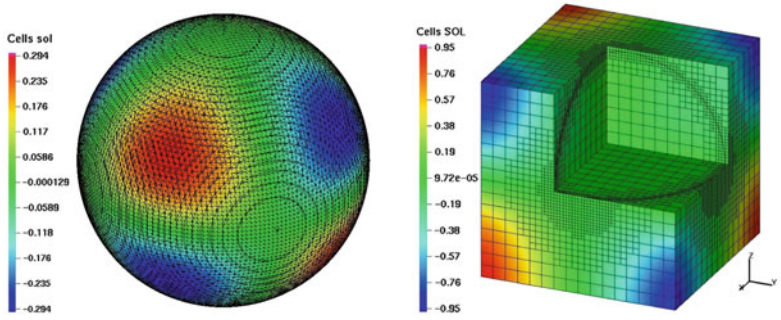


Fig. 1 *Left*: Induced surface mesh and the computed solution on the surface. *Right*: Cutaway of the bulk mesh after one step of local refinement

Table 1 Convergence of numerical solutions in the experiment with a smooth curved Γ and locally refined bulk meshes as in Fig. 1

	#d.o.f.	L^2-norm	rate	H^1-norm	rate	L^∞-norm	rate
3D	120	1.139e-2		1.447e-1		2.817e-2	
	3576	3.457e-3	1.72	5.602e-2	1.37	2.582e-2	0.13
	74176	9.631e-4	1.84	2.111e-2	1.41	7.609e-3	1.76
2D	100	1.043e-2		1.020e-1		1.938e-2	
	1628	1.506e-3	2.79	5.118e-2	0.99	6.467e-3	1.58
	26724	6.134e-4	1.30	2.652e-2	0.95	3.980e-3	0.70

4.2 Steady Analytical Solution for a Triple Fracture Problem

Consider the coupled surface–bulk diffusion problem in the domain $\Omega = [0, 1]^3$ with an embedded piecewise planar Γ. We design Γ to model a branching fracture. In the basic model, $\Gamma = \Gamma(0)$ consists of three planar pieces, $\Gamma(0) = \Gamma_{12} \cup \Gamma_{13} \cup \Gamma_{23}$, $\Gamma_{ij} = \overline{\Omega_i} \cap \overline{\Omega_j}$ $i \neq j$, such that $\Omega_1 = \{\mathbf{x} \in \Omega \,|\, x < 0.5 \text{ and } y > x\}$, $\Omega_2 = \{\mathbf{x} \in \Omega \,|\, x > 0.5 \text{ and } y > x - 1\}$, $\Omega_3 = \Omega \setminus (\overline{\Omega}_1 \cup \overline{\Omega}_2)$.

To model a generic situation when Γ cuts through the background mesh in an arbitrary way, we consider the tessellations of $\Omega = [0, 1]^3$ into three subdomains by a surface $\Gamma(\alpha)$. The surface $\Gamma(\alpha)$ is obtained from $\Gamma(0)$ by applying the clockwise rotation by the angle α around the axis $x = z = 0.5$. We show the results with $\alpha = 20^o$. The resulting tessellation of Ω and surfaces mesh are illustrated in Fig. 2.

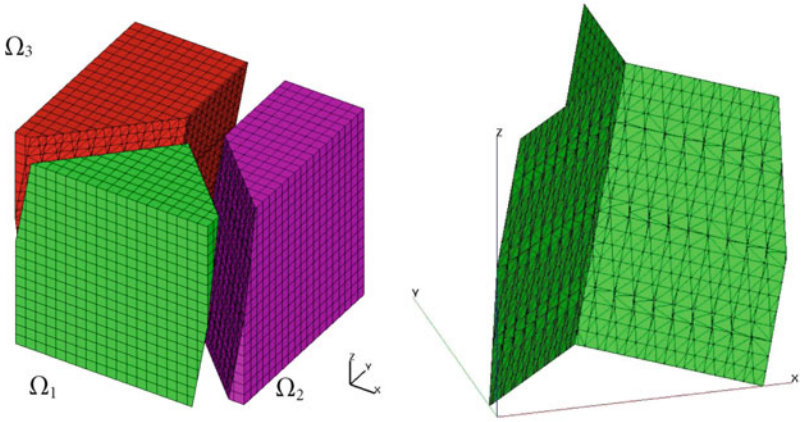


Fig. 2 The figure illustrates the bulk domain with uniform mesh and the surface mesh on the fracture, rotated by 20 degrees

Table 2 The error in the numerical solution for the steady problem with triple fracture, $\alpha = 20$

	#d.o.f.	L^2-norm	rate	L^∞-norm	rate
3D	965	6.319e-3		3.754e-2	
	7872	1.805e-3	1.79	1.280e-2	1.55
	63592	5.623e-4	1.80	3.411e-3	1.90
2D	321	7.792e-3		2.716e-2	
	1692	2.084e-3	1.59	5.400e-3	1.94
	7944	7.019e-4	1.41	2.001e-3	1.29

To define the solution $\{v, u\}$ solving the stationary Eq. (1), we introduce

$$\psi_1 = \begin{cases} 16(y - \frac{1}{2})^4, & y > \frac{1}{2} \\ 0, & y \le \frac{1}{2} \end{cases}, \quad \psi_2 = x - y, \quad \psi_3 = x + y - 1.$$

We define the solution of the basic model problem ($\alpha = 0$)

$$u(\mathbf{x}) = \begin{cases} \sin(2\pi z) \cdot \psi_2(\mathbf{x}) \cdot \phi_3(\mathbf{x}) & \mathbf{x} \in \Omega_1, \\ \sin(2\pi z) \cdot \psi_1(\mathbf{x}) & \mathbf{x} \in \Omega_2, \\ \sin(2\pi z) 2x \cdot \psi_1(\mathbf{x}) & \mathbf{x} \in \Omega_3, \end{cases} \qquad v = u|_{\Gamma(0)}.$$

The solution for the problem with rotated fracture is obtained by applying the same rotation. Other parameters are set to be $\mathbf{w} = \mathbf{w}_\Gamma = 0$, $\phi_i = \phi_\Gamma = 0$, $D_i = I$, $D_{\Gamma,i} = 10I$ for $i = 1..3$, and $d_{23} = 0.1$, $d_{13} = d_{12} = \frac{0.1}{\sqrt{2}}$. An interesting feature of this problem is that the surface Γ is only piecewise smooth. The bulk grid is not fitted to the internal edge $\mathscr{E} = \Gamma_{12} \cap \Gamma_{13} \cap \Gamma_{23}$, and hence the tangential derivatives of v are discontinuous inside certain cubic cells from $\mathscr{T}_h^\Gamma$. We have the situation, when a kink in v is not resolved in the finite element spaces. This is well-known to result in the $\frac{1}{2}$-reduction of convergence order. This suboptimal order for a sequence of uniform background meshes is demonstrated by the results in Table 2.

Acknowledgements This work has been supported by RFBR through the grant 16-31-00527 and by NSF through the Division of Mathematical Sciences grant 1522252.

References

1. Alboin, C., Jaffré, J., Roberts, J.E., Serres, C.: Modeling fractures as interfaces for flow and transport. In: Fluid Flow and Transport in Porous Media, Mathematical and Numerical Treatment, vol. 295, p. 13 (2002). American Mathematical Soc
2. Burman, E., Claus, S., Hansbo, P., Larson, M.G., Massing, A.: Cutfem: Discretizing geometry and partial differential equations. Int. J. Numer. Meth. Eng. **104**(7), 472–501 (2015)
3. Chernyshenko, A.: Generation of octree meshes with cut cells in multiple material domains. Num. Methods Progr. **14**, 229–245 (2013). (in russian)

4. Chernyshenko, A., Vassilevski, Y.: A finite volume scheme with the discrete maximum principle for diffusion equations on polyhedral meshes. In: Fuhrmann, J., Ohlberger, M., Rohde, C. (eds.) Finite Volumes for Complex Applications VII-Methods and Theoretical Aspects, Springer Proceedings in Mathematics and Statistics, vol. 77, pp. 197–205. Springer International Publishing, Switzerland (2014)
5. Chernyshenko, A.Y., Olshanskii, M.A.: An adaptive octree finite element method for pdes posed on surfaces. Comput. Methods Appl. Mech. Eng. **291**, 146–172 (2015)
6. Lipnikov, K., Svyatskiy, D., Vassilevski, Y.: Minimal stencil finite volume scheme with the discrete maximum principle. Russian J. Numer. Anal. Math. Modelling **27**(4), 369–385 (2012)
7. Olshanskii, M., Reusken, A., Grande, J.: A finite element method for elliptic equations on surfaces. SIAM J. Numer. Anal. **47**, 3339–3358 (2009)

A Nonconforming High-Order Method for Nonlinear Poroelasticity

Michele Botti, Daniele A. Di Pietro and Pierre Sochala

Abstract In this work, we introduce a novel algorithm for the quasi-static nonlinear poroelasticity problem describing Darcian flow in a deformable saturated porous medium. The nonlinear elasticity operator is discretized using a Hybrid High-Order method while the heterogeneous diffusion part relies on a Symmetric Weighted Interior Penalty discontinuous Galerkin scheme. The method is valid in two and three space dimensions, delivers an inf-sup stable discretization on general meshes including polyhedral elements and nonmatching interfaces, allows arbitrary approximation orders, and has a reduced cost thanks to the possibility of statically condensing a large subset of the unknowns for linearized versions of the problem. Moreover, the proposed construction can handle rough variations of the permeability coefficient and vanishing specific storage coefficient. Numerical tests demonstrating the performance of the method are provided.

