



Wave Propagation in Fluids

Models and Numerical Techniques

Second Edition

Vincent Guinot

ISTE

WILEY

Wave Propagation in Fluids

Wave Propagation in Fluids

Models and Numerical Techniques

Second Edition

Vincent Guinot



First edition published 2008 by ISTE Ltd and John Wiley & Sons, Inc.
Second updated and revised edition published 2010 in Great Britain and the United States by ISTE Ltd
and John Wiley & Sons, Inc.

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms and licenses issued by the CLA. Enquiries concerning reproduction outside these terms should be sent to the publishers at the undermentioned address:

ISTE Ltd
27-37 St George's Road
London SW19 4EU
UK

www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA

www.wiley.com

© ISTE Ltd 2008, 2010

The rights of Vincent Guinot to be identified as the author of this work have been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

Library of Congress Cataloging-in-Publication Data

Guinot, Vincent.

Wave propagation in fluids : models and numerical techniques / Vincent Guinot. -- 2nd ed., updated and rev.

p. cm.

Includes bibliographical references and index.

ISBN 978-1-84821-213-8

1. Fluids--Mathematics. 2. Wave-motion, Theory of. I. Title.

QA927.G85 2010

532'.05930151--dc22

2010027124

British Library Cataloguing-in-Publication Data

A CIP record for this book is available from the British Library

ISBN 978-1-84821-213-8

Printed and bound in Great Britain by CPI Antony Rowe, Chippenham and Eastbourne.



Table of Contents

Introduction	xv
Chapter 1. Scalar Hyperbolic Conservation Laws in One Dimension of Space	1
1.1. Definitions	1
1.1.1. Hyperbolic scalar conservation laws	1
1.1.2. Derivation from general conservation principles	3
1.1.3. Non-conservation form	6
1.1.4. Characteristic form – Riemann invariants	7
1.2. Determination of the solution	9
1.2.1. Representation in the phase space	9
1.2.2. Initial conditions, boundary conditions	12
1.3. A linear law: the advection equation	14
1.3.1. Physical context – conservation form	14
1.3.2. Characteristic form	16
1.3.3. Example: movement of a contaminant in a river	17
1.3.4. Summary	21
1.4. A convex law: the inviscid Burgers equation	21
1.4.1. Physical context – conservation form	21
1.4.2. Characteristic form	22
1.4.3. Example: propagation of a perturbation in a fluid	24
1.4.4. Summary	28
1.5. Another convex law: the kinematic wave for free-surface hydraulics	28
1.5.1. Physical context – conservation form	28
1.5.2. Non-conservation and characteristic forms	30
1.5.3. Expression of the wave speed	31
1.5.4. Particular case: flow in a rectangular channel	34

1.5.5. Summary	35
1.6. A non-convex conservation law: the Buckley-Leverett equation	35
1.6.1. Physical context – conservation form	35
1.6.2. Characteristic form	38
1.6.3. Example: decontamination of an aquifer	40
1.6.4. Summary	41
1.7. Advection with adsorption/desorption	42
1.7.1. Physical context – conservation form	42
1.7.2. Characteristic form	45
1.7.3. Summary	47
1.8. Summary of Chapter 1	47
1.8.1. What you should remember	47
1.8.2. Application exercises	48
Chapter 2. Hyperbolic Systems of Conservation Laws in One Dimension of Space	53
2.1. Definitions	53
2.1.1. Hyperbolic systems of conservation laws	53
2.1.2. Hyperbolic systems of conservation laws – examples	55
2.1.3. Characteristic form – Riemann invariants	57
2.2. Determination of the solution	59
2.2.1. Domain of influence, domain of dependence	59
2.2.2. Existence and uniqueness of solutions – initial and boundary conditions	61
2.3. A particular case: compressible flows	63
2.3.1. Definition	63
2.3.2. Conservation form	63
2.3.3. Characteristic form	66
2.3.4. Physical interpretation	67
2.4. A linear 2×2 system: the water hammer equations	68
2.4.1. Physical context – assumptions	68
2.4.2. Conservation form	70
2.4.3. Characteristic form – Riemann invariants	75
2.4.4. Calculation of the solution	79
2.4.5. Summary	83
2.5. A nonlinear 2×2 system: the Saint Venant equations	84
2.5.1. Physical context – assumptions	84
2.5.2. Conservation form	85
2.5.3. Characteristic form – Riemann invariants	91
2.5.4. Calculation of solutions	100
2.5.5. Summary	107

2.6. A nonlinear 3×3 system: the Euler equations	108
2.6.1. Physical context – assumptions	108
2.6.2. Conservation form	109
2.6.3. Characteristic form – Riemann invariants	113
2.6.4. Calculation of the solution	117
2.6.5. Summary	121
2.7. Summary of Chapter 2	122
2.7.1. What you should remember	122
2.7.2. Application exercises	123
Chapter 3. Weak Solutions and their Properties	131
3.1. Appearance of discontinuous solutions	131
3.1.1. Governing mechanisms	131
3.1.2. Local invalidity of the characteristic formulation – graphical approach	134
3.1.3. Practical examples of discontinuous flows	136
3.2. Classification of waves	138
3.2.1. Shock wave	138
3.2.2. Rarefaction wave	140
3.2.3. Contact discontinuity	140
3.2.4. Mixed/compound wave	141
3.3. Simple waves	142
3.3.1. Definition and properties	142
3.3.2. Generalized Riemann invariants	143
3.4. Weak solutions and their properties	144
3.4.1. Definitions	144
3.4.2. Non-equivalence between the formulations	145
3.4.3. Jump relationships	146
3.4.4. Non-uniqueness of weak solutions	148
3.4.5. The entropy condition	152
3.4.6. Irreversibility	154
3.4.7. Approximations for the jump relationships	156
3.5. Summary	157
3.5.1. What you should remember	157
3.5.2. Application exercises	158
Chapter 4. The Riemann Problem	161
4.1. Definitions – solution properties	161
4.1.1. The Riemann problem	161
4.1.2. The generalized Riemann problem	162
4.1.3. Solution properties	163

4.2. Solution for scalar conservation laws	165
4.2.1. The linear advection equation	165
4.2.2. The inviscid Burgers equation	166
4.2.3. The Buckley-Leverett equation.	168
4.3. Solution for hyperbolic systems of conservation laws.	173
4.3.1. General principle	173
4.3.2. Application to the water hammer problem: sudden valve failure	174
4.3.3. Free surface flow: the dambreak problem	177
4.3.4. The Euler equations: the shock tube problem	183
4.4. Summary	189
4.4.1. What you should remember.	189
4.4.2. Application exercises	190
Chapter 5. Multidimensional Hyperbolic Systems	193
5.1. Definitions	193
5.1.1. Scalar laws.	193
5.1.2. Two-dimensional hyperbolic systems	195
5.1.3. Three-dimensional hyperbolic systems	196
5.2. Derivation from conservation principles.	197
5.3. Solution properties	200
5.3.1. Two-dimensional hyperbolic systems	200
5.3.2. Three-dimensional hyperbolic systems	206
5.4. Application: the two-dimensional shallow water equations	208
5.4.1. Governing equations	208
5.4.2. The secant plane approach	213
5.4.3. Interpretation – determination of the solution	218
5.5. Summary	221
5.5.1. What you should remember.	221
5.5.2. Application exercises	221
Chapter 6. Finite Difference Methods for Hyperbolic Systems	223
6.1. Discretization of time and space	223
6.1.1. Discretization for one-dimensional problems	223
6.1.2. Multidimensional discretization	224
6.1.3. Explicit schemes, implicit schemes	226
6.2. The method of characteristics (MOC)	227
6.2.1. MOC for scalar hyperbolic laws	227
6.2.2. The MOC for hyperbolic systems of conservation laws	235
6.2.3. Application examples	240
6.3. Upwind schemes for scalar laws	244
6.3.1. The explicit upwind scheme (non-conservative version).	244
6.3.2. The implicit upwind scheme (non-conservative version)	245

6.3.3. Conservative versions of the implicit upwind scheme	247
6.3.4. Application examples	249
6.4. The Preissmann scheme	250
6.4.1. Formulation	250
6.4.2. Estimation of nonlinear terms – algorithmic aspects	253
6.4.3. Numerical applications	254
6.5. Centered schemes	260
6.5.1. The Crank-Nicholson scheme	260
6.5.2. Centered schemes with Runge-Kutta time stepping.	261
6.6. TVD schemes	263
6.6.1. Definitions	263
6.6.2. General formulation of TVD schemes.	264
6.6.3. Harten’s and Sweby’s criteria	266
6.6.4. Classical limiters	268
6.6.5. Computational example	269
6.7. The flux splitting technique	271
6.7.1. Principle of the approach	271
6.7.2. Application to classical schemes	274
6.8. Conservative discretizations: Roe’s matrix	280
6.8.1. Rationale and principle of the approach	280
6.8.2. Expression of Roe’s matrix	281
6.9. Multidimensional problems	284
6.9.1. Explicit alternate directions	284
6.9.2. The ADI method	286
6.9.3. Multidimensional schemes	288
6.10. Summary	289
6.10.1. What you should remember	289
6.10.2. Application exercises	291
Chapter 7. Finite Volume Methods for Hyperbolic Systems	293
7.1. Principle	293
7.1.1. One-dimensional conservation laws	293
7.1.2. Multidimensional conservation laws	295
7.1.3. Application to the two-dimensional shallow water equations	297
7.2. Godunov’s scheme	299
7.2.1. Principle	299
7.2.2. Application to the scalar advection equation	301
7.2.3. Application to the inviscid Burgers equation.	305
7.2.4. Application to the water hammer equations	308
7.3. Higher-order Godunov-type schemes	313
7.3.1. Rationale and principle	313
7.3.2. Example: the MUSCL scheme	316

7.4. EVR approach	319
7.4.1. Principle of the approach	319
7.4.2. Application to the one-dimensional shallow water equations	323
7.5. Summary	326
7.5.1. What you should remember	326
7.5.2. Application exercises	327
Chapter 8. Finite Element Methods for Hyperbolic Systems	329
8.1. Principle for one-dimensional scalar laws	329
8.1.1. Weak form	329
8.1.2. Discretization of space and time	330
8.1.3. Classical shape and test functions	335
8.2. One-dimensional hyperbolic systems	340
8.2.1. Weak formulation	340
8.2.2. Application to the non-conservation form	341
8.3. Extension to multidimensional problems	344
8.3.1. Weak form of the equations	344
8.3.2. Discretization of space	345
8.3.3. Classical shape and test functions	345
8.4. Discontinuous Galerkin techniques	347
8.4.1. Principle of the method	347
8.4.2. Legendre polynomial-based reconstruction	349
8.4.3. Limiting	351
8.4.4. Runge-Kutta time stepping	353
8.5. Application examples	354
8.5.1. The linear advection equation	354
8.5.2. The inviscid Burgers equation	359
8.6. Summary	368
8.6.1. What you should remember	368
8.6.2. Application exercises	369
Chapter 9. Treatment of Source Terms	371
9.1. Introduction	371
9.2. Problem position	372
9.2.1. Example 1: the water hammer equations	372
9.2.2. Example 2: the shallow water equations	374
9.2.3. Stationary solution and C–property	376
9.3. Source term upwinding techniques	377
9.3.1. Principle	377
9.3.2. Application example 1: the water hammer equations	380
9.3.3. Application example 2: the shallow water equations with HLL solver	382

9.4. The quasi-steady wave algorithm	386
9.4.1. Principle	386
9.4.2. Application to the water hammer equations	387
9.4.3. Application to the one-dimensional shallow water equations	387
9.5. Balancing techniques	390
9.5.1. Well-balancing	390
9.5.2. Hydrostatic pressure reconstruction for free surface flow	393
9.5.3. Auxiliary variable-based balancing	395
9.6. Computational example	403
9.7. Summary	408
Chapter 10. Sensitivity Equations for Hyperbolic Systems	411
10.1. Introduction	411
10.2. Forward sensitivity equations for scalar laws	413
10.2.1. Derivation for continuous solutions	413
10.2.2. Conservation, non-conservation and characteristic forms	415
10.2.3. Extension to discontinuous solutions	416
10.2.4. Solution of the Riemann problem	418
10.3. Forward sensitivity equations for hyperbolic systems	422
10.3.1. Governing equations	422
10.3.2. Non-conservation and characteristic forms	424
10.3.3. The Riemann problem	426
10.3.4. Application example: the one-dimensional shallow water sensitivity equations	427
10.4. Adjoint sensitivity equations	435
10.4.1. Introduction	435
10.4.2. Adjoint models for scalar laws	435
10.5. Finite volume solution of the forward sensitivity equations	441
10.5.1. Introduction	441
10.5.2. Discretization	442
10.5.3. A modified HLL Riemann solver for sensitivity solutions	443
10.5.4. Application example: the one-dimensional shallow water equations	446
10.6. Summary	447
Chapter 11. Modeling in Practice	449
11.1. Modeling software	449
11.1.1. Introduction	449
11.1.2. Conservation	450
11.1.3. Solution monotony	453
11.2. Mesh quality	454
11.3. Boundary conditions	459

11.3.1. Number and nature of boundary conditions	459
11.3.2. Prescribed discharge/flow velocity	460
11.3.3. Prescribed pressure/water level	461
11.3.4. Stage-discharge and pressure-discharge relationships.	463
11.4. Numerical parameters	464
11.4.1. Computational time step	464
11.4.2. Scheme centering parameters	465
11.4.3. Iteration control	465
11.5. Simplifications in the governing equations	466
11.5.1. Rationale	466
11.5.2. The Local Partial Inertia (LPI) technique	467
11.5.3. The Reduced Momentum Equation (RME) technique	468
11.5.4. Application examples.	469
11.6. Numerical solution assessment	472
11.6.1. Software solution accuracy	472
11.6.2. Assessing solution convergence	473
11.6.3. Consistency analysis – numerical diffusion and dispersion	474
11.6.4. Stability analysis – phase and amplitude portraits	476
11.7. Getting started with a simulation package	477
Appendix A. Linear Algebra	479
A.1. Definitions	479
A.2. Operations on matrices and vectors	480
A.2.1. Addition	480
A.2.2. Multiplication by a scalar.	481
A.2.3. Matrix product	481
A.2.4. Determinant of a matrix.	482
A.2.5. Inverse of a matrix	482
A.3. Differential operations using matrices and vectors	483
A.3.1. Differentiation	483
A.3.2. Jacobian matrix.	483
A.4. Eigenvalues, eigenvectors	483
A.4.1. Definitions	483
A.4.2. Example	484
Appendix B. Numerical Analysis	487
B.1. Consistency	487
B.1.1. Definitions.	487
B.1.2. Principle of a consistency analysis	487
B.1.3. Numerical diffusion, numerical dispersion.	489
B.2. Stability	491
B.2.1. Definition	491

B.2.2. Principle of a stability analysis	492
B.2.3. Harmonic analysis of analytical solutions	494
B.2.4. Harmonic analysis of numerical solutions	497
B.2.5. Amplitude and phase portraits	499
B.2.6. Extension to systems of equations	501
B.3. Convergence	503
B.3.1. Definition	503
B.3.2. Lax's theorem	503
Appendix C. Approximate Riemann Solvers	505
C.1. The HLL and HLLC solvers	505
C.1.1. The HLL solver	505
C.1.2. The HLLC solver	508
C.2. Roe's solver	511
C.2.1. Principle	511
C.2.2. Algorithmic simplification	513
C.2.3. Entropy violation and fixes	514
C.2.4. Application example: the shallow water equations	514
C.3. The Lax-Friedrichs solver	515
C.4. Approximate-state solvers	516
C.4.1. Principle	516
C.4.2. Shock-based solvers	516
C.4.3. Rarefaction wave-based solvers	517
Appendix D. Summary of the Formulae	521
Bibliography	527
Index	537

Introduction

What is wave propagation?

In a kitchen or in a bathroom, the number of times we turn a tap every day is countless. So is the number of times we watch the liquid stream impacting the sink. The circular flow pattern where the fast and shallow water film diverging from the impact point changes into a deeper, bubbling flow is too familiar to deserve attention. Very few people looking at the circular, bubbling pattern – referred to as a hydraulic jump by hydraulics specialists – are aware that they are contemplating a shock wave.

Closing the tap too quickly may result in a thud sound. This is the audible manifestation of the well-known water hammer phenomenon, a train of pressure waves propagating in the metal pipes as fast as hundreds or thousands of meters per second. The water hammer phenomenon is known to cause considerable damage to hydropower duct systems or water supply networks under the sudden operation of valves, pumps or turbines. The sound is heard because the vibrations of the duct system communicate with the ambient atmosphere, and from there with the operator's ears.

Everyone has once thrown stones into a pond, watching the concentric ripples propagate on the surface. Less visible and much slower than the ripples is the moving groundwater that displaces a pollutant front in a journey that may last for years.

As ubiquitous and familiar as wave propagation may be, the phenomenon is often poorly understood. The reason why intuition so often fails to grasp the mechanisms of wave propagation may lie in the commonly shared, instinctive perception that waves are made of matter. This, however, is not true. In the example of the hydraulic jump in the sink, the water molecules move across an immobile

wave. In the example of the ripples propagating on the free surface of a pond, the waves travel while the water remains immobile.

Waves appear when an object or a system (e.g. the molecules in a fluid, a rigid metallic structure) reacts to a perturbation and transmits it to its neighbors. In many cases, as in the example of the water ripples, the initially perturbed system returns to its initial equilibrium state, while the waves keep propagating. In this respect, waves may be seen as information. The ripples propagating in a pond are a sign that the water molecules “inform” their neighbors that the equilibrium state has been perturbed. A sound is nothing other than information about a perturbation occurring in the atmosphere.

Numerical techniques for wave propagation simulation have been the subject of intensive research over the last 50 years. The advent of fast computers has led to the development of efficient numerical techniques. Engineers and consultants now use simulation software packages for wave propagation on a daily basis. Whether for the purpose of acoustics, aerodynamics, flood wave propagation or contaminant transport studies, computer-based simulation tools have become indispensable to almost all domains of engineering. Such tools, however, remain instruments operated by human beings to execute tedious, repetitive operations previously carried out by hand. They cannot, and hopefully never will, replace the expert’s judgment and experience. Human presence remains necessary for the sound assessment of the relevance and accuracy of modeling results. Such an assessment, however, is possible only provided that the very specific type of reasoning required for the correct understanding of wave propagation phenomena has been acquired.

The main purpose of this book is to contribute to a better understanding of wave propagation phenomena and the most commonly used numerical techniques for its simulation. The first three chapters deal with the physics and mathematics of wave propagation. Chapters 4, 5 and 10 provide insight into more theoretical notions, used in specific numerical techniques. Chapters 6 to 9 are devoted to finite difference, finite volume and finite element techniques. Chapter 11 is devoted to practical advice for the modeler. Basic notions of linear algebra and numerical methods are presented in Appendices A to C. The various formulae used in the present book are summarized in Appendix D.

What is the intended readership of this book?

This book is intended for students of professional and research master’s programs and those engaged in doctoral studies, the curriculum of which contains hydraulics and/or fluid mechanics-related subjects. Engineers and developers in the

field of fluid mechanics and hydraulics are also a potential target group. This book was written with the following objectives:

- (i) To introduce the physics of wave propagation, the governing assumptions and the derivation of the governing equations (in other words, the modeling process) in various domains of fluid mechanics. The application fields are as diverse as contaminant transport, open channel and free surface hydraulics, or aerodynamics.
- (ii) To explain how the behavior of the physical systems can be analyzed using very simple mathematical techniques, thus allowing practical problems to be solved.
- (iii) To introduce the main families of numerical techniques used in most simulation software packages. As today's practicing engineers cannot afford not to master modeling packages, a basic knowledge of the existing numerical techniques appears as an indispensable engineering skill.

How should this book be read?

Most of the chapters are made up of three parts:

- the first part of the chapter is devoted to the theoretical notions applied in the remainder of the chapter;
- the second part deals with the application of these theoretical notions to various hydraulics and fluid mechanics equations;
- the third part provides a summary of the key points developed in the chapter, as well as suggestions of application exercises.

The main purpose of the application exercises is to test the reader's ability to reuse the notions developed in the chapter and apply them to practical problems. The solutions to the exercises may be accessed at the following URL:
<http://vincentguinot.free.fr/waves/exercises.htm>.

Try to resist the temptation to read the solution immediately. Solving the exercise by yourself should be the primary objective. The solution text is provided only as an aid, in case you cannot find a way to start and for you to check the validity of your reasoning after completing the exercise.

Chapter 1

Scalar Hyperbolic Conservation Laws in One Dimension of Space

1.1. Definitions

1.1.1. *Hyperbolic scalar conservation laws*

A one-dimensional hyperbolic scalar conservation law is a Partial Differential Equation (PDE) that can be written in the form:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} = S \quad [1.1]$$

where t and x are respectively the time- and space-coordinates, U is the so-called conserved variable, F is the flux and S is the source term. Equation [1.1] is said to be the conservation, or divergent, form of the conservation law. The following definitions are used:

- the flux F is the amount of U that passes at the abscissa x per unit time due to the fact that U (also called the transported variable) is being displaced;
- the source term S is the amount of U that appears per unit time and per unit volume, irrespective of the amount transported via the flux F . If U represents the concentration in a given chemical substance, the source term may express degradation phenomena, or radioactive decay. S is positive when the conserved variable appears in the domain, negative if U disappears from the domain;

2 Wave Propagation in Fluids

– the conservation law is said to be scalar because it deals with only one dependent variable. When several equations in form [1.1] are satisfied simultaneously, the term “system of conservation laws” is used. Systems of conservation laws are dealt with in Chapter 2.

Only hyperbolic conservation laws are dealt with in what follows. The conservation law is said to be hyperbolic if the flux F is a function of U (and none of its derivatives) and, possibly, of x and t . Such a dependence is expressed in the form:

$$\left. \begin{aligned} F &= F(U, x, t) \\ S &= S(U, x, t) \end{aligned} \right\} \quad [1.2]$$

The function $F(U, x, t)$ is called the “flux function”.

NOTE.– The expression $F(U, x, t)$ in equation [1.2] indicates that F depends on U at the abscissa x at the time t and does not depend on such quantities as derivatives of U with respect to time or space. For instance, the following expression:

$$F = aU \quad [1.3]$$

is a permissible expression [1.2] for F , while the following, diffusion flux:

$$F = -D \frac{\partial U}{\partial x} \quad [1.4]$$

where D is the diffusion coefficient, does not yield a hyperbolic conservation law because the flux F is a function of the first-order derivative of U with respect to space.

In the case of a zero source term, equation [1.1] becomes

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} = 0 \quad [1.5]$$

In such a case (see section 1.1.2), U is neither created nor destroyed over the domain. The total amount of U over the domain varies only due to the difference between the incoming and outgoing fluxes at the boundaries of the domain.

Depending on the expression of the flux function, the conservation law is said to be convex, concave or non-convex (Figure 1.1):

– the law is convex when the second-order derivative $\partial^2 F / \partial U^2$ of the flux function with respect to U is positive for all U ;

– the law is concave when the second-order derivative $\partial^2 F / \partial U^2$ of the flux function with respect to U is negative for all U ;

– the law is said to be non-convex when the sign of the second-order derivative $\partial^2 F / \partial U^2$ of the flux function with respect to U changes with U .

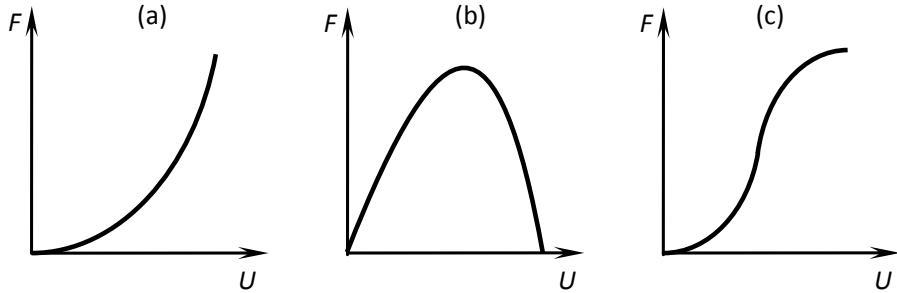


Figure 1.1. Typical examples of flux functions: convex (a), concave (b), non-convex (c)

1.1.2. Derivation from general conservation principles

The conservation form [1.1] is derived from a balance over a control volume of unit section defined between x_0 and $x_0 + \delta x$ (Figure 1.2). The balance is carried out over the control volume between two times t_0 and $t_0 + \delta t$. The variation in the total amount of U contained in the control volume is then related to the derivatives $\partial U / \partial t$ and $\partial F / \partial x$ in the limit of vanishing δt and δx .

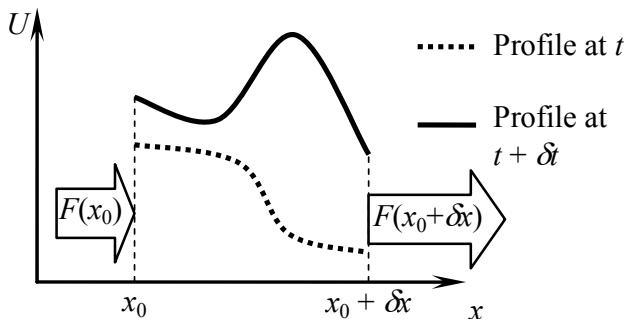


Figure 1.2. Definition sketch for the balance over a control volume

4 Wave Propagation in Fluids

The total amount $M(t_0)$ of U contained in the control volume at $t = t_0$ is defined as:

$$M(t_0) = \int_{x_0}^{x_0 + \delta x} U(x, t_0) dx \quad [1.6]$$

At $t = t_0 + \delta t$, the total amount of U contained in the control volume is:

$$M(t_0 + \delta t) = \int_{x_0}^{x_0 + \delta x} U(x, t_0 + \delta t) dx \quad [1.7]$$

The variation δS in the amount of U induced by the source term S over the domain between t_0 and $t_0 + \delta t$ is given by:

$$\delta S = \int_{t_0}^{t_0 + \delta t} \int_{x_0}^{x_0 + \delta x} S(U, x, t) dx dt \quad [1.8]$$

The amount $\delta F(x_0)$ of U brought by the flux F across the left-hand side boundary of the control volume between t_0 and $t_0 + \delta t$ is given by:

$$\delta F(x_0) = \int_t^{t + \delta t} F(x_0, t) dt \quad [1.9]$$

A quantity $\delta F(x_0 + \delta x)$ leaves the domain across the right-hand side boundary:

$$\delta F(x_0 + \delta x) = \int_t^{t + \delta t} F(x_0 + \delta x, t) dt \quad [1.10]$$

Stating the conservation of U over the control volume $[x_0, x_0 + \delta x]$ between t_0 and $t_0 + \delta t$, the following equality is obtained:

$$M(t_0 + \delta t) = M(t_0) + \delta F(x_0) - \delta F(x_0 + \delta x) + \delta S \quad [1.11]$$

Substituting equations [1.6] – [1.10] into equation [1.11] leads to:

$$\begin{aligned} \int_{x_0}^{x_0 + \delta x} [U(x, t_0 + \delta t) - U(x, t_0)] dx &= \int_{t_0}^{t_0 + \delta t} [F(x_0, t) - F(x_0 + \delta x, t)] dt \\ &+ \int_{t_0}^{t_0 + \delta t} \int_{x_0}^{x_0 + \delta x} S(x, t) dx dt \end{aligned} \quad [1.12]$$

A first-order Taylor series expansion around (x_0, t_0) gives:

$$\left. \begin{aligned} U(x_0, t_0 + \delta t) - U(x_0, t_0) &= \delta t \frac{\partial U}{\partial t} + O(\delta t^2) \\ F(x_0, t_0) - F(x_0 + \delta x, t_0) &= -\delta x \frac{\partial F}{\partial x} + O(\delta x^2) \end{aligned} \right\} [1.13]$$

where the quantities $O(\delta t^2)$ and $O(\delta x^2)$ are second- or higher-order polynomials with respect to δt and δx respectively. These polynomials contain the second- and higher-order derivatives of U and F with respect to t and x . When δt and δx tend to zero, the polynomial $O(\delta t^2)$ becomes negligible compared to the quantity $\delta t \partial U / \partial t$ because δt^2 decreases faster than δt . The polynomial $O(\delta x^2)$ becomes negligible compared to $\delta x \partial F / \partial x$ for the same reason. Relationships [1.13] thus become:

$$\left. \begin{aligned} U(x_0, t_0 + \delta t) - U(x_0, t_0) &\approx_{\delta t \rightarrow 0} \delta t \frac{\partial U}{\partial t} \\ F(x_0, t_0) - F(x_0 + \delta x, t_0) &\approx_{\delta x \rightarrow 0} -\delta x \frac{\partial F}{\partial x} \end{aligned} \right\} [1.14]$$

A similar reasoning leads to the following equivalence:

$$\int_{t_0}^{t_0 + \delta t} \int_{x_0}^{x_0 + \delta x} S(x, t) dx dt \underset{\delta t \rightarrow 0}{\underset{\delta x \rightarrow 0}{\approx}} \delta t \delta x S [1.15]$$

Substituting equations [1.14] and [1.15] into equation [1.12] leads to

$$\delta t \frac{\partial U}{\partial t} \delta x = -\delta x \frac{\partial F}{\partial x} \delta t + \delta t \delta x S [1.16]$$

Dividing equation [1.16] by $\delta t \delta x$ yields the conservation form [1.1], recalled here:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} = S$$

6 Wave Propagation in Fluids

The following remarks can be made:

- the Partial Differential Equation (PDE) [1.1] is a particular case of the more general, integral equation [1.12]. Equation [1.1] is obtained from equation [1.12] using the assumption that ∂t and ∂x tend to zero. Equation [1.12] is the so-called weak form of equation [1.1] (see Chapter 3 for more details);
- the conservation form [1.1] is based on the implicit assumption that F is differentiable with respect to x and U is differentiable with respect to t . Consequently, [1.1] is meaningful only when U is continuous in space and time. In contrast, equation [1.12] is meaningful even when U is discontinuous in space and/or time. This has consequences on the calculation of discontinuous solutions, as shown in Chapter 3.