Keywords Nonlinear poroelasticity · Hybrid high-order · Discontinuous galerkin · General meshes

Classifications 65M08 · 65N30 · 74B20 · 76S05

M. Botti (✉) · D.A. Di Pietro
Université de Montpellier, Institut Montpéllierain Alexander Grothendieck,
Place Eugène Bataillon, Montpellier 34095, France
e-mail: michele.botti@umontpellier.fr

D.A. Di Pietro
e-mail: daniele.di-pietro@umontpellier.fr

P. Sochala
Bureau de Recherches Géologiques Et Minières, 3, Avenue Claude-Guillemin,
45060 Orléans, France
e-mail: p.sochala@brgm.fr

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_56

1 Introduction

We consider in this work the nonlinear poroelasticity model obtained by generalizing the linear Biot's consolidation model of [1, 7] to nonlinear stress-strain constitutive laws. Our original motivation comes from applications in geosciences, where the support of polyhedral meshes and nonconforming interfaces is crucial.

Let $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, denote a bounded connected polyhedral domain with Lipschitz boundary $\partial\Omega$ and outward normal $\mathbf{n}$. For a given finite time $t_F > 0$, volumetric load $\mathbf{f}$, fluid source g, the considered nonlinear poroelasticity problem consists in finding a vector-valued displacement field $\mathbf{u}$ and a scalar-valued pore pressure field p solution of

$$-\nabla\cdot\sigma(\cdot, \nabla_s\mathbf{u}) + \nabla p = \mathbf{f} \qquad \text{in } \Omega \times (0, t_F), \tag{1a}$$

$$c_0\partial_t p + \partial_t\nabla\cdot\mathbf{u} - \nabla\cdot(\kappa(\cdot)\nabla p) = g \qquad \text{in } \Omega \times (0, t_F), \tag{1b}$$

where ∇_s denotes the symmetric gradient, $c_0 \geq 0$ is the constrained specific storage coefficient, and $\kappa : \Omega \to (0, \overline{\kappa}]$ is the scalar-valued permeability field. Eqs. (1a) and (1b) express, respectively, the momentum equilibrium and the fluid mass balance. For the sake of simplicity, we assume that κ is piecewise constant on a partition P_Ω of Ω into bounded disjoint polyhedra and we consider the following homogeneous boundary conditions:

$$\mathbf{u} = \mathbf{0} \qquad \text{on } \partial\Omega \times (0, t_F), \tag{1c}$$

$$(\kappa(\cdot)\nabla p)\cdot\mathbf{n} = 0 \qquad \text{on } \partial\Omega \times (0, t_F). \tag{1d}$$

The treatment of more general permeability fields and boundary conditions is possible up to minor modifications. Initial conditions are set prescribing $\mathbf{u}(\cdot, 0) = \mathbf{u}^0$ and, if $c_0 > 0$, $p(\cdot, 0) = p^0$. In the incompressible case $c_0 = 0$, we also need the following compatibility condition on g and zero-average constraint on p:

$$\int_\Omega g(\cdot, t) = 0 \quad \text{and} \quad \int_\Omega p(\cdot, t) = 0 \qquad \forall t \in (0, t_F). \tag{1e}$$

We assume that the symmetric stress tensor $\sigma : \Omega \times \mathbb{R}^{d\times d}_{\text{sym}} \to \mathbb{R}^{d\times d}_{\text{sym}}$ is a Caratheodory function such that there exist real numbers $\overline{\sigma}, \underline{\sigma} \in (0, +\infty)$ and, for a.e. $\mathbf{x} \in \Omega$, and all $\tau, \eta \in \mathbb{R}^{d\times d}_{\text{sym}}$, the following conditions hold:

$$\|\sigma(\mathbf{x}, \tau) - \sigma(\mathbf{x}, \mathbf{0})\|_{d\times d} \leq \overline{\sigma}\|\tau\|_{d\times d}, \quad \text{(growth)} \tag{2a}$$

$$\sigma(\mathbf{x}, \tau) : \tau \geq \underline{\sigma}\|\tau\|^2_{d\times d}, \quad \text{(coercivity)} \tag{2b}$$

$$(\sigma(\mathbf{x}, \tau) - \sigma(\mathbf{x}, \eta)) : (\tau - \eta) \geq 0, \quad \text{(monotonicity)} \tag{2c}$$

where $\tau : \eta := \sum_{i,j=1}^d \tau_{i,j}\eta_{i,j}$ and $\|\tau\|^2_{d\times d} = \tau : \tau$.

2 Mesh and Notation

Denote by $\mathscr{H} \subset \mathbb{R}_*^+$ a countable set having 0 as unique accumulation point. We consider refined mesh sequences $(\mathscr{T}_h)_{h\in\mathscr{H}}$ where each $\mathscr{T}_h$ is a finite collection of disjoint open polyhedral elements T with boundary ∂T such that $\overline{\Omega} = \bigcup_{T\in\mathscr{T}_h} \overline{T}$ and $h = \max_{T\in\mathscr{T}_h} h_T$ with h_T diameter of T. We assume that mesh regularity holds in the sense of [4, Definition 1.38] and that, for all $h \in \mathscr{H}$, $\mathscr{T}_h$ is compatible with the partition P_Ω on which the permeability coefficient κ is piecewise constant, so that jumps of the permeability coefficient do not occur inside mesh elements.

Mesh faces are hyperplanar subsets of $\overline{\Omega}$ with positive $(d-1)$-dimensional Hausdorff measure and disjoint interiors. Interfaces are collected in the set $\mathscr{F}_h^{\mathrm{i}}$, boundary faces in $\mathscr{F}_h^{\mathrm{b}}$, and we assume that $\mathscr{F}_h := \mathscr{F}_h^{\mathrm{i}} \cup \mathscr{F}_h^{\mathrm{b}}$ is such that $\bigcup_{T\in\mathscr{T}_h} \partial T = \bigcup_{F\in\mathscr{F}_h} F$. For all $T \in \mathscr{T}_h$, $\mathscr{F}_T := \{F \in \mathscr{F}_h \mid F \subset \partial T\}$ denotes the set of faces contained in ∂T and, for all $F \in \mathscr{F}_T$, $\mathbf{n}_{TF}$ is the unit normal to F pointing out of T.

For $X \subset \overline{\Omega}$, we denote by $\|\cdot\|_X$ the norm in $L^2(X;\mathbb{R})$, $L^2(X;\mathbb{R}^d)$, and $L^2(X;\mathbb{R}^{d\times d})$. For $l \geq 0$, the space $\mathbb{P}^l(X;\mathbb{R})$ is spanned by the restriction to X of polynomials of total degree l. On regular mesh sequences, we have the following optimal approximation property for the L^2-projector $\pi_X^l : L^1(X;\mathbb{R}) \to \mathbb{P}^l(X;\mathbb{R})$: There exists $C_{\mathrm{ap}} > 0$ such that, for all $h \in \mathscr{H}$, all $T \in \mathscr{T}_h$, all $s \in \{1,\dots,l+1\}$, and all $v \in H^s(T;\mathbb{R})$,

$$|v - \pi_T^l v|_{H^m(T;\mathbb{R})} \leq C_{\mathrm{ap}} h_T^{s-m} |v|_{H^s(T;\mathbb{R})} \qquad \forall m \in \{0,\dots,s\}. \tag{3}$$

Other geometric and analytic results on regular meshes can be found in [4, Chap. 1] and [3]. In what follows, for an integer $l \geq 0$, we denote by $\mathbb{P}^l(\mathscr{T}_h;\mathbb{R})$, $\mathbb{P}^l(\mathscr{T}_h;\mathbb{R}^d)$, and $\mathbb{P}^l(\mathscr{T}_h;\mathbb{R}^{d\times d})$, respectively, the space of scalar-, vector-, and tensor-valued broken polynomials of total degree l on $\mathscr{T}_h$. The space of broken vector-valued polynomials of total degree l on the mesh skeleton is denoted by $\mathbb{P}^l(\mathscr{F}_h;\mathbb{R}^d)$.