1.1.3. Non-conservation form

Equation [1.1] can be rewritten in the so-called non-conservation form that involves only derivatives of U . The non-conservation form of equation [1.1] is:

$$\frac{\partial U}{\partial t} + \lambda \frac{\partial U}{\partial x} = S' \quad [1.17]$$

where λ is called the wave speed, and S' is a source term that may be identical (but not necessarily) to the source term S in equation [1.1]. Equation [1.17] is obtained from equation [1.1] by rewriting the derivative $\partial F / \partial x$ as:

$$\frac{\partial F}{\partial x} = \frac{\partial F}{\partial U} \frac{\partial U}{\partial x} + F' \quad [1.18]$$

where the term $F' = (\partial F / \partial x)_{U=\text{Const}}$ contains all the derivatives of F other than the derivative with respect to U . The expression of F being known, $\partial F / \partial U$ and F' are easily determined. Substituting equation [1.18] into equation [1.1] yields:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial U} \frac{\partial U}{\partial x} + F' = S \quad [1.19]$$

that is:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial U} \frac{\partial U}{\partial x} = S - F' \quad [1.20]$$

Comparing equation [1.20] to equation [1.17] leads to the following definitions for λ and S' :

$$\left. \begin{array}{l} \lambda = \frac{\partial F}{\partial U} \\ S' = S - \left(\frac{\partial F}{\partial x} \right)_{U=\text{Const}} \end{array} \right\} \quad [1.21]$$

The expressions of F and S being known, the knowledge of U at any point in time and space allows λ and S' to be calculated directly. From definition [1.21], in the case where the variations in F are due to variations in U only, $F' = 0$ and S' is identical to S .

Example: assume that the flux function F is defined as in equation [1.3], recalled here:

$$F = aU$$

where a is a function of x and t . Equation [1.18] then becomes:

$$\frac{\partial F}{\partial x} = \frac{\partial}{\partial x}[a(x, t)U] = a \frac{\partial U}{\partial x} + U \frac{\partial a}{\partial x} \quad [1.22]$$

and λ and F' are given by:

$$\left. \begin{array}{l} \lambda = a \\ F' = U \frac{\partial a}{\partial x} \end{array} \right\} \quad [1.23]$$

If a does not depend on x , $F' = 0$ because $\partial a / \partial x = 0$.

1.1.4. Characteristic form – Riemann invariants

Writing a conservation law in non-conservation form leads to the notions of characteristic form and the Riemann invariant. Such notions are essential to the understanding of hyperbolic conservation laws. A very convenient way of determining the behavior of the solutions of hyperbolic conservation laws consists of identifying invariant quantities (that is, quantities that do not change) along certain trajectories, also called the “characteristic curves” (or more simply the “characteristics”). The solution is calculated by “following” the invariants along the

8 Wave Propagation in Fluids

characteristics, which allows the value of U to be determined at any point. To do so, the non-conservation form [1.17] is used:

$$\frac{\partial U}{\partial t} + \lambda \frac{\partial U}{\partial x} = S'$$

The purpose is to derive the expression of the variation δU in U observed by an observer travelling at a given speed v . A small time interval δt is considered, over which the traveler moves by a distance $\delta x = v \delta t$. The variation δU “seen” by the observer is given by:

$$\delta U = \frac{\partial U}{\partial t} \delta t + \frac{\partial U}{\partial x} \delta x = \left(\frac{\partial U}{\partial t} + v \frac{\partial U}{\partial x} \right) \delta t \quad [1.24]$$

Note that from the observer's point of view, U is a function of time only, because the observer's location $x(t)$ is defined by $dx/dt = v$. When δt tends to zero, the ratio $\delta U / \delta t$ tends to the so-called total derivative dU/dt . Therefore equation [1.24] becomes:

$$\frac{\delta U}{\delta t} \underset{\delta t \rightarrow 0}{\approx} \frac{dU}{dt} = \frac{\partial U}{\partial t} + v \frac{\partial U}{\partial x} \quad \text{for } \frac{dx}{dt} = v \quad [1.25]$$

In the particular case of an observer moving at a speed λ , equation [1.25] becomes:

$$\frac{dU}{dt} = \frac{\partial U}{\partial t} + \lambda \frac{\partial U}{\partial x} \quad \text{for } \frac{dx}{dt} = \lambda \quad [1.26]$$

Comparing equations [1.26] and [1.17] leads to:

$$\frac{dU}{dt} = S' \quad \text{for } \frac{dx}{dt} = \lambda \quad [1.27]$$

Equation [1.27] is the so-called characteristic form of equation [1.1]. The trajectory, the equation of which is $dx/dt = \lambda$, is called a characteristic. λ is called the wave speed.

S' being a function of U , x and t , its value may be calculated at any point (x, t) if the value of U is known. The first-order Ordinary Differential Equation (ODE) [1.27] is applicable along the characteristic.

In the (important) particular case where the source term S' is zero, equation [1.17] becomes:

$$\frac{\partial U}{\partial t} + \lambda \frac{\partial U}{\partial x} = 0 \quad [1.28]$$

and equation [1.27] becomes:

$$\frac{dU}{dt} = 0 \text{ for } \frac{dx}{dt} = \lambda \quad [1.29]$$

Equation [1.29] can also be written as:

$$U = \text{Const} \text{ for } \frac{dx}{dt} = \lambda \quad [1.30]$$

Consequently, the quantity U is invariant to an observer moving at the speed λ . U is called a Riemann invariant.

The physical meaning of the wave speed is the following. The wave speed is the speed at which the variations in U (and not U itself) propagate. A perturbation appearing in the profile of U at a given time propagates at the speed λ . The wave speed can be viewed as the speed at which “information”, or “signals” created by variations in U , propagate in space.

1.2. Determination of the solution

1.2.1. Representation in the phase space

The phase space is a very useful tool in the determination of the behavior of the solutions of hyperbolic conservation laws. The term “phase space” indicates the (x, t) plane formed by the space coordinate x and the time coordinate t (Figure 1.3).

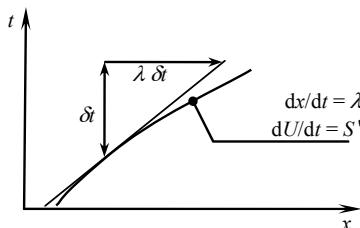


Figure 1.3. Representation of characteristic curves in the phase space

The trajectory $\frac{dx}{dt} = \lambda$ is represented by a curve in the phase space. The distance δx covered by the characteristic over a time interval δt is given by $\delta x = \lambda \delta t$, therefore the slope of the line is $\delta t / \delta x = 1/\lambda$. Note that the sign of λ may change with time depending on the variations in U and the expressions of λ and S' . When λ becomes zero the tangent to the characteristic curve is vertical in the phase space (Figure 1.4a). In contrast, an extremum with respect to time is not physically permissible (Figure 1.4b) because “travelling backwards in time” is not possible.

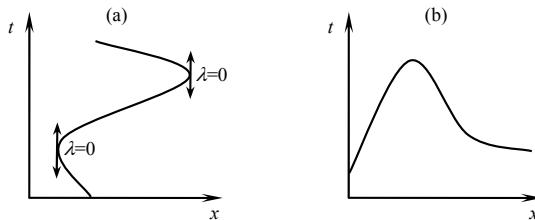


Figure 1.4. Physically permissible (a) and non-permissible (b) characteristics

The representation in the phase space may be used to determine the behavior of the solutions of conservation law [1.1] using the so-called “method of characteristics”. The following simple case is considered:

- the source term S in equation [1.1] is zero;
- the flux depends only on U , therefore $F' = 0$ in equations [1.18] – [1.20].

The characteristic form [1.27] then reduces to equation [1.30], recalled here:

$$U = \text{Const} \quad \text{for} \quad \frac{dx}{dt} = \lambda$$

F being a function of U only, λ is also a function of U only. Consequently, if U is constant along a characteristic line, λ is also constant and the characteristic is a straight line in the phase space (Figure 1.5). Assume that the profile $U(x, t_0)$ is known for all x at the time t_0 . The purpose is to determine the profile $U(x, t_1)$ for all x at the time $t_1 > t_0$. Consider the point A, the abscissa of which is denoted by x_A , at which the value of U at (x_A, t_0) is denoted by U_A . Since the wave speed λ depends on U only, the characteristic passing at A is a straight line. Its (constant) wave speed is $\lambda_A = \partial F / \partial U (U_A)$. At time t_1 , the characteristic has moved to point A', the abscissa $x_{A'}$ of which is given by:

$$x_{A'} = x_A + (t_1 - t_0)\lambda_A \quad [1.31]$$

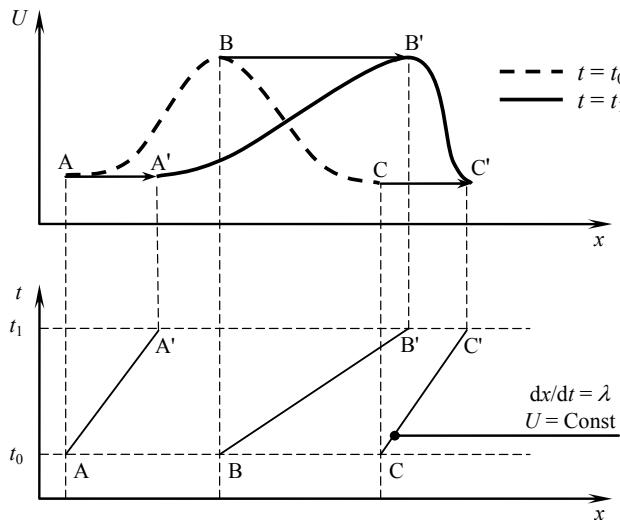


Figure 1.5. Representation of the characteristics in the phase space (bottom) and behavior of the physical profile (top) in the particular case $F' = S = 0$

From the property of invariance of U along the characteristic, U remains unchanged between A and A':

$$U_{A'} = U_A \quad [1.32]$$

Extending the reasoning above to any value of x , the following relationship is obtained:

$$U(x + \lambda\Delta t, t + \Delta t) = U(x, t) \quad [1.33]$$

where Δt represents the quantity $(t_1 - t_0)$ and λ is estimated at (x, t) .

Figure 1.5 shows how the method of characteristics can be used to determine the evolution of a given profile [ABC]. The figure is drawn assuming that λ is an increasing function of U . Therefore, point B moves faster than points A and C because U_B is larger than U_A and U_C . Consequently, the region [AB] tends to spread in time, while the region [BC] becomes narrower. After a certain amount of time point B catches up with point C and the solution becomes discontinuous at point $B' = C'$. The derivatives $\partial U / \partial t$ and $\partial U / \partial x$ are no longer defined and a specific treatment must be applied to determine the solution at later times. Such a treatment is detailed in Chapter 3.

In the general case, S and F' are non-zero. Then relationship [1.33] cannot be used because:

- U is not invariant along a characteristic line;
- the characteristics are therefore curved lines, the slope of which depends on the local value of x and U .

Therefore, no simple relationship can be derived between the initial profile at $t = t_0$ and the final profile at $t = t_1$. In most cases, the solution must be computed approximately using numerical methods. Such methods are dealt with in Chapters 6 and 7.

1.2.2. Initial conditions, boundary conditions

In practical applications, the solution of equation [1.1] is sought over a domain of finite length. A key issue is the amount of information needed for the calculation of U at a point $M(x, t)$ in the domain. This question is best answered using the phase space (Figure 1.6). The solution domain is assumed to extend from $x = 0$ to $x = L$.

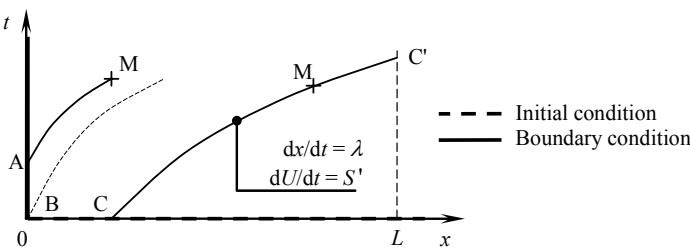


Figure 1.6. Initial and boundary conditions in the phase space

For the sake of clarity, the wave speed λ is assumed to be positive over the entire domain (the case where the sign of the wave speed changes is examined at the end of the section). Two possibilities arise:

- If point M is located on the right-hand side of the characteristic that passes at point B ($x = 0, t = 0$), there exists a point C on the line ($t = 0$) such that the characteristic passing at C passes at M . Point C is called the foot of the characteristic at $t = 0$. If the value of U is known at point C , U can be computed along the characteristic line by solving the characteristic form [1.27] using any analytical or numerical method. Therefore, the value of U can be computed at any point M located on the right-hand side of the characteristic that passes at B ($0, 0$), provided

that $U(x, 0)$ is known for all x between 0 and L . The function that describes the profile $U(x, 0)$ is called the initial condition. It is expressed as follows:

$$U_0(x) = U(x, t = 0), \quad x \in [0, L] \quad [1.34]$$

– If point M is located on the left-hand side of the characteristic passing at B, the value of U at M cannot be calculated from the initial condition and the knowledge of the value of U at all points A along the line ($x = 0$) is necessary. The function that describes the profile $U(0, t)$ is called a boundary condition. In the case of a positive λ , the characteristics enter the domain on the left-hand side and the left boundary condition must be used. It is expressed as follows:

$$U_b(t) = U(x = 0, t), \quad t > 0 \quad [1.35]$$

Note that a boundary condition can be prescribed only if the characteristics enter the domain. In the situation illustrated by Figure 1.6, prescribing a boundary condition at the point C' would be meaningless because the value of U at C' is entirely determined by the initial condition at C via the characteristic form [1.27] and cannot be prescribed independently of it. Depending on the variations of λ with U, x and t , the number of boundary conditions needed to determine U uniquely over the domain $[0, L]$ may be 0, 1 or 2 (see Figure 1.7).