3 Discretization

In this section we define the discrete counterparts of the elasticity and Darcy operators and of the hydro-mechanical coupling terms.

3.1 Nonlinear Elasticity Operator

The discretization of the nonlinear elasticity operator is based on the Hybrid High-Order method of [5]. Let a polynomial degree $k \geq 1$ be fixed. The degrees of freedom (DOFs) for the displacement are collected in the space $\underline{\mathbf{U}}_h^k := \mathbb{P}^k(\mathscr{T}_h;\mathbb{R}^d) \times \mathbb{P}^k(\mathscr{F}_h;\mathbb{R}^d)$. To account for the Dirichlet condition (1c) we define the subspace

$$\underline{\mathbf{U}}_{h,0}^k := \left\{ \underline{\mathbf{v}}_h := \left((\mathbf{v}_T)_{T\in\mathscr{T}_h}, (\mathbf{v}_F)_{F\in\mathscr{F}_h} \right) \in \underline{\mathbf{U}}_h^k \mid \mathbf{v}_F = \mathbf{0} \quad \forall F \in \mathscr{F}_h^{\mathrm{b}} \right\},$$

equipped with the discrete strain norm

$$\|\underline{\mathbf{v}}_h\|_{\varepsilon,h} := \left(\sum_{T\in\mathscr{T}_h} \|\underline{\mathbf{v}}_h\|_{\varepsilon,T}^2 \right)^{1/2}, \quad \|\underline{\mathbf{v}}_h\|_{\varepsilon,T}^2 := \|\nabla_{\mathrm{s}}\mathbf{v}_T\|_T^2 + \sum_{F\in\mathscr{F}_T} \frac{\|\mathbf{v}_F - \mathbf{v}_T\|_F^2}{h_F}.$$

The DOFs corresponding to a function $\mathbf{v} \in H_0^1(\Omega;\mathbb{R}^d)$ are obtained by means of the reduction map $\underline{\mathbf{I}}_h^k : H_0^1(\Omega;\mathbb{R}^d) \to \underline{\mathbf{U}}_{h,0}^k$ such that $\underline{\mathbf{I}}_h^k \mathbf{v} := \left((\boldsymbol{\pi}_T^k \mathbf{v})_{T\in\mathscr{T}_h}, (\boldsymbol{\pi}_F^k \mathbf{v})_{F\in\mathscr{F}_h} \right)$. Using the H^1-stability of the L^2-projector and the trace inequality [4, Lemma 1.49], we infer the existence of $C_{\mathrm{st}} > 0$ independent of h such that, for all $\mathbf{v} \in H_0^1(\Omega;\mathbb{R}^d)$,

$$\|\underline{\mathbf{I}}_h^k \mathbf{v}\|_{\varepsilon,h} \le C_{\mathrm{st}} \|\mathbf{v}\|_{H^1(\Omega;\mathbb{R}^d)}. \tag{4}$$

For all $T \in \mathscr{T}_h$, we denote by $\underline{\mathbf{U}}_T^k$ and $\underline{\mathbf{I}}_T^k$ the restrictions to T of $\underline{\mathbf{U}}_h^k$ and $\underline{\mathbf{I}}_h^k$, and we define the local symmetric gradient reconstruction $\mathsf{G}_{\mathrm{s},T}^k : \underline{\mathbf{U}}_T^k \to \mathbb{P}^k(T;\mathbb{R}_{\mathrm{sym}}^{d\times d})$ as the unique solution of the pure traction problem: For a given $\underline{\mathbf{v}}_T = \left(\mathbf{v}_T, (\mathbf{v}_F)_{F\in\mathscr{F}_T} \right) \in \underline{\mathbf{U}}_T^k$, find $\mathsf{G}_{\mathrm{s},T}^k \underline{\mathbf{v}}_T \in \mathbb{P}^k(T;\mathbb{R}_{\mathrm{sym}}^{d\times d})$ such that, for all $\tau \in \mathbb{P}^k(T;\mathbb{R}_{\mathrm{sym}}^{d\times d})$,

$$\int_T \mathsf{G}_{\mathrm{s},T}^k \underline{\mathbf{v}}_T : \tau = -\int_T \mathbf{v}_T \cdot (\nabla\cdot\tau) + \sum_{F\in\mathscr{F}_T} \int_F \mathbf{v}_F \cdot (\tau \mathbf{n}_{TF}). \tag{5}$$

The definition of $\mathsf{G}_{\mathrm{s},T}^k$ is justified by the following commuting property that, combined with (3), shows that $\mathsf{G}_{\mathrm{s},T}^k \underline{\mathbf{I}}_T^k$ has optimal approximation properties in $\mathbb{P}^k(T;\mathbb{R}_{\mathrm{sym}}^{d\times d})$.

Lemma 1 *For all $T \in \mathscr{T}_h$ and all $\mathbf{v} \in H^1(T;\mathbb{R}^d)$, $\mathsf{G}_{\mathrm{s},T}^k \underline{\mathbf{I}}_T^k \mathbf{v} = \boldsymbol{\pi}_T^k(\nabla_{\mathrm{s}}\mathbf{v})$.*

Proof Let $T \in \mathscr{T}_h$ and $\mathbf{v} \in H^1(T;\mathbb{R}^d)$. For all $\tau \in \mathbb{P}^k(T;\mathbb{R}_{\mathrm{sym}}^{d\times d})$, we have

$$\begin{aligned}\int_T \mathsf{G}_{\mathrm{s},T}^k \underline{\mathbf{I}}_T^k \mathbf{v} : \tau &= -\int_T \boldsymbol{\pi}_T^k \mathbf{v} \cdot (\nabla\cdot\tau) + \sum_{F\in\mathscr{F}_T} \int_F \boldsymbol{\pi}_F^k \mathbf{v} \cdot (\tau \mathbf{n}_{TF}) \\ &= -\int_T \mathbf{v} \cdot (\nabla\cdot\tau) + \sum_{F\in\mathscr{F}_T} \int_F \mathbf{v} \cdot (\tau \mathbf{n}_{TF}) = \int_T \boldsymbol{\pi}_T^k(\nabla_{\mathrm{s}}\mathbf{v}) : \tau.\end{aligned}$$

□

From $\mathsf{G}_{\mathrm{s},T}^k$ we define the local displacement reconstruction operator $\mathbf{r}_T^{k+1} : \underline{\mathbf{U}}_T^k \to \mathbb{P}^{k+1}(T;\mathbb{R}^d)$ such that, for all $\underline{\mathbf{v}}_T \in \underline{\mathbf{U}}_T^k$ and all $\mathbf{w} \in \mathbb{P}^{k+1}(T;\mathbb{R}^d)$, it holds

$$\int_T (\nabla_{\mathrm{s}} \mathbf{r}_T^{k+1} \underline{\mathbf{v}}_T - \mathsf{G}_{\mathrm{s},T}^k \underline{\mathbf{v}}_T) : \nabla_{\mathrm{s}} \mathbf{w} = 0$$

$$\int_T \mathbf{r}_T^{k+1} \underline{\mathbf{v}}_T = \int_T \mathbf{v}_T, \qquad \int_T \nabla_{\mathrm{ss}} \mathbf{r}_T^{k+1} \underline{\mathbf{v}}_T = \sum_{F \in \mathscr{F}_T} \int_F \tfrac{1}{2} \left(\mathbf{n}_{TF} \otimes \mathbf{v}_F - \mathbf{v}_F \otimes \mathbf{n}_{TF} \right),$$

where ∇_{ss} denotes the skew-symmetric part of the gradient operator.