In configuration (a) the characteristics leave the domain at both boundaries ($x = 0$) and ($x = L$). The value of U at both boundaries is determined entirely by the initial condition $U(x, 0)$. In configuration (b) the left-hand boundary condition is needed because the characteristics enter the domain at $x = 0$ and the value of U at this location cannot be determined from the values inside the domain. In contrast, the knowledge of U at the right-hand boundary is not required because U is determined uniquely from the value of U inside the domain. In configuration (c) the characteristics enter the domain at both $x = 0$ and $x = L$. Consequently, two boundary conditions are needed, one at each end of the domain, because $U(0, t)$ and $U(L, t)$ cannot be determined from inside the domain and must therefore be specified independently in the form of boundary conditions.

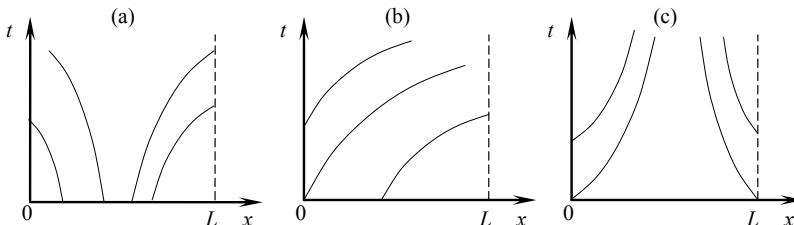


Figure 1.7. Number of boundary conditions needed depending on the variations of the wave speed: none (a), one (b), two (c)

1.3. A linear law: the advection equation

1.3.1. Physical context – conservation form

The linear advection equation is the simplest possible hyperbolic conservation law. It is found in many domains of fluid mechanics because it expresses a widespread phenomenon, the transport of a given quantity in a moving fluid. The transported variable may be the temperature of the fluid, the concentration in a given chemical, etc. The expression “advection” is often understood as the advection of a passive scalar, that is, a quantity that does not influence the behavior of the flow by which it is transported. In a number of cases however, the transported quantity influences the velocity field, a phenomenon known as coupling. This is the case of the inviscid Burgers equation dealt with in section 1.4.

In this section, a passive scalar is considered. The example of a chemical substance dissolved in water with a concentration variable in space and time is used. The water is assumed to flow in a channel, the transverse dimensions of which are assumed to be negligible compared to the longitudinal dimension. The channel may be an open channel (a river, a canal) or a closed channel (a conduit) with a cross-sectional area variable in space and time. The assumption of negligible transverse dimensions for the channel allows the assumption of a one-dimensional, longitudinal flow and transport process to be used. The channel is represented as a one-dimensional object. The space coordinate is the curvilinear abscissa (Figure 1.8).

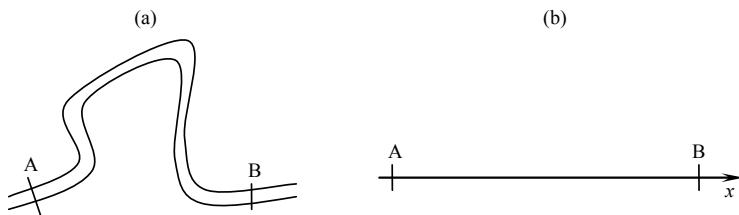


Figure 1.8. One-dimensional representation of a channel, the transverse dimension of which can be considered negligible. Reality (a) and model (b)

The governing PDE for the one-dimensional transport of a dissolved substance is derived by carrying out a balance as in section 1.1.2. The total quantity $M(t)$ of substance (the “mass” as introduced in section 1.1.2) over an elementary channel slice of length δx (Figure 1.9) is given by

$$M = C \delta V \quad [1.36]$$

where C is the concentration of the dissolved substance and δV is the volume of the elementary channel slice, given by:

$$\delta V = A \delta x \quad [1.37]$$

where A is the cross-sectional area of the channel (Figure 1.9).

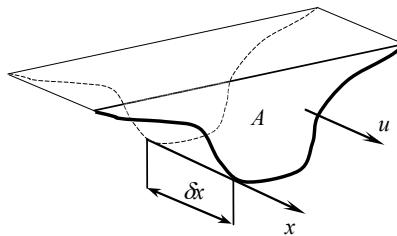


Figure 1.9. Perspective view of an elementary channel section

The amount $\delta F(x_0)$ of dissolved chemical that passes at x_0 during an elementary time interval δt is given by:

$$\delta F(x_0) = A u C \delta t \quad [1.38]$$

where u is the flow velocity. Using the same reasoning as in equations [1.11] to [1.16] with a zero source term, the PDE that describes the conservation of mass (also called the continuity equation) is obtained:

$$\frac{\partial}{\partial t}(AC) + \frac{\partial}{\partial x}(AuC) = 0 \quad [1.39]$$

Equation [1.39] can be written in the form [1.1] by defining the conserved variable U , the flux F and the source term S as:

$$\left. \begin{aligned} U &= AC \\ F &= QC \\ S &= 0 \end{aligned} \right\} \quad [1.40]$$

where $Q = Au$ is the so-called liquid discharge.

1.3.2. Characteristic form

Several approaches may be used to rewrite equation [1.39] in characteristic form. A first approach consists of defining the conserved quantity as AC and rewriting equations [1.39] to [1.40] as:

$$\frac{\partial U}{\partial t} + \frac{\partial}{\partial x}(uU) = 0 \quad [1.41]$$

noting that $\partial / \partial x(uU) = u \partial U / \partial x + U \partial u / \partial x$, equation [1.41] becomes:

$$\frac{\partial U}{\partial t} + u \frac{\partial U}{\partial x} = -U \frac{\partial u}{\partial x} \quad [1.42]$$

As shown in section 1.1.4 (see equations [1.24] to [1.27]), equation [1.42] is equivalent to:

$$\frac{dU}{dt} = -U \frac{\partial u}{\partial x} \quad \text{for } \frac{dx}{dt} = u \quad [1.43]$$

Equation [1.43] is of limited interest because U does not appear as an invariant quantity along a characteristic line.

In a second approach, equation [1.39] is rewritten with respect to the concentration C by developing the derivatives:

$$A \frac{\partial C}{\partial t} + C \frac{\partial A}{\partial t} + Q \frac{\partial C}{\partial x} + C \frac{\partial Q}{\partial x} = 0 \quad [1.44]$$

Equation [1.44] is rewritten as:

$$A \frac{\partial C}{\partial t} + Q \frac{\partial C}{\partial x} = -\left(\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x}\right)C \quad [1.45]$$

Noting that the continuity equation for the flow can be written as:

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0 \quad [1.46]$$

substituting equation [1.46] into equation [1.45] yields the following equation:

$$A \frac{\partial C}{\partial t} + Q \frac{\partial C}{\partial x} = 0 \quad [1.47]$$

Dividing by A and noting that $Q/A = u$ leads to:

$$\frac{\partial C}{\partial t} + u \frac{\partial C}{\partial x} = 0 \quad [1.48]$$

From the developments carried out in section 1.1.4 (see equations [1.24] to [1.27]), equation [1.48] is known to be equivalent to the following characteristic form:

$$\frac{dC}{dt} = 0 \quad \text{for} \quad \frac{dx}{dt} = u \quad [1.49]$$

Equation [1.49] is equivalent to

$$C = \text{Const} \quad \text{for} \quad \frac{dx}{dt} = u \quad [1.50]$$

Equation [1.50] is an interesting alternative to equation [1.43] because it allows a Riemann invariant to be derived. The Riemann invariant is the concentration of the dissolved substance. Note that the conserved quantity (the mass AC per unit length of channel) is not identical to the invariant quantity (the concentration).

1.3.3. Example: movement of a contaminant in a river

The difference between a conserved quantity and an invariant quantity is best illustrated by the following example. Consider a river in which a contaminant is transported without degradation or external inflow. The cross-sectional area of the river is variable, with a sudden narrowing at point N. The cross-sectional areas upstream and downstream of the narrowing are denoted by A_1 and A_2 respectively. The discharge Q is assumed to be constant in both space and time. The numerical parameters are summarized in Table 1.1.

Symbol	Meaning	Value
A_1	Cross-sectional area of the channel upstream of the narrowing	100 m ²
A_2	Cross-sectional area of the channel downstream of the narrowing	25 m ²
C_1	Concentration upstream of the narrowing	1 g/l
C_{us}	Upstream boundary condition ($x = 0$)	0 g/l
L_1	Length of the contaminant cloud upstream of the narrowing	100 m
Q	Liquid discharge	10 m ³ /s

Table 1.1. Parameters of the problem

The initial contaminant concentration is assumed to be zero everywhere, except over a segment [AB] where it is uniformly equal to C_1 . The segment [AB] is assumed to be located upstream of N at $t = 0$ (Figure 1.10). The upstream boundary condition is a zero concentration, so that no additional contaminant enters the domain from the upstream boundary.

As shown by the data in Table 1.1 the flow velocity is 0.1 m/s upstream of the narrowing and 0.25 m/s downstream of it. Denoting by $t_{A'}$ and $t_{B'}$ the times at which the characteristics passing at A and B reach the narrowing N, the following relationship holds:

$$\frac{x_B - x_A}{t_{A'} - t_{B'}} = u_1 = \frac{Q}{A_1} \quad [1.51]$$

Equation [1.51] can be rewritten as:

$$t_{A'} - t_{B'} = \frac{(x_B - x_A)A_1}{Q} = \frac{A_1 L_1}{Q} \quad [1.52]$$

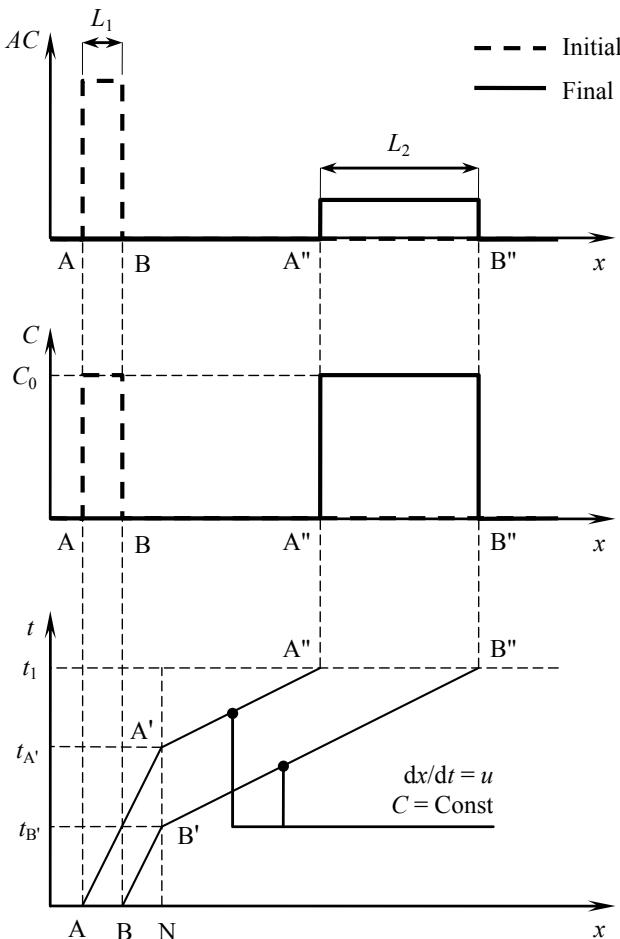


Figure 1.10. Behavior of the solution in the physical space (top, middle) and the phase space (bottom)

Consider now a time t_1 after the contaminant cloud has left point R. The front and rear of the cloud are denoted by A'' and B'' respectively. The abscissas $x_{A''}$ and $x_{B''}$ are related to the times $t_{A'}$ and $t_{B'}$ as follows:

$$\frac{x_{B''} - x_{A''}}{t_{A'} - t_{B'}} = u_2 = \frac{Q}{A_2} \quad [1.53]$$

Denoting with L_2 the width of the cloud downstream of the narrowing, the following relationship holds:

$$L_2 = \frac{Q}{A_2} (t_{A'} - t_{B'}) = \frac{A_1}{A_2} L_1 \quad [1.54]$$

From the parameters in Table 1.1 the cloud is 4 times larger downstream of the narrowing than upstream of it. Note that:

- the concentration C is invariant along a characteristic line. Therefore the concentration profile remains piecewise constant. The integral of the concentration with respect to x at $t = 0$ is given by:

$$\int_0^{+\infty} C(x, 0) dx = L_1 C_0 \quad [1.55]$$

The integral at $t = t_1$ is given by:

$$\int_0^{+\infty} C(x, t_1) dx = L_2 C_0 = \frac{A_1}{A_2} L_1 C_0 \quad [1.56]$$

where equation [1.54] is used to relate L_2 and L_1 . Although the source term is zero and no contaminant enters the domain across the boundary, the integral of C with respect to x is not constant. The concentration is an invariant quantity, it is not a conserved quantity;

- the mass of contaminant per unit length is equal to AC . At $t = 0$ the integral of AC with respect to x is:

$$\int_0^{+\infty} A(x) C(x, 0) dx = A_1 L_1 C_0 \quad [1.57]$$

At $t = t_1$ the integral is equal to:

$$\int_0^{+\infty} A(x) C(x, t_1) dx = A_2 L_2 C_0 = A_2 \frac{A_1}{A_2} L_1 C_0 = A_1 L_1 C_0 \quad [1.58]$$

where equation [1.54] is used again to provide a relationship between L_1 and L_2 . In contrast with the integral of C , the integral of AC (that is, the total mass of solute contained in the river reach) is conserved. However, AC is equal to $A_1 C_0$ upstream of N and is equal to $A_2 C_0$ downstream of N. Consequently, AC is a conserved quantity, not an invariant quantity.

1.3.4. Summary

The conservation form of the linear advection equation for a contaminant in a channel is equation [1.39]. It can be written in the form [1.1] by defining the conserved variable, the flux and the source terms as in equation [1.40].

The non-conservation form of equation [1.39] is equation [1.48] and its characteristic form is equation [1.49].

The conserved variable is the mass of contaminant AC per unit length of channel. The invariant quantity is the contaminant concentration C .