The discretization of the nonlinear elasticity operator is realized by the function $a_h : \underline{\mathbf{U}}_h^k \times \underline{\mathbf{U}}_h^k \to \mathbb{R}$ defined such that, for all $\underline{\mathbf{w}}_h, \underline{\mathbf{v}}_h \in \underline{\mathbf{U}}_h^k$,

$$a_h(\underline{\mathbf{w}}_h, \underline{\mathbf{v}}_h) := \sum_{T \in \mathscr{T}_h} \left(\int_T \sigma(\cdot, \mathsf{G}_{\mathrm{s},T}^k \underline{\mathbf{u}}_T) : \mathsf{G}_{\mathrm{s},T}^k \underline{\mathbf{v}}_T + \sum_{F \in \mathscr{F}_T} \frac{\gamma}{h_F} \int_F \Delta_{TF}^k \underline{\mathbf{u}}_T \cdot \Delta_{TF}^k \underline{\mathbf{v}}_T \right), \tag{6}$$

where we penalize in a least-square sense the face-based residual $\Delta_{TF}^k : \underline{\mathbf{U}}_T^k \to \mathbb{P}^k(F; \mathbb{R}^d)$ such that, for all $T \in \mathscr{T}_h$, all $\underline{\mathbf{v}}_T \in \underline{\mathbf{U}}_T^k$, and all $F \in \mathscr{F}_T$,

$$\Delta_{TF}^k \underline{\mathbf{v}}_T := \boldsymbol{\pi}_F^k (\mathbf{r}_T^{k+1} \underline{\mathbf{v}}_T - \mathbf{v}_F) - \boldsymbol{\pi}_T^k (\mathbf{r}_T^{k+1} \underline{\mathbf{v}}_T - \mathbf{v}_T).$$

This definition ensures that Δ_{TF}^k vanishes whenever its argument is of the form $\underline{\mathbf{I}}_T^k \mathbf{w}$ with $\mathbf{w} \in \mathbb{P}^{k+1}(T; \mathbb{R}^d)$, a crucial property to obtain high-order error estimates (cf. [2, Theorem 12]). A possible choice for the scaling parameter $\gamma > 0$ in (6) is $\gamma = \overline{\sigma}$. For all $\underline{\mathbf{v}}_h \in \underline{\mathbf{U}}_{h,0}^k$, it holds (the proof follows from [5, Lemma 4]):

$$C_{\mathrm{eq}}^{-1} \|\underline{\mathbf{v}}_h\|_{\varepsilon,h}^2 \le \sum_{T \in \mathscr{T}_h} \left(\|\mathsf{G}_{\mathrm{s},T}^k \underline{\mathbf{v}}_T\|_T^2 + \sum_{F \in \mathscr{F}_T} \frac{\gamma}{h_F} \|\Delta_{TF}^k \underline{\mathbf{v}}_T\|_F^2 \right) \le C_{\mathrm{eq}} \|\underline{\mathbf{v}}_h\|_{\varepsilon,h}^2,$$

where $C_{\mathrm{eq}} > 0$ is independent of h. By (2b), this implies the coercivity of a_h.

3.2 Darcy Operator

The discretization of the Darcy operator is based on the Symmetric Weighted Interior Penalty method of [6], cf. also [4, Sect. 4.5]. At each time step, the discrete pore pressure is sought in the broken polynomial space

$$P_h^k := \begin{cases} \mathbb{P}^k(\mathscr{T}_h; \mathbb{R}) & \text{if } c_0 > 0, \\ \mathbb{P}_0^k(\mathscr{T}_h; \mathbb{R}) := \left\{ q_h \in \mathbb{P}^k(\mathscr{T}_h; \mathbb{R}) \mid \int_\Omega q_h = 0 \right\} & \text{if } c_0 = 0. \end{cases}$$

For all $F \in \mathscr{F}_h^{\mathrm{i}}$ and all $q_h \in \mathbb{P}^k(\mathscr{T}_h;\mathbb{R})$, we define the jump and average operators such that, denoting by q_T and κ_T the restrictions of q_h and κ to an element $T \in \mathscr{T}_h$,

$$[q_h]_F := q_{T_{F,1}} - q_{T_{F,2}}, \qquad \{q_h\}_F := \frac{\kappa_{T_{F,2}}}{\kappa_{T_{F,1}}+\kappa_{T_{F,2}}} q_{T_{F,1}} + \frac{\kappa_{T_{F,1}}}{\kappa_{T_{F,1}}+\kappa_{T_{F,2}}} q_{T_{F,2}},$$

where $T_{F,1}, T_{F,2} \in \mathscr{T}_h$ are such that $F \subset T_{F,1} \cap T_{F,2}$. The bilinear form c_h on $P_h^k \times P_h^k$ is defined such that, for all $q_h, r_h \in P_h^k$,

$$\begin{aligned} c_h(r_h,q_h) := \int_\Omega \kappa \boldsymbol{\nabla}_h r_h \cdot \boldsymbol{\nabla}_h q_h + \sum_{F\in\mathscr{F}_h^{\mathrm{i}}} \frac{2\varsigma\kappa_{T_{F,1}}\kappa_{T_{F,2}}}{h_F(\kappa_{T_{F,1}}+\kappa_{T_{F,2}})} \int_F [r_h]_F [q_h]_F \\ - \sum_{F\in\mathscr{F}_h^{\mathrm{i}}} \int_F ([r_h]_F \{\kappa\boldsymbol{\nabla}_h q_h\}_F + [q_h]_F \{\kappa\boldsymbol{\nabla}_h r_h\}_F)\cdot \mathbf{n}_{T_{F,1}F}, \end{aligned} \tag{7}$$

where $\boldsymbol{\nabla}_h$ denotes the broken gradient and $\varsigma > 0$ is a user-defined penalty parameter chosen large enough to ensure the coercivity of c_h (cf. [4, Lemma 4.51]). In the numerical tests of Sect. 5, we took $\varsigma = (N_\partial + 0.1)k^2$, with N_∂ equal to the maximum number of faces between the elements in $\mathscr{T}_h$. The fact that the boundary terms only appear on internal faces in (7) reflects the Neumann boundary condition (1d).

3.3 Hydro-Mechanical Coupling

The hydro-mechanical coupling is realized by means of the bilinear form b_h on $\underline{\mathbf{U}}_{h,0}^k \times \mathbb{P}^k(\mathscr{T}_h;\mathbb{R})$ such that, for all $\underline{\mathbf{v}}_h \in \underline{\mathbf{U}}_{h,0}^k$ and all $q_h \in \mathbb{P}^k(\mathscr{T}_h;\mathbb{R})$,

$$b_h(\underline{\mathbf{v}}_h, q_h) := -\sum_{T\in\mathscr{T}_h} \int_T \mathsf{G}_{\mathrm{s},T}^k \underline{\mathbf{v}}_T : q_{h|T}\mathsf{I}_d, \tag{8}$$

where $\mathsf{I}_d \in \mathbb{R}^{d\times d}_{\mathrm{sym}}$ is the identity matrix. A simple verification shows that there exists $C_{\mathrm{bd}} > 0$ independent of h such that $b_h(\underline{\mathbf{v}}_h, q_h) \le C_{\mathrm{bd}} \|\underline{\mathbf{v}}_h\|_{\varepsilon,h} \|q_h\|_\Omega$. Additionally, using definition (5) of $\mathsf{G}_{\mathrm{s},T}^k$, it can be proved that, for all $\underline{\mathbf{v}}_h \in \underline{\mathbf{U}}_{h,0}^k$, $b_h(\underline{\mathbf{v}}_h, 1) = 0$. The following inf-sup condition expresses the stability of the coupling:

Proposition 1 *There is a real β independent of h such that, for all $q_h \in \mathbb{P}_0^k(\mathscr{T}_h;\mathbb{R})$,*

$$\|q_h\|_\Omega \le \beta \sup_{\underline{\mathbf{v}}_h \in \underline{\mathbf{U}}_{h,0}^k\setminus\{\mathbf{0}\}} \frac{b_h(\underline{\mathbf{v}}_h, q_h)}{\|\underline{\mathbf{v}}_h\|_{\varepsilon,h}}. \tag{9}$$

Proof Let $q_h \in \mathbb{P}_0^k(\mathscr{T}_h;\mathbb{R})$. There is $\mathbf{v}_{q_h} \in H_0^1(\Omega;\mathbb{R}^d)$ such that $\boldsymbol{\nabla}\cdot\mathbf{v}_{q_h} = q_h$ and $\|\mathbf{v}_{q_h}\|_{H^1(\Omega;\mathbb{R}^d)} \le C_{\mathrm{sj}}\|q_h\|_\Omega$, with $C_{\mathrm{sj}} > 0$ independent of h. Owing to (4) we get

$$\|\mathbf{I}_h^k \mathbf{v}_{q_h}\|_{\varepsilon,h} \le C_{\mathrm{st}} \|\mathbf{v}_{q_h}\|_{H^1(\Omega;\mathbb{R}^d)} \le C_{\mathrm{st}} C_{\mathrm{sj}} \|q_h\|_\Omega.$$

Therefore, using the commuting property of Lemma 1, denoting by S the supremum in (9), and using the previous inequality, it is inferred that

$$\|q_h\|_\Omega^2 = \sum_{T\in\mathscr{T}_h}\int_T (\nabla_{\mathrm{s}}\mathbf{v}_{q_h} : q_h \mathsf{I}_d)_{|T} = -b_h(\underline{\mathbf{I}}_h^k \mathbf{v}_{q_h}, q_h) \le \mathsf{S}\|\underline{\mathbf{I}}_h^k \mathbf{v}_{q_h}\|_{\varepsilon,h} \le C_{\mathrm{st}} C_{\mathrm{sj}} \mathsf{S}\|q_h\|_\Omega.$$

□

If a HHO discretization were used also for the Darcy operator, only cell DOFs would be controlled by the inf-sup condition.

4 Formulation of the Method

For the time discretization, we consider a uniform mesh of the time interval $(0, t_{\mathrm{F}})$ of step $\tau := t_{\mathrm{F}}/N$ with $N \in \mathbb{N}^*$, and introduce the discrete times $t^n := n\tau$ for all $0 \le n \le N$. For any $\varphi \in C^l([0, t_{\mathrm{F}}]; V)$, we set $\varphi^n := \varphi(t^n) \in V$ and let, for all $1 \le n \le N$,

$$\delta_t \varphi^n := \frac{\varphi^n - \varphi^{n-1}}{\tau} \in V.$$

For all $1 \le n \le N$, the discrete solution $(\underline{\mathbf{u}}_h^n, p_h^n) \in \underline{\mathbf{U}}_{h,0}^k \times P_h^k$ at time t^n is such that, for all $(\underline{\mathbf{v}}_h, q_h) \in \underline{\mathbf{U}}_{h,0}^k \times \mathbb{P}^k(\mathscr{T}_h; \mathbb{R})$,

$$a_h(\underline{\mathbf{u}}_h^n, \underline{\mathbf{v}}_h) + b_h(\underline{\mathbf{v}}_h, p_h^n) = \sum_{T\in\mathscr{T}_h}\int_T \mathbf{f}^n \cdot \mathbf{v}_T, \quad (10a)$$

$$c_0 \int_\Omega (\delta_t p_h^n) q_h - b_h(\delta_t \underline{\mathbf{u}}_h^n, q_h) + c_h(p_h^n, q_h) = \int_\Omega g^n q_h. \quad (10b)$$

If $c_0 = 0$, we set the initial discrete displacement as $\underline{\mathbf{u}}_h^0 = \underline{\mathbf{I}}_h^k \mathbf{u}^0$. If $c_0 > 0$, the usual way to enforce the initial condition is to compute a displacement from the given initial pressure p^0. We let $p_h^0 := \pi_h^k p^0$ and set $\underline{\mathbf{u}}_h^0 \in \underline{\mathbf{U}}_{h,0}^k$ as the solution of

$$a_h(\underline{\mathbf{u}}_h^0, \underline{\mathbf{v}}_h) = \sum_{T\in\mathscr{T}_h}\int_T \mathbf{f}^0 \cdot \mathbf{v}_T - b_h(\underline{\mathbf{v}}_h, p_h^0) \qquad \forall \underline{\mathbf{v}}_h \in \underline{\mathbf{U}}_{h,0}^k.$$

At each time step n the discrete nonlinear Eq. (10) are solved by the Newton's method using as initial guess the solution at step $n-1$. At each Newton's iteration the Jacobian matrix is computed analytically and in the linearized system the displacement element unknowns can be statically condensed (cf. [2, Sect. 5]).

5 Numerical Results

We consider a regular exact solution in order to assess the convergence of the method for polynomial degree $k = 1$. Specifically, we solve problem (1) in the square domain $\Omega = (0, 1)^2$ with $t_F = 1$ and physical parameters $c_0 = 0$ and $\kappa = 1$. As nonlinear constitutive law we take the Hencky–Mises relation given by

$$\sigma(\nabla_s \mathbf{u}) = (2\mathrm{e}^{-\mathrm{dev}(\nabla_s \mathbf{u})} - 1)\,\mathrm{tr}(\nabla_s \mathbf{u})\mathsf{I}_d + (4 - 2\mathrm{e}^{-\mathrm{dev}(\nabla_s \mathbf{u})})\nabla_s \mathbf{u},$$

where $\mathrm{tr}(\boldsymbol{\tau}) := \boldsymbol{\tau} : \mathsf{I}_d$ and $\mathrm{dev}(\boldsymbol{\tau}) = \mathrm{tr}(\boldsymbol{\tau}^2) - \frac{1}{d}\,\mathrm{tr}(\boldsymbol{\tau})^2$ are the trace and deviatoric operators. It can be checked that the previous stress-strain relation satisfies (2). The exact displacement $\mathbf{u}$ and exact pressure p are given by

$$\mathbf{u}(\mathbf{x}, t) = t^2\big(\sin(\pi x_1)\sin(\pi x_2),\ \sin(\pi x_1)\sin(\pi x_2)\big),$$
$$p(\mathbf{x}, t) = -\pi^{-1} t\,(\sin(\pi x_1)\cos(\pi x_2) + \cos(\pi x_1)\sin(\pi x_2)).$$

The volumetric load $\mathbf{f}$, the source term g, and the boundary conditions are inferred from the exact solutions. The time step τ on the coarsest mesh is 0.2 and it decreases with the mesh size h according to $\tau_1/\tau_2 = 2h_1/h_2$. In Fig. 1 we display the convergence results obtained on two mesh families. The method exhibits second order convergence with respect to the mesh size h for both the energy norm of the displacement and the L^2-norm of the pressure at final time N. Further numerical tests, including higher-order approximation, will be considered in a future publication.

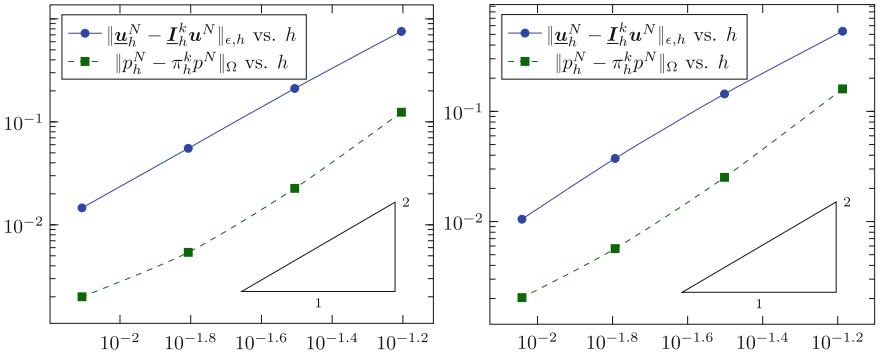


Fig. 1 Convergence tests on a Cartesian mesh family (*left*) and on a Voronoi mesh family (*right*)

Acknowledgements The work of M. Botti was partially supported by Labex NUMEV (ANR-10-LABX-20) ref. 2014-2-006. The work of D. A. Di Pietro was partially supported by *Agence Nationale de la Recherche* grant HHOMM (ref. ANR-15-CE40-0005).

References

1. Biot, M.A.: General theory of threedimensional consolidation. J. Appl. Phys. **12**(2), 155–164 (1941)
2. Boffi, D., Botti, M., Di Pietro, D.A.: A nonconforming high-order method for the biot problem on general meshes. SIAM J. Sci. Comp. **38**(3), A1508–A1537 (2016)
3. Di Pietro, D.A., Droniou, J.: A hybrid high-order method for Leray–Lions elliptic equations on general meshes. Math. Comp. (2017). doi:10.1090/mcom/3180
4. Di Pietro, D.A., Ern, A.: Mathematical Aspects of Discontinuous Galerkin Methods, Mathématiques & Applications, vol. 69. Springer, Heidelberg (2012)
5. Di Pietro, D.A., Ern, A.: A hybrid high-order locking-free method for linear elasticity on general meshes. Comput. Meth. Appl. Mech. Eng. **283**, 1–21 (2015)
6. Di Pietro, D.A., Ern, A., Guermond, J.L.: Discontinuous Galerkin methods for anisotropic semi-definite diffusion with advection. SIAM J. Numer. Anal. **46**(2), 805–831 (2008)
7. Terzaghi, K.: Theoretical Soil Mechanics. Wiley, New York (1943)

New Criteria for Mesh Adaptation in Finite Volume Simulation of Planar Ionization Wavefront Propagation

Hanen Amor, Fayssal Benkhaldoun, Tarek Ghoudi, Imad Kissami and Mohammed Seaid

Abstract Adaptive unstructured finite volume methods for ionization waves are receiving increased attention mainly because of their ability to provide a flexible spatial discretization. Hence, some areas can be resolved in great detail while not over-resolving other areas. Our purpose is to examine the numerical performance of a new criteria for mesh adaptation which account only for the elliptic equation for the electric potential. The proposed adaptive finite volume method has important advantages in the discretization of the gradient fluxes and diffusion terms using unstructured grids and satisfies the conservation property. Numerical results are presented for a propagation of ionization waves in a rectangular domain.