1.4. A convex law: the inviscid Burgers equation

1.4.1. Physical context – conservation form

The Burgers equation was first introduced [BUR 48] as a simple model to account for nonlinear advection and diffusion. A transformation was proposed independently by Hopf [HOP 50] and Cole [COL 51], whereby the Burgers equation is transformed into a linear PDE. In what follows, only the inviscid form of the equation is considered. If the initial and boundary conditions are not too complex, an analytical solution can be derived for the inviscid Burgers equation. This equation is often used to assess the performance of numerical methods for hyperbolic PDEs.

The inviscid Burgers equation can be viewed as a restriction of the Euler equations of gas dynamics introduced in Chapter 2. It is derived from the momentum equation under the assumption of negligible external forces, pressure gradients and momentum diffusion. If this is the case, the variations in the fluid velocity are due only to the initial distribution of momentum in the fluid. In contrast with the advection equation dealt with in section 1.3, the inviscid Burgers equation describes the advection of an active scalar, in that the value of u influences its own propagation speed.

Consider a slice of fluid, the (infinitesimal) length of which is denoted by δx and the cross-sectional area of which is equal to unity (Figure 1.11). The mass δm and the momentum δq of the slice of fluid are given by:

$$\left. \begin{aligned} \delta m &= \rho \delta x \\ \delta q &= \rho u \delta x = u \delta m \end{aligned} \right\} \quad [1.59]$$

where u and ρ are the fluid velocity and density respectively.

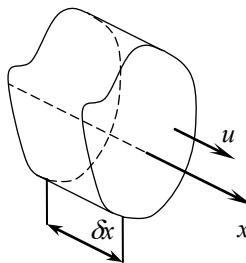


Figure 1.11. Perspective view of an elementary volume of width Δx and unit cross-sectional area

Both the density and the momentum are transported at the speed u of the fluid. The mass that crosses the unit section over an elementary time interval Δt is given by:

$$\delta F_m = \rho u \Delta t = q \Delta t \quad [1.60]$$

The momentum that crosses the unit section over the time interval Δt is given by:

$$\delta F_q = qu \Delta t = \rho u^2 \Delta t \quad [1.61]$$

Applying the reasoning [1.11–16] to the conservation of the density ρ leads to the so-called continuity equation:

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x} (\rho u) = 0 \quad [1.62]$$

while the principle of conservation of the momentum $q = \rho u$ leads to the so-called momentum equation:

$$\frac{\partial}{\partial t} (\rho u) + \frac{\partial}{\partial x} (\rho u^2) = 0 \quad [1.63]$$

1.4.2. Characteristic form

The characteristic form of the inviscid Burgers equation is obtained by developing equation [1.63] into:

$$\rho \frac{\partial u}{\partial t} + u \frac{\partial \rho}{\partial t} + u \frac{\partial}{\partial x}(\rho u) + \rho u \frac{\partial u}{\partial x} = 0 \quad [1.64]$$

Substituting equation [1.62] into equation [1.64] yields the following equation:

$$\rho \frac{\partial u}{\partial t} + \rho u \frac{\partial u}{\partial x} = 0 \quad [1.65]$$

Dividing by the density leads to an equation in u in non-conservation form:

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = 0 \quad [1.66]$$

As shown in section 1.1.4 (see equations [1.24] to [1.27]), equation [1.66] is equivalent to:

$$\frac{du}{dt} = 0 \quad \text{for } \frac{dx}{dt} = u \quad [1.67]$$

that is:

$$u = \text{Const} \quad \text{for } \frac{dx}{dt} = u \quad [1.68]$$

The conserved quantities are the density ρ and the momentum $q = \rho u$, the invariant quantity is u . Note however that u may be seen as a conserved quantity in equation [1.66], which can be rewritten in conservation form by noting that $u \, du = d(u^2/2)$:

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} \left(\frac{u^2}{2} \right) = 0 \quad [1.69]$$

Equation [1.69] is the so-called conservation form of the inviscid Burgers equation. It can be written in the form [1.1] by defining U and F as follows:

$$\left. \begin{aligned} U &= u \\ F &= \frac{U^2}{2} \end{aligned} \right\} \quad [1.70]$$

Equation [1.69] is a convex conservation law because the flux function $F(U)$ is a quadratic function with a positive second-order derivative. The quantity U is conserved and invariant. Moreover, the wave speed is equal to $dF/dU = U = u$. Since u is an invariant along the characteristic curves, these are straight lines in the phase space.

1.4.3. Example: propagation of a perturbation in a fluid

The current example deals with the variations in the fluid density and velocity arising from an initial perturbation in the velocity profile. In what follows, the time t is assumed to be small enough for the density and the velocity profiles to remain continuous, that is, the derivatives of ρ and u with respect to time and space are assumed to take finite values. If this is not the case, the analysis hereafter becomes invalid and a specific treatment must be applied to the discontinuities (also called shocks). Such a treatment is not detailed in this chapter. It is covered in detail in Chapter 3.

Consider the following initial condition. At $t = 0$, the density is uniformly equal to ρ_0 . The velocity profile is uniformly equal to u_0 , except over a segment [AC], where it is triangular (Figure 1.12). The initial velocity profile is symmetrical with a width $2L$ and the maximum value (also called peak velocity) is denoted by u_1 . The parameters of the problem are summarized in Table 1.2. The velocity being positive everywhere, the characteristics move from left to right. Therefore the upstream boundary is the left-hand boundary, $x = 0$.

Symbol	Meaning	Value
u_0	Initial flow velocity outside the perturbed region [AC]	1 m/s
u_1	Initial peak velocity	2 m/s
u_{us}	Flow velocity at the upstream boundary ($x = 0$)	1 m/s
L	Half-length of the initial velocity perturbation	1 m
ρ_0	Initial density	1 kg/m ³
ρ_{us}	Fluid density at the upstream boundary ($x = 0$)	1 kg/m ³

Table 1.2. Problem parameters

The characteristic form [1.67] allows the behavior of the flow velocity to be determined. Points A and C move at the speed u_0 , while point B moves at the speed u_1 . Therefore point B moves faster than A and C. The profile becomes smoother between A and B and steeper between B and C.

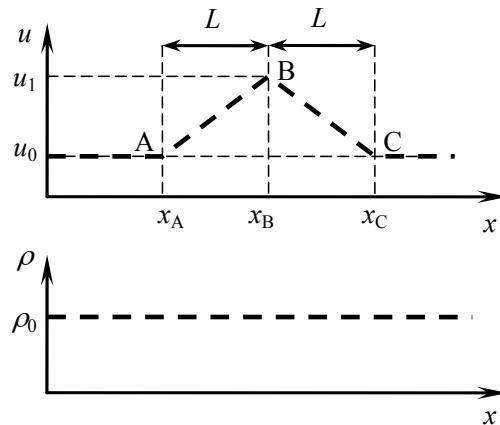


Figure 1.12. Initial profiles for the velocity (top) and the density (bottom)

The profile can easily be shown to remain triangular. Indeed, differentiating equation [1.66] with respect to x leads to the following equation:

$$\frac{\partial^2 u}{\partial x \partial t} + \frac{\partial}{\partial x} \left(u \frac{\partial u}{\partial x} \right) = 0 \quad [1.71]$$

Swapping the derivatives in the first term of equation [1.71] and developing the second term yields:

$$\frac{\partial^2 u}{\partial t \partial x} + \left(\frac{\partial u}{\partial x} \right)^2 + u \frac{\partial^2 u}{\partial x^2} = 0 \quad [1.72]$$

which gives the following PDE in $\partial u / \partial x$:

$$\frac{\partial u}{\partial t} \left(\frac{\partial u}{\partial x} \right) + u \frac{\partial}{\partial x} \left(\frac{\partial u}{\partial x} \right) = - \left(\frac{\partial u}{\partial x} \right)^2 \quad [1.73]$$

The reasoning developed in section 1.1.4 allows equation [1.73] to be rewritten in the form:

$$\frac{d}{dt} \left(\frac{\partial u}{\partial x} \right) = - \left(\frac{\partial u}{\partial x} \right)^2 \quad \text{for } \frac{dx}{dt} = u \quad [1.74]$$

The solution of equation [1.74] is:

$$\frac{\partial u}{\partial x}(t) = \frac{u'_0}{1 + (t - t_0) u'_0} \quad \text{for } \frac{dx}{dt} = u \quad [1.75]$$

where u'_0 is the value of $\partial u / \partial x$ at $t = t_0$ along the characteristic line (Figure 1.13).

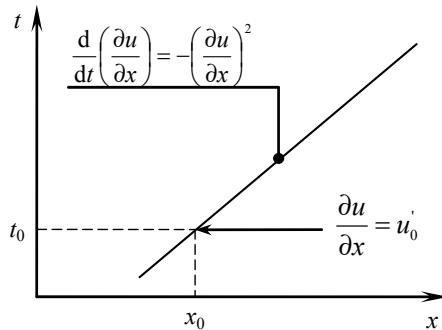


Figure 1.13. Determining the variations in $\partial u / \partial x$ along a characteristic from the initial velocity profile

The function $u'_0(x)$ is given by:

$$u'_0(x) = u(x, t_0) = \begin{cases} 0 & \text{for } x < x_A \\ \frac{u_1 - u_0}{L} & \text{for } x_A \leq x < x_B \\ \frac{u_0 - u_1}{L} & \text{for } x_B \leq x < x_C \\ 0 & \text{for } x \geq x_C \end{cases} \quad [1.76]$$

This function is piecewise constant and remains such at later times. Applying equation [1.75] along a characteristic line and using equation [1.76] for the initial condition leads to:

$$\frac{\partial u}{\partial x}(x, t) = \begin{cases} 0 & \text{for } x < x_{A'} \\ \frac{u_1 - u_0}{L + (u_1 - u_0)t} & \text{for } x_{A'} \leq x < x_{B'} \\ \frac{u_0 - u_1}{L + (u_0 - u_1)t} & \text{for } x_{B'} \leq x < x_{C'} \\ 0 & \text{for } x \geq x_{C'} \end{cases} \quad [1.77]$$

where the abscissas $x_{A'}$, $x_{B'}$ and $x_{C'}$ are the abscissas at the time t of the characteristics passing at A, B and C at the time $t_0 = 0$ (Figure 1.14):

$$\left. \begin{array}{l} x_{A'} = x_A + u_0 t \\ x_{B'} = x_B + u_1 t \\ x_{C'} = x_C + u_0 t \end{array} \right\} \quad [1.78]$$

Equation [1.77] is that of a piecewise constant derivative. Consequently, the profile of u is piecewise linear. Note that, since $u_1 > u_0$, the quantity $L + (u_1 - u_0)t$ in the second equation [1.77] increases with time and the magnitude of $\partial u / \partial x$ decreases with time between A' and B'. Conversely, $L + (u_0 - u_1)t$ in the third equation [1.77] decreases with time and the magnitude of $\partial u / \partial x$ increases with time between B' and C'. In other words, the profile [A'B'] becomes flatter, while the profile [B'C'] becomes steeper. From the third equation [1.77] the derivative becomes infinite for a time $t = t_{\max}$ given by:

$$L + (u_0 - u_1) t_{\max} = 0 \quad [1.79]$$

that is:

$$t_{\max} = \frac{L}{u_1 - u_0} \quad [1.80]$$

At $t = t_{\max}$ the segment [B'C'] becomes vertical. The velocity profile is no longer continuous, and both the conservation form [1.1] and the non-conservation form [1.17] become invalid. A special treatment is needed to allow the behavior of the solution to be analyzed at later times. Such a treatment is described in Chapter 3.

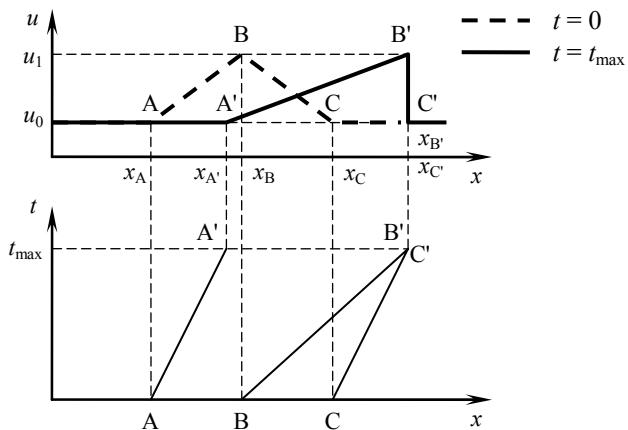


Figure 1.14. Velocity profile in the physical space (top) and in the phase space (bottom) for $t = 0$ and $t = t_{\max}$

1.4.4. Summary

The conservation form of the inviscid Burgers equation is given by equation [1.69]. It can be written as in equation [1.1] by defining the conserved variable and the flux as in equations [1.70].

The non-conservation and characteristic forms of the equation are given by equations [1.66] and [1.67] respectively.

Both the conserved variable and the Riemann invariant are equal to the fluid velocity u .

1.5. Another convex law: the kinematic wave for free-surface hydraulics

1.5.1. Physical context – conservation form

The kinematic wave is a simplified form of the open channel flow equations (Figure 1.15). It is a scalar equation, which makes it easy to solve. A summary of the conditions under which the kinematic wave provides a valid approximation to open channel flow can be found in [LIG 55]. The underlying assumptions of the kinematic wave equations are the following:

- Assumption (A1): the transverse dimensions of the channel are negligible compared to its longitudinal dimensions and the flow can be considered as one-dimensional.

- Assumption (A2): inertia is negligible and the channel slope is steep enough for the energy slope to be equivalent to the channel bed slope.

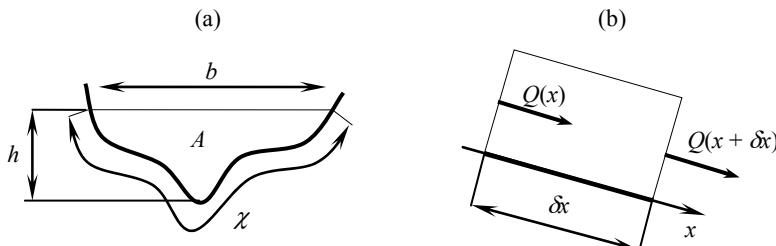


Figure 1.15. Flow in an open channel. Cross-sectional view (a), side view (b)

Under such conditions, the energy loss originating from friction against the walls is balanced with the energy gained from the slope. The friction term is assumed to obey Strickler's law:

$$Q = K_{\text{Str}} A R_H^{2/3} S_f^{1/2} \quad [1.81]$$

where A is the cross-sectional area of the channel, K_{Str} is Strickler's friction coefficient, R_H is the hydraulic radius and S_f is the slope of the energy line. The hydraulic radius is defined as the ratio of the cross-sectional area A to the wetted perimeter χ :

$$R_H = A / \chi \quad [1.82]$$

Since the slope of the energy line is assumed to be equal to the bottom slope S_0 , equation [1.81] becomes:

$$Q = K_{\text{Str}} \frac{A^{5/3}}{\chi^{2/3}} S_0^{1/2} \quad [1.83]$$

A mass balance over a control volume defined between abscissas x_0 and $x_0 + \delta x$ (Figure 1.15b) leads to the continuity equation (see section 1.2):

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0 \quad [1.84]$$

where Q is given by equation [1.83]. Note that Q is a known function of A at all points, because S_0 is known at all points and the law that relates the wetted perimeter χ to A is a known function of the geometry of the channel.