Keywords Ionization waves · Finite volume methods · Error estimators

H. Amor
PRP-DGE/SEDRAN/BERAM, IRSN, 31, Avenue de la Division Leclerc,
92260 Fontenay Aux Roses, France
e-mail: hanen.amor@irsn.fr

F. Benkhaldoun · T. Ghoudi · I. Kissami (✉)
LAGA, Université Paris 13, Sorbonne Paris Cité, 99 Av J.B. Clement,
93430 Villetaneuse, France
e-mail: kissami.imad@lipn.univ-paris13.fr

F. Benkhaldoun
e-mail: fayssal@math.univ-paris13.frr

T. Ghoudi
e-mail: ghoudi@math.univ-paris13.frr

M. Seaid
School of Engineering and Computing Sciences, University of Durham,
Durham, UK
e-mail: m.seaid@durham.ac.uk

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-319-57394-6_57

1 Introduction

Ionization waves occur when non-ionized matter is exposed to a high intensity electric field and they appear in various forms depending on the electric field and on the pressure and volume of the medium. Streamers are among the active area of applications for ionization waves. Among these applications one can mention the treatment of contaminated media, see for instance [3] and references therein. The governing equations vary from one application to another but the common link between all these models is the presence of both convective and diffusive differential operators. In the present study a transient convection-diffusion equation is considered for the electron density and a steady diffusion equation is solved for the electric potential. The model is coupled through the electric field and a set of empirical equations.

Numerical simulation of ionization waves often presents difficulties due to their nonlinear form, presence of the source terms, coupling between the electron density and electric field. In addition, the difficulty in these models comes from the presence of both diffusion and convection terms which need spatial discretization on the same elements. A finite volume method has been successfully applied to these models in [1], where a strategy of dynamic mesh adaptation is implemented in order to capture the very stiff phenomenon of streamer discharge ignition and propagation. In [1] the adaptation is based on an empirical physical criteria which can be the gradient of the electron density, the drift velocity or a mix of both. In the present study we use the same method but choosing a new criteria for mesh adaptation in which a rigorous mathematical a posteriori error estimate is implemented. Note that all the remaining part of our former adaptive code algorithm stays the same. Although the system is fully coupled, the adaptation procedure is based only on the solution of Poisson problem related to the electric potential in the system. This gradient of the potential gives the electric field which is used to define the drift velocity in the transport equation of electron density. Numerical results are presented for a test example of propagation of an ionization wave in a rectangular domain subject to an electric field created by a difference in potentials between its walls. Results presented in this paper demonstrate the performance of the proposed adaptive finite volume method for this class of ionization wave front propagations.

2 Equations for Ionization Wave

In the current study we are interested in a simple two-dimensional model for negative streamer investigated in [3] among others. In what follows, we use boldface notation to denote vectors. Hence, for the electron density n_e, the ion density n_i, the electric potential V and the electric field $\mathbf{E}$, we solve the following equations

$$\begin{aligned} \frac{\partial n_e}{\partial t} + \operatorname{div}\left(n_e \mathbf{v}_e - D_e \nabla n_e\right) &= S_e, \\ \frac{\partial n_i}{\partial t} &= S_i^+, \\ -\operatorname{div}\left(\varepsilon \nabla V\right) &= f = e\left(n_i - n_e\right), \\ \mathbf{E} &= -\nabla V, \end{aligned} \tag{1}$$

where $\mathbf{v}_e$ is the electron drift velocity, D_e the diffusive coefficient, ε the dielectric constant, e the electron charge, and S_e and S_i^+ are source terms. The electron drift velocity $\mathbf{v}_e$ is a function of the electric field $\mathbf{E}$ and depends on the ratio $||\mathbf{E}||/N$, with N the neutral gas density. We refer to [1] for the definition of the rectangular computational domain Ω, initial and boundary conditions and the form of the physical parameters. Note that Eq. (1) have to be solved in a time interval [0, 57 ns].

3 Adaptive Finite Volume Methods

To discretize system (1) in space, we use a cell-centered finite volume method for which an upwind method is used for the convective terms and a Diamond scheme is used for the diffusive terms. In the current study, details on these techniques have been kept minimal and for a detailed formulation of these methods we refer the reader to [1]. Thus, a spatial discretization of the equations of electron and ion densities in system (1) yields the following system of ordinary differential equations (ODE) related to the unknowns on a given cell T_i (triangle here):

$$\begin{aligned} \frac{dn_{e_i}}{dt} &= -\frac{1}{\mu(T_i)} \sum_{j=1}^{m} \oint_{\sigma_{ij}} \left(n_{e_{ij}} \mathbf{v}_{e_{ij}} \cdot \mathbf{n}_{ij} ds - D_{e_{ij}} \nabla n_{e_{ij}} \cdot \mathbf{n}_{ij} ds\right) + S_{e_i}, \\ \frac{dn_{i_i}}{dt} &= S_{i_i}^+, \end{aligned} \tag{2}$$

where $\mu(T_i)$ is the volume of the triangle T_i, m the number of faces of the cell i, $\mathbf{n}_{ij}$ and $D_{e_{ij}}$ are respectively the outward unit normal vector and the diffusion coefficient on the face σ_{ij} between the cells T_i and T_j, and ds is its differential length. Other variables denoted by the subscripts ij represent variables on the face σ_{ij}. Note that the convective flux term in (2) is simply approximated by an upwind scheme and to discretize the Poisson problem for the electric potential in (1) and the diffusion term in (2) we use a finite volume approximation given by

$$\sum_{j=1}^{m} \oint_{\sigma_{ij}} \varepsilon_{ij} \nabla V_{ij} \cdot \mathbf{n}_{ij} ds \approx \sum_{j=1}^{m} \varepsilon_{ij} |\sigma_{ij}| \nabla_{|_{\sigma_{ij}}} V_{ij} \cdot \mathbf{n}_{ij} = \mu(T_i) f_i,$$

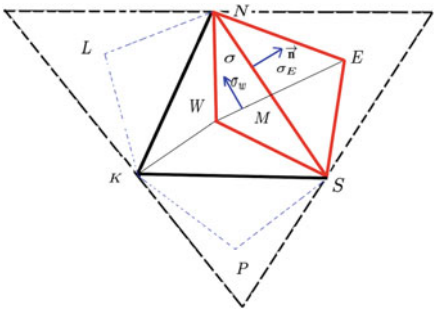


Fig. 1 Illustration of control volumes and diamond cell used in the space discretization

where f is the source term in the equation of electric potential in (1). To reconstruct the flux terms across a face σ_{ij} we consider the approach proposed in [1]. First we consider control volumes in Fig. 1. To simplify, the main cell T_i is now the west cell W and a given neighbor T_j is an east cell E. We denote by σ the edge separating the two cells W and E. Then we use a conservative reconstruction such that the flux F_{WE} between the cells W and E is equal to the flux F_{EW} between the cells E and W ($F_\sigma = F_{WE} = F_{EW}$). This yields

$$F_{WE} = -\varepsilon_{WE}\nabla_{WE}V \cdot |\sigma|\mathbf{n}, \qquad F_{EW} = -\varepsilon_{EW}\nabla_{EW}V \cdot |\sigma|(-\mathbf{n}),$$

where the gradients $\nabla_{WE}V$ and $\nabla_{EW}V$ are assumed to be constants on the two half-diamond cells (NWS) and (SEN) (see Fig. 1), V_M is the unknown value of V in the center M of the face (SN). Thus, one obtains (note that here $n_W = n_E = n_{WE}$):

$$\begin{aligned} \nabla_{WE}V &= \frac{1}{2|NWS|}\left((V_N - V_S)|\sigma_W|\mathbf{n_W} + (V_W - V_M)|\sigma|\mathbf{n}\right) \\ \nabla_{EW}V &= \frac{1}{2|SEN|}\left((V_N - V_S)|\sigma_E|\mathbf{n_E} + (V_E - V_M)|\sigma|\mathbf{n}\right) \end{aligned} \tag{3}$$

where V_N and V_S are the values of V at the vertices N and S, respectively. These values are computed by the least squares method

$$V_N = \sum_{p=1}^{r(N)} \alpha_p(N)V_p, \qquad V_S = \sum_{p=1}^{r(S)} \alpha_p(S)V_p,$$

where V_p is the value of V in the cell T_p, $r(N)$ is the number of cells surrounding vertex N, $\alpha_p(N)$ are weights associated with the least square method. Similarly, $r(S)$ and $\alpha_p(S)$ are the number of cells and weights for the vertex S.
To conclude, we impose the value of V_M such that $F_{WE} = F_{EW}$.
Finally, the convective flux is computed by a simple upwind scheme.
Note that the above finite volume discretization is only first-order accurate. A second-order accuracy can be achieved by considering the MUSCL and limiters techniques,

see details on these reconstructions in [1]. The time integration of the semi-discrete system (1) is carried out using a second-order explicit Runge-Kutta scheme. Because these schemes are explicit the time step has to satisfy the canonical CFL condition to guarantee the stability of the method. The implementation of this scheme for solving (1) is straightforward and it is omitted here, we only give a schematic view of the global algorithm:

Algorithm 1: Algorithm of code ADAPT for Streamer Propagation

```
1  W = (ne, ni);
2  Read mesh data;
3  Initialize conditions and create constants;
4  for each time iteration do
5      if the physical solution has evolved enough;
6        Compute Error Estimates and Adaptation Criteria;
7        Adapt the mesh: Refine where necessary and coarsen where necessary;
8        Construct matrix of linear system;
9      end if;
10     Compute Right-Hand-Side;
11     Solve linear system using MUMPS;
12     Compute fluxes of convection, diffusion and source term;
13     Update solution: W^n ⟶ W^{n+1};
14     Apply boundary conditions;
15 end
16 Save results ;
```

4 Error Estimates for Mesh Adaptation

To improve the performance of the finite volume method we incorporate a dynamics mesh adaptation using error estimates for the Poisson problem on electric potential in (1), written in a compact form as

$$-\operatorname{div}(\varepsilon \nabla V) = f, \tag{4}$$

where $f = e\,(n_i - n_e)$. We consider a posteriori error estimators developed by Vohralik in [4] for the diffusion equation (4) in complex geometries based on a conforming flux reconstruction using the approximate solution. Note that here the primal and dual meshes are the same ($\mathcal{T}_h = \mathcal{D}_h$) since we are using a cell-centred method. The sub-triangulation $\mathcal{S}_h$ of [4] is obtained by subdividing a given cell $W = (SNK)$ into 6 sub-triangles using the 3 vertices, the center of the cell W, and the 3 midpoints of the cell edges (see Fig. 1). Next, we solve the diffusion (4) using the cell-centered finite volume method based on the reconstruction of a discrete gradient at the mesh interfaces to compute the diffusive fluxes. Thus we choose $t_h \in H(div, \Omega)$, such

that:

$$\int_D \nabla \cdot t_h dx = \int_D f(x) dx,$$

where D is a given cell (triangle of the mesh here), t_h is chosen in the Raviart-Thomas approximation space of lowest degree per quarter of diamond, and the estimation of the a posteriori error with the energy norm is defined by [4]

$$|||V - V_h||| \leq \left(\sum_{D \in \mathcal{D}_h} (\eta_{R,D} + \eta_{DF,D})^2 \right)^{\frac{1}{2}}, \tag{5}$$

where the energy norm is given by

$$|||V - V_h||| = ||\varepsilon^{\frac{1}{2}} \nabla (V - V_h)||.$$

The residual error estimate $\eta_{R,D}$ and the flux error estimate $\eta_{DF,D}$ in (5) are defined for each element $D \in \mathcal{D}_h$ as

$$\eta_{R,D} = ||\varepsilon^{\frac{1}{2}} \nabla V_h + \varepsilon^{-\frac{1}{2}} t_h||_D, \qquad \eta_{DF,D} = m_{D,\varepsilon} ||f - \nabla \cdot t_h||_D,$$

where $m_{D,\varepsilon}$ the volume of the element D, and ∇V_h is defined by (3).

Here we choose the method of direct prescription to get t_h. Hence, on each one of the 6 sub-triangles of the cell D, which happens to be quarters of a diamond, we define

$$t_h \cdot n_\sigma = -\varepsilon \nabla V_h \cdot n_\sigma,$$

For instance on the sub-triangle NWM, the following system is solved:

$$\begin{aligned} t_h \cdot \mathbf{n}_{NW} &= -0.5\varepsilon \nabla^-_{NW}(V) \cdot \mathbf{n}_{NW} - 0.5\varepsilon \nabla^+_{NW}(V) \cdot \mathbf{n}_{NW}, \\ t_h \cdot \mathbf{n}_{WM} &= -\varepsilon \nabla_{WM}(V) \cdot \mathbf{n}_{WM} \\ t_h \cdot \mathbf{n}_{MN} &= -\varepsilon \nabla_{MN}(V) \cdot \mathbf{n}_{MN} \end{aligned}$$

where we note for a given side (AB), $\mathbf{n}_{AB}$ the outward normal to the side, $\nabla^-_{AB}(V)$ and $\nabla^+_{AB}(V)$, the values of the approximate gradient on left and right quarter of diamond sharing the side (AB).

5 Numerical Example

To assess the numerical performance of the proposed finite volume method we solve the test example of discharge propagation in a homogeneous electric field described in [1]. Hence, we solve Eq. (1) in the rectangular domain $[0, 1] \times [0, 0.5]$ subjected

to a plane anode with $V = 25000\,\mathrm{v}$ on the left and a plane cathode with $V = 0\,\mathrm{v}$ on the right whereas periodic boundary conditions are imposed on the upper and lower walls. Initially,

$$n_e(x, y, 0) = n_i(x, y, 0) = 10^{16} \times e^{-\frac{(x-0.2)^2 + (y-0.25)^2}{\sigma^2}} + 10^9,$$

with $\sigma = 0.01$. All our simulations are carried on a Pentium IV 2.66 GHz having 1 GB of RAM running C^{++} codes. For the mesh adaptation we used our technique used for combustion simulations in [2].

In Fig. 2 we present the numerical results obtained using an error estimate for the adaptive meshes, distribution of electron densities n_e and the velocity fields $\mathbf{v}_e$ at three different times. It is clear that the initial Gaussian pulse for the electron and the ion densities creates a disturbance in the electric field which is necessary for formation of the discharge. The proposed error estimate for the finite volume method has clearly resolved this test example and accurately captured the physical features.

To further illustrate these effects we display Fig. 3 with cross-sections The results on the very fine mesh are assumed as reference solutions and relative errors are calculated and summarized in Table 1. For the considered test example the adaptive finite volume method is highly accurate and efficient, compare the errors and computational times in Table 1.

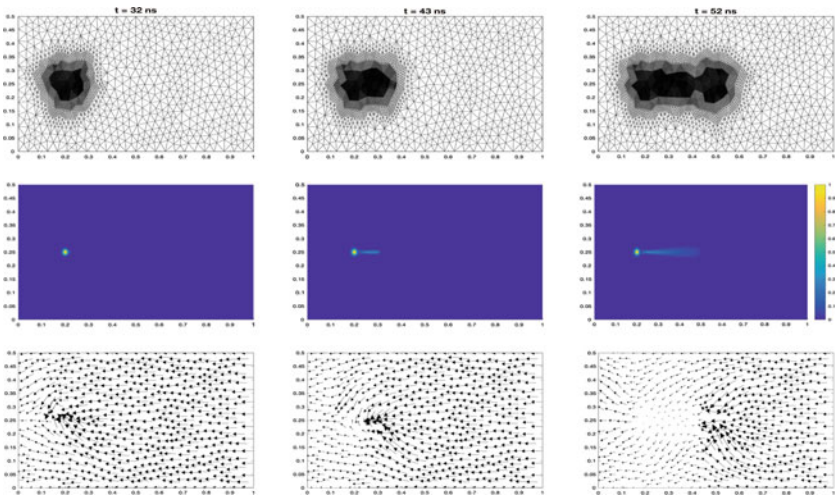


Fig. 2 Adaptive meshes (first row), electron density distributions (second row) and velocity fields (third row) for propagation of a discharge in the rectangular domain at times $t = 35$, 47 and 57 ns

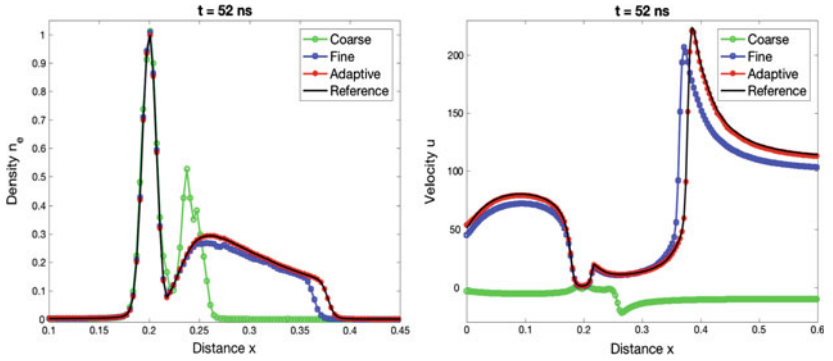


Fig. 3 Cross-section at $x = 0.25$ cm of the electron density (*left*) and velocity (*right*) at time 52 ns

Table 1 Mesh statistics, relative errors and computational times for propagation of a discharge in the rectangular domain at time $t = 52$ ns

	# of cells	# of nodes	Error in n_e	Error in u	CPU time (in seconds)
Coarse	43674	22061	0.5881	1.1296	2700
Fine	276992	139057	0.0863	1.0976	24300
Adaptive	35372	17827	0.0024	0.0124	10854
Reference	1107968	555105	—	—	214000

6 Conclusion

We have proposed a new mesh adaptation for finite volume solution of ionization waves in planar electric field. The finite volume method uses an Upwind reconstruction for the convective terms and a cell-centered method for the diffusion terms. The criteria for mesh adaptation is based only on an error estimate of the diffusion equation for the electric potential. Numerical results have been presented for a test problem of a propagation of ionization waves in a rectangular domain. Comparisons to simulations on fixed meshes show that the current adaptive finite volumes scheme offers a potential to produce highly accurate solutions with low computational cost. Note that the originality of the work presented here is confirmed by the obtention of precise and reliable results for discharge propagation using dynamic mesh refinement which is based on a rigorous mathematical criterion. This criterion results from solving elliptic Poisson problem alone and will be supplemented in the future by a more global criterion, taking into account the a posteriori error for the transport terms as well as the error introduced by the time integration scheme.