1.5.2. Non-conservation and characteristic forms

Equation [1.84] can be written in non-conservation form by developing the derivative of Q so as to involve derivatives of A , K_{Str} and S_0 :

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial A} \frac{\partial A}{\partial x} + \frac{\partial Q}{\partial K_{\text{Str}}} \frac{\partial K_{\text{Str}}}{\partial x} + \frac{\partial Q}{\partial S_0} \frac{\partial S_0}{\partial x} = 0 \quad [1.85]$$

Note that the purpose is to rewrite equation [1.85] in the form [1.17], recalled here:

$$\frac{\partial U}{\partial t} + \lambda \frac{\partial U}{\partial x} = S'$$

To do so, it is sufficient to define U as the cross-sectional area A and λ and S' as follows:

$$\left. \begin{aligned} \lambda &= \frac{\partial Q}{\partial A} \\ S' &= -\frac{\partial Q}{\partial K_{\text{Str}}} \frac{\partial K_{\text{Str}}}{\partial x} - \frac{\partial Q}{\partial S_0} \frac{\partial S_0}{\partial x} \end{aligned} \right\} \quad [1.86]$$

Equation [1.85] becomes:

$$\frac{\partial A}{\partial t} + \lambda \frac{\partial A}{\partial x} = S' \quad [1.87]$$

The variations in the cross-sectional area A propagate at the speed λ , as illustrated by the characteristic form of the equation:

$$\frac{dA}{dt} = S' \text{ for } \frac{dx}{dt} = \lambda \quad [1.88]$$

Note that the cross-sectional area is a conserved variable (see equation [1.84]) but it is not necessarily an invariant quantity (note the source term in equation [1.88]). If S' is non-zero, A is not invariant along a characteristic line.

Also note that, Q being a function of A via relationship [1.81], its variations also propagate at the speed λ . This can be shown in a more rigorous way by rewriting equation [1.87] as:

$$\frac{\partial A}{\partial Q} \frac{\partial Q}{\partial t} + \lambda \frac{\partial A}{\partial Q} \frac{\partial Q}{\partial x} = S' \quad [1.89]$$

Dividing by $\partial A / \partial Q$ yields:

$$\frac{\partial Q}{\partial t} + \lambda \frac{\partial Q}{\partial x} = \frac{\partial Q}{\partial A} S' \quad [1.90]$$

Substituting the first equation [1.86] into equation [1.90] leads to:

$$\frac{\partial Q}{\partial t} + \lambda \frac{\partial Q}{\partial x} = \lambda S' \quad [1.91]$$

which leads to the characteristic form for Q :

$$\frac{dQ}{dt} = \lambda S' \quad \text{for} \quad \frac{dx}{dt} = \lambda \quad [1.92]$$

1.5.3. Expression of the wave speed

The wave speed is equal to $\lambda = \partial Q / \partial A$. This derivative is difficult to estimate in the general case because the relationship between the wetted perimeter and the cross-sectional area can rarely be determined analytically. The difficulty can be removed by involving the water depth h (that is, the difference between the elevation of the free surface and the lowest point of the river bed):

$$\lambda = \frac{\partial Q}{\partial h} \frac{\partial h}{\partial A} = \frac{\partial Q}{\partial h} \left(\frac{\partial A}{\partial h} \right)^{-1} \quad [1.93]$$

The analytical expression of $\partial A / \partial h$ is derived very easily by noting that an infinitesimal variation dh in the water depth results in an infinitesimal variation dA in the cross-sectional area (Figure 1.16):

$$dA = b \, dh \quad [1.94]$$

hence the relationship:

$$\frac{\partial A}{\partial h} = b \quad [1.95]$$

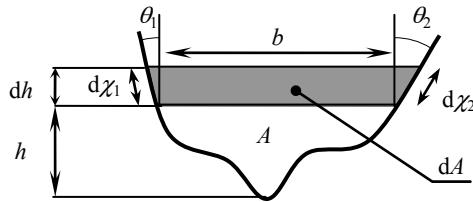


Figure 1.16. Definition sketch for the derivatives of A and R_H with respect to h

The derivative $\partial Q / \partial h$ is obtained from equation [1.83]:

$$\begin{aligned} \frac{\partial Q}{\partial h} &= K_{\text{Str}} S_0^{1/2} \frac{\partial}{\partial h} \left(\frac{A^{5/3}}{\chi^{2/3}} \right) \\ &= K_{\text{Str}} S_0^{1/2} \left[\frac{5}{3} \left(\frac{A}{\chi} \right)^{2/3} \frac{\partial A}{\partial h} - \frac{2}{3} \left(\frac{A}{\chi} \right)^{5/3} \frac{\partial \chi}{\partial h} \right] \end{aligned} \quad [1.96]$$

The expression of $\partial \chi / \partial h$ is determined by considering the variation $d\chi$ caused by an infinitesimal variation dh in the water depth:

$$d\chi = \left(\frac{1}{\cos \theta_1} + \frac{1}{\cos \theta_2} \right) dh \quad [1.97]$$

which leads to:

$$\frac{\partial \chi}{\partial h} = \frac{1}{\cos \theta_1} + \frac{1}{\cos \theta_2} \quad [1.98]$$

Substituting equations [1.95] and [1.98] into equations [1.96] and [1.93] yields the final expression of λ :

$$\lambda = K_{\text{Str}} S_0^{1/2} \left[\frac{5}{3} R_H^{2/3} - \frac{2 R_H^{5/3}}{3b} \left(\frac{1}{\cos \theta_1} + \frac{1}{\cos \theta_2} \right) \right] \quad [1.99]$$

Note that the flow velocity u is given by:

$$u = \frac{Q}{A} = K_{\text{Str}} S_0^{1/2} R_H^{2/3} \quad [1.100]$$

This allows equation [1.99] to be rewritten as:

$$\lambda = \frac{5}{3} u - \frac{2R_H}{3b} \left(\frac{1}{\cos \theta_1} + \frac{1}{\cos \theta_2} \right) u \quad [1.101]$$

The following remarks can be made:

- The wave speed is different from the flow velocity. In the general case, λ is larger than u . In the particular case of a very wide rectangular channel, the ratio of the wave speed to the flow velocity is exactly 5/3 (see section 1.5.4).
- The steeper the slope, the faster the wave for a given water depth. A large Strickler coefficient, which corresponds to a small friction term, also induces a large wave speed. Friction therefore contributes to a reduced wave speed.

– For a given value of the hydraulic radius, non-vertical embankments contribute to reduce the wave speed. Indeed, the milder the embankments, the larger the tangents of the angles θ_1 and θ_2 and, from equation [1.99], the smaller the value of λ . Note that this is true not only when the channel widens from the bottom to the top (i.e. for positive θ_1 and θ_2), but also when the channel narrows down (i.e. for negative θ_1 and θ_2). This is because, for a given increase in h , non-vertical embankments induce a larger increase in the wetted perimeter (hence, in the friction) than vertical embankments would. Friction that is exerted at the walls increases faster when the embankments are not vertical, which contributes to reducing the wave speed.

– However, a section that widens from the bottom to the top yields a larger wave speed than a section that narrows down from the bottom to the top because the hydraulic radius increases faster with h .

– In the general case, the profile of a perturbation that propagates in the channel is subject to deformation in time. This is because λ is a function of R_H and b and therefore of A . Consequently, if A is variable, λ is also variable, which induces a deformation in the profile $A(x)$. This also means that the profile shape of Q is also altered as it travels in the channel. The assumption of a constant shape for the perturbation is valid only if the amplitude of the perturbation is small enough for the wave speed to be considered constant.

– A perturbation with a small amplitude may also be subjected to deformation as it travels in the channel. Indeed, if the Strickler coefficient or the bed slope are

variable in space, the source term S' in equation [1.17] is non-zero (see equation [1.86]) and A is not an invariant along a characteristic line.

To summarize, the assumption of zero deformation for a perturbation is valid only if the amplitude of the perturbation is small enough and if the parameters that govern friction (i.e. the Strickler coefficient and the channel slope) are constant in space.

1.5.4. Particular case: flow in a rectangular channel

Consider a rectangular channel, the (constant) width b of which is very large compared to the water depth h . The Strickler coefficient and the slope are assumed to be constant. In this case the area and the hydraulic radius are given by:

$$\left. \begin{aligned} A &= bh \\ R_H &= \frac{bh}{b + 2h} \end{aligned} \right\} \quad [1.102]$$

If h is very small compared to b , the hydraulic radius is equivalent to the water depth h . This assumption is known as the “wide channel approximation”. The expression for the discharge is then simplified into:

$$Q = K_{\text{Str}} b S_0^{1/2} h^{5/3} = \frac{K_{\text{Str}} S_0^{1/2}}{b^{2/3}} A^{5/3} \quad [1.103]$$

Substituting equation [1.103] into equation [1.84] leads to the conservation form:

$$\frac{\partial A}{\partial t} + \frac{\partial}{\partial x} \left(\frac{K_{\text{Str}} S_0^{1/2}}{b^{2/3}} A^{5/3} \right) = 0 \quad [1.104]$$

The non-conservation form of equation [1.104] is:

$$\frac{\partial A}{\partial t} + \frac{5}{3} \frac{K_{\text{Str}} S_0^{1/2} A^{2/3}}{b^{2/3}} \frac{\partial A}{\partial x} + \frac{A^{5/3}}{b^{2/3}} \frac{\partial}{\partial x} \left(K_{\text{Str}} S_0^{1/2} \right) = 0 \quad [1.105]$$

which is equivalent to the non-conservation form [1.87], recalled here:

$$\frac{\partial A}{\partial t} + \lambda \frac{\partial A}{\partial x} = S'$$

provided that λ and S' are defined as:

$$\left. \begin{aligned} \lambda &= \frac{5}{3} \frac{K_{\text{Str}} S_0^{1/2} A^{2/3}}{b^{2/3}} \\ S' &= -\frac{A^{5/3}}{b^{2/3}} \frac{\partial}{\partial x} \left(K_{\text{Str}} S_0^{1/2} \right) = 0 \end{aligned} \right\} [1.106]$$

From equation [1.103], we have:

$$\lambda = \frac{5}{3} \frac{Q}{A} = \frac{5}{3} u [1.107]$$

1.5.5. Summary

The conservation form of the kinematic wave equation is given by equation [1.84]. Its non-conservation form is equation [1.87]. Its characteristic form is equation [1.88].

The cross-sectional area A and the discharge Q propagate at the same wave speed λ given by equations [1.99] and [1.101]. The wave speed increases with the channel slope, the Strickler coefficient and the slope of the embankments. It is an increasing function of the water depth h .

The cross-sectional area is the conserved variable in the conservation equation. It is not an invariant along the characteristic lines when the channel slope or the Strickler coefficient are variable in space.

1.6. A non-convex conservation law: the Buckley-Leverett equation

1.6.1. Physical context – conservation form

The Buckley-Leverett equation [BUC 42] is one of the simplest existing models for the propagation of non-aqueous phase liquids (typically, hydrocarbons) and water in porous media. The displacement of a single fluid in a porous medium can be described by Darcy's law, initially derived for water:

$$V_d = -K \frac{\partial H}{\partial x} [1.108]$$

where K is the hydraulic conductivity, V_d is the so-called Darcy velocity and H is the hydraulic head, defined as:

$$H = \frac{p}{\rho g} + z + \frac{v^2}{2g} \quad [1.109]$$

where g is the gravitational acceleration, p is the pressure and ρ is the density. Darcy's law reflects the assumption that the water flows slowly enough in the soil pores for the head loss to be proportional to the velocity.

In other words, the flow regime is assumed to be laminar, in contrast with the turbulent regime, where the head loss is assumed to be proportional to the square of the velocity. When the flow velocity is very small, the term $v^2/(2g)$ can be neglected in equation [1.109] and the expression of the hydraulic head becomes:

$$H \approx \frac{p}{\rho g} + z \quad [1.110]$$

Note that the hydraulic conductivity K is a function of both the geometric properties of the porous medium and the physical characteristics of the fluid. A smaller pore size induces a higher friction against the grains, which means that a larger pressure gradient is needed for the fluid to move at a given velocity.

Darcy's law is applicable to soils saturated with water, but also to porous media saturated with other liquids, with the difference that the conductivity K must be changed to reflect the difference between the density and viscosity of the fluid and that of water.

When the porous medium contains both water and a hydrocarbon, Darcy's law is not applicable. There are two reasons for this:

- the water and the hydrocarbon are said to be immiscible. They are present as two separate fluids in the soil. No dissolution or mixing occurs between these two liquids;

- the surface tension of a hydrocarbon is different from that of water. The interface between the water and the hydrocarbon is curved and the pressure is different on both sides of the interface. The pressure being larger in the hydrocarbon, the hydrocarbon appears to be “trapped” in the aqueous phase (Figure 1.17). The hydrocarbon is said to be a non-wetting fluid.

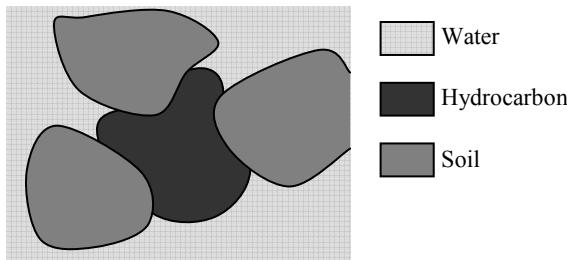


Figure 1.17. Hydrocarbon-water interface in a porous medium

Since the pressure is different in the water and in the hydrocarbon, the motion of these two fluids should not be expected to obey the same equation. In principle, an equation such as [1.108] should be needed for each of the liquids, with an additional relationship between the pressure jump across the water-hydrocarbon interface and the saturation in each liquid. Such models exist and are used on a daily basis in petroleum reservoir simulation. The most sophisticated models available describe three-phase behaviors: the water, the hydrocarbon and a gas phase with a volatilized fraction of hydrocarbon. Such models are very time-consuming and require powerful computers. In 1942, Buckley and Leverett had nevertheless proposed a single equation that accounts for the essential properties of water and hydrocarbon motion in porous media.

The Buckley-Leverett equation is a scalar law. Its flux function takes the form:

$$F = \frac{s^2}{s^2 + (1-s)^2 b_{BL}} V_d \quad [1.111]$$

where F is the water flux, b_{BL} is a shape parameter (see Figure 1.8) and s is the saturation, defined as the ratio of the volume filled with water to the porosity (that is, the total volume that can be occupied by a fluid).

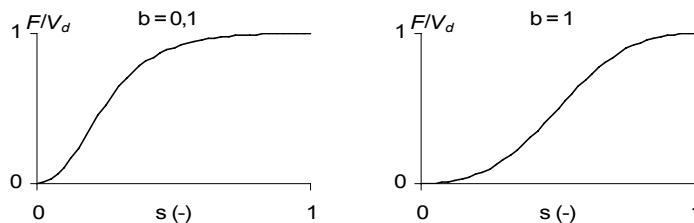


Figure 1.18. Flux function for the Buckley-Leverett equation

By definition, the saturation s lies within the interval $[0, 1]$. The volume occupied by the hydrocarbon is $1 - s$. Figure 1.18 illustrates the variations of the flux function F with the saturation for two values of the coefficient b_{BL} . When $b_{BL} = 1$ the flux function is symmetrical with respect to the point $s = 1/2$. An important feature of F is that its derivative with respect to s is equal to zero for $s = 0$ and $s = 1$. This results in specific aspects of hydrocarbon mobility, as detailed in section 1.6.2.

The Buckley-Leverett equation is derived from a balance over a slice of soil of unit cross-sectional area, the width of which is denoted by δx (Figure 1.11). The mass M of the water contained in this volume is given by:

$$M = \rho s \delta x \quad [1.112]$$

where ρ is the water density (assumed constant hereafter). The mass of water that passes at the abscissa x_0 per unit time is given by:

$$\delta F(x_0) = \rho F(x_0) \delta t \quad [1.113]$$

By reasoning as in section 1.1.2 (see equations [1.11] to [1.16]), the following equation is obtained:

$$\frac{\partial}{\partial t}(\rho s) + \frac{\partial}{\partial x}(\rho F) = 0 \quad [1.114]$$

Dividing by ρ yields the conservation form of the Buckley-Leverett equation:

$$\frac{\partial s}{\partial t} + \frac{\partial F}{\partial x} = 0 \quad [1.115]$$

where F is defined as in equation [1.111].

1.6.2. Characteristic form

The characteristic form of the equation is obtained by first rewriting equation [1.115] in non-conservation form

$$\frac{\partial s}{\partial t} + \lambda \frac{\partial s}{\partial x} = 0 \quad [1.116]$$

where $\lambda = \partial F / \partial s$. The expression of λ is obtained by differentiating equation [1.111] with respect to s :

$$\lambda = \frac{\partial F}{\partial s} = 2 \frac{(1-s)s}{[s^2 + (1-s)^2 b_{BL}]^2} b_{BL} V_d \quad [1.117]$$

Figure 1.19 illustrates the variations of λ for two different values of the shape parameter b_{BL} . Note that for $b_{BL} = 1$, the curve $\lambda(s)$ is symmetrical with respect to the axis $s = 1/2$.

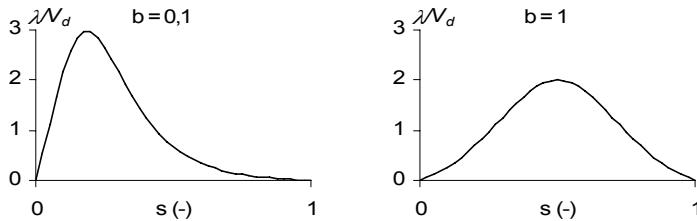


Figure 1.19. Wave speed as a function of the water saturation for the Buckley-Leverett equation

A similar reasoning to that of section 1.1.4 (see equations [1.24] to [1.27]) allows equation [1.116] to be rewritten in characteristic form as:

$$\frac{ds}{dt} = 0 \quad \text{for} \quad \frac{dx}{dt} = \lambda \quad [1.118]$$

an alternative writing for which is:

$$s = \text{Const} \quad \text{for} \quad \frac{dx}{dt} = \lambda \quad [1.119]$$

The water saturation s is a conserved quantity. It is also a Riemann invariant along the characteristics. From a physical point of view, λ is the speed at which the interface between the water and the hydrocarbon moves. The following points may be noted:

- the wave speed λ is equal to 0 for $s = 0$. This reflects a well-known property of hydrocarbons as non-wetting fluids, the mobility of which is very low at small water saturations;
- λ is maximum for an intermediate value s_{\max} of s between 0 and 1. The hydrocarbon achieves its maximum mobility for intermediate values of s ;

$-\lambda$ is equal to 0 for $s = 1$. As a well-known consequence of hydrocarbon mobility properties, decontaminating totally an aquifer initially contaminated by hydrocarbons is almost impossible. A classical decontamination method consists of injecting pure water at one or several points in the aquifer, while the immiscible mixture is pumped from the aquifer at other points. Such a method is not totally efficient in that the mobility of the hydrocarbon decreases with its saturation. Removing the remaining 2% of hydrocarbons from an aquifer may be more time-consuming than the removal of the first 98% in some cases.

1.6.3. Example: decontamination of an aquifer

Consider an aquifer of width L partly contaminated by a hydrocarbon. The initial water saturation in the aquifer is assumed to be uniformly equal to s_0 . For the sake of clarity, s_0 is assumed to be larger than the saturation s_{\max} for which λ is maximum. Pure water is injected at the left-hand boundary of the aquifer ($x = 0$) from $t = 0$ onwards. The Darcy velocity is denoted by V_d . The water saturation thus rises instantaneously to $s = 1$ at the left-hand boundary. The saturation profile is made continuous by assuming that s decreases linearly from $s = 1$ to $s = s_0$ between $x = 0$ and $x = \varepsilon$ (Figure 1.20).



Figure 1.20. Using a continuous profile between $x = 0$ and $x = \varepsilon$ to represent a discontinuous saturation profile

The width ε of the saturation ramp can be made arbitrarily small, a discontinuous profile being obtained for $\varepsilon = 0$. The solution at $t > 0$ is obtained using a diagram in the phase space (Figure 1.21). Consider a point A initially located at $x = x_A$, such that $s_0 < s_A < 1$. The saturation being invariant along a characteristic line, λ is also invariant and the characteristics are straight lines in the phase space.

$$x_{A'} = x_A + \lambda(s_A)t = \lambda(s_A)t \quad [1.120]$$

Since s is an invariant along the characteristic line, the saturation at A' is identical to the saturation at A. The coordinates $(x_{A'}, s_{A'})$ of A' in the phase space are given by:

$$\left. \begin{aligned} x_{A'} &= \lambda(s_A) t \\ s_{A'} &= s_A \end{aligned} \right\} [1.121]$$

As a consequence of equation [1.121], the point $s = 1$ does not move because $\lambda = 0$ for $s = 1$. The saturation profile decreases from $s = 1$ at $x = 0$ to $s = s_0$ at $x = \lambda(s_0) t$. The saturation remains uniformly equal to s_0 for all $x > x_0$ (Figure 1.21).

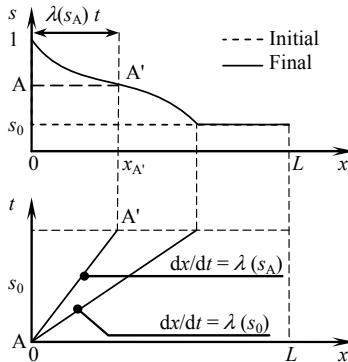


Figure 1.21. Representation of the solution in the physical space (top) and the phase space (bottom)

The complete equation of the profile at $t > 0$ is:

$$s = \begin{cases} \lambda^{-1}(x/t) & \text{for } x \leq \lambda(s_0) t \\ s_0 & \text{for } x \geq \lambda(s_0) t \end{cases} [1.122]$$

where λ^{-1} denotes the reciprocal function of $\lambda(s)$. $\lambda^{-1}(x/t)$ represents the value of s for which $\lambda = x/t$.

Note that the solution depends on x/t only. The solution is said to be self-similar. The self-similarity property of the solution is an important feature of the Riemann problem, an initial-value problem used in a number of numerical methods for wave propagation (see Chapter 4 for more details).

1.6.4. Summary

The conservation form of the Buckley-Leverett equation is given by equation [1.115]. It can be written in the general form [1.1] by defining U as the water saturation and the flux F as in equation [1.111].

The non-conservation form of the equation is given as in equation [1.116–117]. Its characteristic form is given by equation [1.118].

The water saturation is the conserved quantity. It is also a Riemann invariant for the Buckley-Leverett equation.

1.7. Advection with adsorption/desorption

1.7.1. Physical context – conservation form

The linear advection equation is treated in detail in section 1.2. It expresses the advective transport of a non-reactive, conservative substance (also called a “tracer”). The situation is now considered where the transported substance is subjected to adsorption/desorption on a substratum. This is typically the case with pesticides in groundwater.

Adsorption/desorption is a physical-chemical phenomenon that reflects the ability of the transported substance to interact with a substratum and, possibly, to remain fixed on it. The substance is transported at a concentration C_T in a liquid or gas phase that is in direct contact with the substratum (Figure 1.22). Part of the substance reacts with the substratum and is immobilized. The quantity of immobile substance may be expressed as a mass concentration (that is, the mass of adsorbed substance per unit mass of substratum) denoted by C_A . As in any chemical reaction, the adsorption phenomenon does not take place instantaneously. It is conditioned by a reaction kinetics. In many cases however, the adsorption kinetics is rapid enough to justify the assumption that the concentrations in the fluid and the adsorbed substance are at equilibrium.

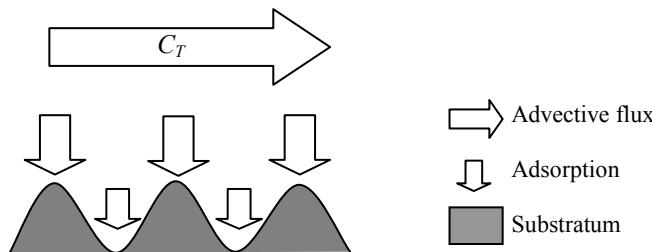


Figure 1.22. Definition sketch of the adsorption phenomenon

The concentration C_A in the immobile phase is generally not proportional to the transported concentration C_T because the number of adsorption sites in the substratum is not infinite. Moreover, all of the available adsorption sites are not

equally accessible to the dissolved substance. For small values of C_T , the most accessible adsorption sites are occupied first and C_A increases quickly with C_T . As C_T increases, the adsorption sites become less and less accessible and the function that relates C_A to C_T tends to an asymptote. Three main models are available for adsorption/desorption (Figure 1.23):

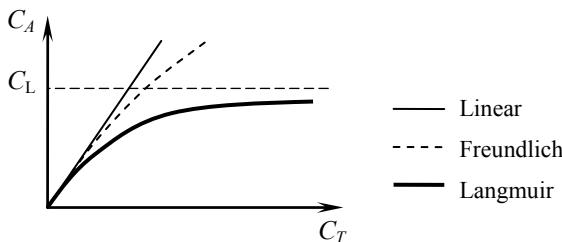


Figure 1.23. Linear, Freundlich and Langmuir adsorption isotherms

– in the linear law C_A is assumed to be directly proportional to C_T :

$$C_A = k_{\text{lin}} C_T \quad [1.123]$$

where k_{lin} is a constant;

– the Freundlich adsorption model uses the implicit assumption that the number of adsorption sites is infinite. However, the adsorption sites become less and less accessible as the concentration in the immobile phase (and therefore in the mobile phase) increases:

$$C_A = k_F C_T^b \quad [1.124]$$

where b is an exponent smaller than unity and k_F is a constant;

– the Langmuir adsorption model is based on the assumption of a limited number of equally accessible adsorption sites. The number of sites being limited, there is a physical limit to the adsorbed concentration C_A . The maximum concentration is reached asymptotically when the transported concentration C_T tends to infinity. The resulting law takes the form:

$$C_A = C_L \frac{k_L C_T}{1 + k_L C_T} \quad [1.125]$$

where C_L is the maximum permissible concentration in the immobile phase and k_L is a constant.

Consider a substance dissolved in water. The water flows in a soil with which the dissolved substance is likely to react (Figure 1.24). It is assumed that the soil is fully saturated with water, with no gas phase into which the substance may volatilize. The substance is dissolved at a concentration C_T and immobilized on the substratum at a mass concentration C_A . The conservation form of the transport equation is derived as detailed in section 1.1.2 via a balance on a control volume of unit section, the length δx of which is made arbitrarily small to lead to the differential form of the equation. The Darcy velocity (i.e. the liquid discharge per unit cross-section in the soil) is assumed to be constant in what follows.