References

1. Benkhaldoun, F., Fot, J., Hassouni, K., Karel, J.: Simulation of planar ionization wave front propagation on an unstructured adaptive grid. J. Comput. Appl. Math. **236**, 4623–4634 (2012)
2. Elmahi, I., Benkhaldoun, F., Borghi, R., Raghay, S.: Ignition of fuel issuing from a porous cylinder located adjacent to a heated wall: a numerical study. Combust. Theor. Model. **8**, 789–809 (2004)
3. Montijn, C., Hundsdorfer, W., Ebert, U.: An adaptive grid refinement strategy for the simulation of negative streamers. J. Comput. Phys. **219**, 801–835 (2006)
4. Vohralík, M.: Guaranteed and fully robust a posteriori error estimates for conforming discretizations of diffusion problems with discontinuous coefficients. J. Sci. Comput. **46**, 397–438 (2011)

Erratum to: Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems

Clément Cancès and Pascal Omnes

Erratum to:

C. Cancès and P. Omnes (Eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, https://doi.org/10.1007/978-3-319-57394-6

The original version of the book was inadvertently published with incorrect table, figure position and equation, which have been corrected as follows

In Chap. 9, Table 1, incomplete table lines have been corrected and the text "N/A" has been center aligned by merging all the empty cells.

In Chap. 37, the position of sub-figures of Fig. 4 has been interchanged.

In Chap. 54, the equation $b_{\vec{u},T}(\underline{u}_T, \underline{v}_T) := (u_T, \mathrm{G}_{\vec{u},T}^{k+1}\underline{u}_T, v_T)_T + s_{\vec{u},T}(\underline{u}_T, \underline{v}_T)$ has been corrected as $-b_{\vec{u},T}(\underline{u}_T, \underline{v}_T) := -(u_T, \mathrm{G}_{\vec{u},T}^{k+1}\underline{v}_T)_T + s_{\vec{u},T}(\underline{u}_T, \underline{v}_T)$

The updated online version of the book can be found at
https://doi.org/10.1007/978-3-319-57394-6_9
https://doi.org/10.1007/978-3-319-57394-6_37
https://doi.org/10.1007/978-3-319-57394-6_54

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, https://doi.org/10.1007/978-3-319-57394-6_58

Author Index

A
Abbate, Emanuela, 227
Abgrall, Rémi, 257
Ahusborde, Etienne, 407
Aïssiouene, Nora, 219
Amaziane, Brahim, 407
Ambartsumyan, Ilona, 377
Amor, Hanen, 547
Anthonissen, Martijn, 437, 467
Audusse, Emmanuel, 209

B
Bürger, Raimund, 189
Bacigaluppi, Paola, 257
Basara, Branislav, 81
Beaude, Laurence, 317
Behrens, Jörn, 237, 247
Beisiegel, Nicole, 237
Beltman, René, 467
Benkhaldoun, Fayssal, 137, 547
Birgle, Nabil, 387
Botti, Michele, 537
Boubekeur, Mohamed, 137
Boukili, Hamza, 109
Bourdarias, Christian, 33, 101
Boyaval, Sébastien, 163, 477
Brenner, Konstantin, 317
Brenner, Pierre , 295
Bristeau, Marie-Odile, 219

C
Campos Pinto, Martin, 265
Cancès, Clément, 327
Castro, Manuel J., 23, 119
Chalons, Christophe, 63, 71
Chave, Florent, 517
Cheng, Hanz Martin, 367
Chernyshenko, Alexey, 527
Colas, Clément, 53
Corot, Théo, 43
Coulette, David, 171

D
Daude, Frédéric, 3
De Laage de Meux, Benoît, 33
Demay, Charles, 33
Di Pietro, Daniele A., 517, 537
Do, Minh Hieu, 209
Droniou, Jérôme, 367
Duvigneau, Régis, 71

E
El Ossmani, Mustapha, 407
Emmanuel, Franck, 171
Enchéry, Guillaume, 477
Erath, Christoph, 337
Escalante, Cipriano, 119

F
Ferrand, Martin, 53
Fetzer, Thomas, 347
Fiorini, Camilla, 71
Flemisch, Bernd, 347, 417
Frolkovič, Peter, 81
Fuhrmann, Jürgen, 397, 497

G
Gallardo, José M., 23
Galon, pascal, 3
Gerbi, Stéphane, 33, 101
Gersbacher, Christoph, 447
Ghidaglia, Jean-Michel, 155
Ghoudi, tarek, 547
Gläser, Dennis, 417
Glitzky, Annegret, 397
Godlewski, Edwige, 219
Goudon, Thierry, 91
Goutal, Nicole, 275

C. Cancès and P. Omnes (eds.), *Finite Volumes for Complex Applications VIII—Hyperbolic, Elliptic and Parabolic Problems*, Springer Proceedings in Mathematics & Statistics 200, DOI 10.1007/978-3-662-54033-6

Granjeon, Didier, 327
Grapsas, Dionysis, 285
Grüninger, Christoph, 347
Guhlke, Clemens, 497

H
Hérard, Jean-Marc, 3, 33, 53, 109
Hahn, Jooyoung, 81
Helluy, Philippe, 171
Helmig, Rainer, 347, 417
Herbin, Raphaèle, 285
Hubert, Florence, 13

I
Iampietro, david, 3
Iollo, Angelo, 227

J
Jeschke, Anja, 247

K
Kane, Birane, 447
Khattatov, Eldar, 377
Kissami, imad, 547
Klöfkorn, Robert, 447
Koren, Barry, 467
Kröker, Ilja, 189
Kramarenko, Vasiliy, 507
Kumar, Nikhil, 457
Kumar, Sarvesh, 307
Kumbaro, Anela, 155
Kwant, Ruben, 437

L
Latché, Jean-Claude, 285
Le, Minh Hoang, 275
Liero, Matthias, 397
Linke, Alexander, 457
Lipnikov, Konstantin, 427
Llobell, Julie, 91
Lopez, Simon, 317
Lteif, Ralph, 101
Lukáčová-Medvid'ová,Mária, 179

M
Mangeney, Anne, 219
Marche, Fabien, 517
Marquina, Antonio, 23
Masson, Roland, 317, 387
Mehrenberger, Michel, 171
Melis, Ward, 145
Mikula, Karol, 81
Minjeaud, Sebastian, 91
Morales de Luna, TomÁs, 119

N
Navoret, Laurent, 171
Nikitin, Kirill, 507

O
Ohlberger, Mario, 357
Olshanskii, Maxim, 527
Omnes, Pascal, 209

P
Parés, Carlos, 219
Penel, Yohan, 209
Peton, Nicolas, 327
Pont, Grégoire, 295
Puppo, Gabriella, 227

Q
Quaglia, Laurent, 487

R
Rey, Thomas, 145
Rosemeier, Juliane, 179
Ruiz Baier, Ricardo, 307

S
Sainte-Marie, Jacques, 219
Samaey, Giovanni, 145
Sanchez, Riad, 477
Sandilya, Ruchi, 307
Schindler, Felix, 357
Schneider, Martin, 417
Schorr, Robert, 337
Seaid, Mohammed, 547
Smai, Farid, 321
Sochala, Pierre, 537
Spichtinger, Peter, 179
Stauffert, Maxime, 63
Svyatskiy, Daniil, 427

T
Ten Thije Boonkkamp, Jan, 437, 457
Tesson, Rémi, 13
Tokareva, Svetlana, 127
Toro, Eleuterio, 127
Tran, Quang-Huy, 327, 477
Trenty, Laurent, 387

U
Ung, Philippe, 275

V
Vassilevski, Yuri, 507, 527
Vater, Stefan, 237, 247

W
Wiebe, Bettina, 179
Wolf, Sylvie, 327

Y
Yotov, Ivan, 377

Z
Zakerzadeh, Hamed, 199
Zhang, Lei, 155