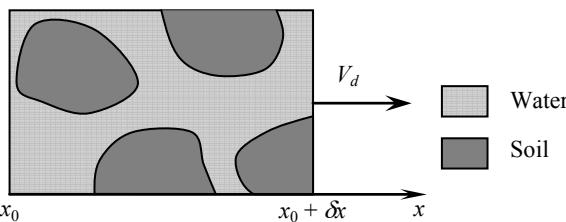


Figure 1.24. Transport and adsorption of a reactive substance in a soil

The mass M of substance contained in the control volume is the sum of two terms:

- the mass M_T of the product in the mobile phase (that is, the water) is given by

$$M_T = \theta C_T \delta x \quad [1.126]$$

where θ is the water content;

- the mass M_A of substance contained in the immobile phase is:

$$M_A = \rho_A C_A \delta x \quad [1.127]$$

where ρ_A is the soil density. The total mass of substance contained in the control volume is:

$$M = M_T + M_A = (C_T + \rho_A C_A) \delta x \quad [1.128]$$

The advection flux is given by the product of the liquid discharge and the concentration in the mobile phase, that is:

$$F = V_d C_T \quad [1.129]$$

Using the same reasoning as in section 1.1.2, the following equation is obtained for mass conservation:

$$\frac{\partial M}{\partial t} + \frac{\partial}{\partial x}(V_d C_T) = 0 \quad [1.130]$$

Since C_A is a (known) function of C_T , the quantity $C_T + \rho_A C_A$ is related to C_T by a one-to-one relationship. The so-called retardation factor R_F is introduced:

$$R_F = \frac{M}{\theta C_T} = \frac{\theta C_T + \rho_A C_A}{\theta C_T} = 1 + \frac{\rho_A C_A}{\theta C_T} \quad [1.131]$$

Equation [1.130] becomes:

$$\frac{\partial M}{\partial t} + \frac{\partial}{\partial x}\left(\frac{V_d}{\theta R_F} M\right) = 0 \quad [1.132]$$

1.7.2. Characteristic form

Differentiating the flux F with respect to the conserved variable M leads to the characteristic form of equation [1.132]:

$$\frac{\partial M}{\partial t} + \frac{V_d}{\theta R_F} \frac{\partial M}{\partial x} - \frac{V_d M}{\theta R_F^2} \frac{\partial R_F}{\partial x} = 0 \quad [1.133]$$

The derivative of the retardation factor R_F is expressed as a function of $\partial M / \partial x$:

$$\frac{\partial R_F}{\partial x} = \frac{dR_F}{dM} \frac{\partial M}{\partial x} \quad [1.134]$$

The quantity dF/dM is determined solely from M because R_F is a function of C_T (see equation [1.131]) and C_T is related to M by a one-to-one relationship like equations [1.123] to [1.125]. Substituting equation [1.134] into equation [1.133] leads to the following equation in non-conservation form:

$$\frac{\partial M}{\partial t} + \left(\frac{1}{R_F} - \frac{M}{R_F^2} \frac{dR_F}{dM} \right) \frac{V_d}{\theta} \frac{\partial M}{\partial x} = 0 \quad [1.135]$$

that can be rewritten in the form [1.17] by defining the wave speed λ as follows:

$$\lambda = \left(\frac{1}{R_F} - \frac{M}{R_F^2} \frac{dR_F}{dM} \right) \frac{V_d}{\theta} \quad [1.136]$$

Note that V_d/θ is the so-called pore velocity, that is, the speed at which the water molecules travel in the soil.

Also note that the total mass M and the concentrations C_T and C_A move at the same speed. Indeed, equation [1.135] can also be rewritten in the form [1.17] as:

$$\frac{\partial M}{\partial t} + \lambda \frac{\partial M}{\partial x} = 0 \quad [1.137]$$

Introducing the derivative of M with respect to C_T leads to:

$$\frac{\partial M}{\partial C_T} \frac{\partial C_T}{\partial t} + \lambda \frac{\partial M}{\partial C_T} \frac{\partial C_T}{\partial x} = 0 \quad [1.138]$$

which can be simplified into:

$$\frac{\partial C_T}{\partial t} + \lambda \frac{\partial C_T}{\partial x} = 0 \quad [1.139]$$

Equation [1.139] is equivalent to the following characteristic form:

$$\frac{dM}{dt} = \frac{dC_T}{dt} = 0 \quad \text{for } \frac{dx}{dt} = \lambda \quad [1.140]$$

When chemicals such as pesticides are dealt with, the solute concentration remains small enough for the retardation factor to be reasonably considered to be almost constant. The wave speed λ can therefore be approximated as follows:

$$\lambda \approx \frac{V_d}{\theta R_F} \quad [1.141]$$

It should be remembered however that equation [1.141] is only an approximation and that the complete expression of R_F is given by equation [1.136]. It is clearly visible from equations [1.136] and [1.141] that the speed at which the concentration propagates is not the pore velocity but the pore velocity divided by a factor that can be approximated with R_F . This is why R_F is called the retardation factor.

1.7.3. Summary

The conservation form of the advection equation with adsorption-desorption is given by equation [1.132]. Its non-conservation form is equation [1.137], its characteristic form is equation [1.140].

The conserved variable is the total mass of product (in the mobile and in the immobile phase) per unit volume of soil. This quantity is also an invariant along the characteristic curves.

The concentration in the liquid phase being related solely to the total mass of product per unit volume of soil, it is also an invariant that moves at the same speed.

The wave speed of the concentration wave is smaller than that of the liquid in which it is transported. The wave speed is equal to the velocity of the liquid phase divided by the retardation factor.

1.8. Summary of Chapter 1

1.8.1. What you should remember

Scalar hyperbolic conservation laws may be written in conservation, non-conservation or characteristic form. Such equations express the movement of a given quantity at a finite speed, called the wave speed.

The conservation form is derived from a balance over a control volume, the size of which is made arbitrarily small, over a time interval that is decreased to zero so as to allow the time and space derivatives of the conserved quantity and the flux function to be introduced into the equation (see section 1.1.2).

The expression of the wave speed is obtained by differentiating the flux function with respect to the conserved variable (see section 1.1.4). The trajectory, the speed of which is the wave speed, is called a characteristic. A quantity that remains constant along a characteristic is called a Riemann invariant (see section 1.1.4).

The solution can be determined solely over a subdomain $[0, L]$ of space provided that the initial condition (that is, the value of the variable at all points of the domain) and the boundary conditions are known. Zero, one or two boundary conditions may be needed depending on the flow configuration (see section 1.2.2).

The conserved variable is not necessarily an invariant along the characteristic lines (see the examples of the linear advection in section 1.3, the inviscid Burgers equation in section 1.4, the kinematic wave in section 1.5). However, the conserved

variable may also be a Riemann invariant under certain conditions (see sections 1.5, 1.6 and 1.7).

In the general case the wave speed is different from the fluid velocity (see sections 1.5 to 1.7), although exceptions exist (see sections 1.3 and 1.4).

1.8.2. Application exercises

1.8.2.1. Exercise 1.1: the inviscid Burgers equation

Solve the Burgers equation for the following initial conditions:

- the flow velocity is uniformly equal to u_1 for $x < x_1$. It is uniformly equal to u_2 for $x > x_2 > x_1$. It increases linearly from $u_1 < u_2$ to u_2 between x_1 and x_2 ;
- the density is uniformly equal to ρ_0 over the domain.

1) Derive the analytical formula for the profiles $u(x)$ and $\rho(x)$ at $t > 0$.

2) Sketch the behavior of the solution in the physical space and in the phase space.

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

1.8.2.2. Exercise 1.2: the kinematic wave equation

Derive the analytical formula for the wave speed λ in a channel, the cross-section of which is a trapezium of bottom width b_0 , the embankments of which make an angle θ with the vertical (Figure 1.25). Provide a graphical representation of λ as a function of the water depth h and carry out the numerical application for the set of parameters provided in Table 1.3. Note that the case $b_0 = 0$ corresponds to a triangular channel, in which case only positive values of θ are meaningful.

Symbol	Meaning	Value
b_0	Channel bottom width	0 m, 10 m, 40 m
K_{Str}	Strickler coefficient	40
S_0	Channel bed slope	1%
θ	Angle between the embankments and the vertical	$-30^\circ, 0^\circ, 30^\circ, 60^\circ$

Table 1.3. Parameters for Exercise 1.2

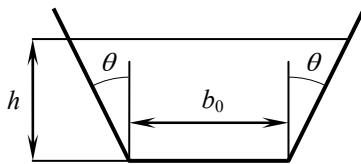


Figure 1.25. Definition sketch for the trapezoidal channel

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

1.8.2.3. Exercise 1.3: the kinematic wave equation

Consider a rectangular channel where the width b and the Strickler coefficient K_{Str} are constant. The channel is divided into three zones, the boundaries of which are located at x_1 and x_2 , with $x_1 < x_2$ (Figure 1.26):

- For $x < x_1$ the slope is uniformly equal to $S_{0,1}$.
- For $x > x_2$ the slope is uniformly equal to $S_{0,2}$.

– For x between x_1 and x_2 the slope varies continuously from $S_{0,1}$ to $S_{0,2}$. No analytical expression is available for the variations of the slope as a function of the longitudinal coordinate.

At $t = 0$, the discharge Q is uniform over the domain, which is equivalent to assuming steady state.

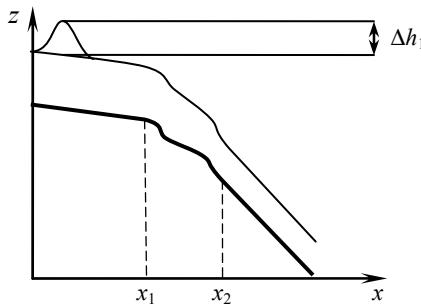


Figure 1.26. Propagation of a wave in a channel with variable slope

A discharge perturbation ΔQ_1 appears at the upstream end of the channel and propagates downstream (in the direction of positive x). The discharge perturbation is

assumed to be very small compared to the initial discharge Q , with the consequence that the wave speed can be assumed to be independent from time in the various parts of the channel.

- 1) Express the height Δh_1 of the perturbation in the upstream part of the channel ($x < x_1$) as a function of b , K_{Str} , Q , $S_{0.1}$ and ΔQ_1 .
- 2) Determine the height Δh_2 of the perturbation when it reaches $x = x_2$ and the corresponding discharge perturbation ΔQ_2 .

Carry out the numerical application for the parameter values given in Table 1.4.

Symbol	Meaning	Value
b	Channel width	40 m
K_{Str}	Strickler coefficient	40
Q	Initial discharge	10 m ³ /s
$S_{0.1}$	Channel slope upstream of x_1	0.1%
$S_{0.2}$	Channel slope downstream of x_2	1%
ΔQ_1	Discharge perturbation	1 m ³ /s

Table 1.4. Parameters for Exercise 1.3

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

1.8.2.4. Exercise 1.4: the Buckley-Leverett equation

Consider an aquifer of length L contaminated by a hydrocarbon at an initial water saturation of 50% (Figure 1.27). The remediation technique consists of injecting pure water at the left-hand boundary of the aquifer ($x = 0$). The Darcy velocity V is constant.

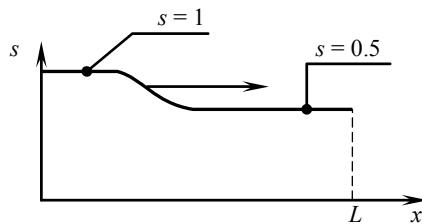


Figure 1.27. Injecting pure water to decontaminate an aquifer

- 1) Determine the shape and the expression of the saturation profile as a function of space (hint: try to express x as a function of s as opposed to the usual formulation $s(x)$).
- 2) Determine the time after which the average hydrocarbon saturation is 5%, 1% and 0.5%. Carry out the numerical application for the parameter values in Table 1.5.

Symbol	Meaning	Value
b_{BL}	Shape parameter for the Buckley-Leverett flux function	1
L	Length of the aquifer	200 m
s_0	Initial water saturation in the aquifer	0.5
s_L	Water saturation at the left-hand boundary ($x = 0$)	1
V	Darcy velocity	1 m/day

Table 1.5. Parameters for Exercise 1.4

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

1.8.2.5. Exercise 1.5: linear advection with adsorption-desorption

Consider an aquifer of length L initially contaminated by a solute with an initial concentration C_0 . The solute is adsorbed on the soil grains following Langmuir's adsorption law, the parameters of which can be found in Table 1.6.

Symbol	Meaning	Value
C_0	Initial concentration in the liquid phase	1.5 kg/m ³
C_L	Maximum adsorbed mass concentration	10 ⁻⁴ g/g
k_L	Constant in Langmuir's law	5 l/g
L	Length of the aquifer	200 m
V	Darcy velocity	1 m/day
θ	Water content (porosity) in the aquifer	0.25
ρ_A	Soil bulk density	1,500 kg/m ³

Table 1.6. Parameters for Exercise 1.6

The aquifer is decontaminated by injecting pure water at the left-hand boundary of the aquifer starting at $t = 0$ (Figure 1.28). The Darcy velocity is denoted by V . The approximation [1.141] is assumed to be valid.

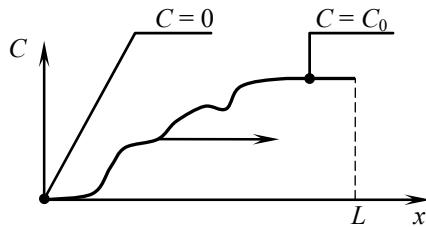


Figure 1.28. Decontaminating an aquifer by injecting pure water

1) Determine the expression and the numerical value of the time T_d at which the contaminant concentration starts to decrease at the right-hand boundary of the aquifer ($x = L$).

2) Determine the expression for the profile for $t > 0$ (searching tip: try to express x as a function of C_T as opposed to the more classical approach $C_T(x)$).

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

Chapter 2

Hyperbolic Systems of Conservation Laws in One Dimension of Space

2.1. Definitions

2.1.1. *Hyperbolic systems of conservation laws*

A system of conservation laws is a system of partial differential equations (PDEs) that can be written in conservation form as

$$\left. \begin{array}{l} \frac{\partial U_1}{\partial t} + \frac{\partial F_1}{\partial x} = S_1 \\ \vdots \\ \frac{\partial U_p}{\partial t} + \frac{\partial F_p}{\partial x} = S_p \\ \vdots \\ \frac{\partial U_m}{\partial t} + \frac{\partial F_m}{\partial x} = S_m \end{array} \right\} [2.1]$$

where the variables U_1, U_2, \dots, U_m are the m conserved variables, the quantities F_1, F_2, \dots, F_m are the corresponding fluxes and S_1, S_2, \dots, S_m are the corresponding source terms. System [2.1] of m equations for m unknowns is often referred to as an “ $m \times m$ system”. In most publications dealing with systems of conservation laws, the following vector notation is used:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} = \mathbf{S} \quad [2.2]$$

where \mathbf{U} , \mathbf{F} and \mathbf{S} are vectors of size m defined as:

$$\mathbf{U} = \begin{bmatrix} U_1 \\ \vdots \\ U_p \\ \vdots \\ U_m \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} F_1 \\ \vdots \\ F_p \\ \vdots \\ F_m \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} S_1 \\ \vdots \\ S_p \\ \vdots \\ S_m \end{bmatrix} \quad [2.3]$$

The components of the vectors \mathbf{F} and \mathbf{S} are functions of the components of \mathbf{U} and, possibly, of t and x . The following notation may also be used:

$$\left. \begin{array}{l} \mathbf{U} = [U_1, \dots, U_p, \dots, U_m]^T \\ \mathbf{F} = [F_1, \dots, F_p, \dots, F_m]^T \\ \mathbf{S} = [S_1, \dots, S_p, \dots, S_m]^T \end{array} \right\} \quad [2.4]$$

where T denotes the transposition operator. The notations [2.1] and [2.2] and [2.3] are equivalent in that the derivative of a vector is the vector formed by the derivatives of its individual components. Any reader who is not familiar with the notation of linear algebra should refer to Appendix A, which summarizes the basic notation and principles of linear algebra.

Equation [2.2] is called the conservation form of the system by analogy with the scalar conservation form [1.1] introduced in Chapter 1. The conservation form [2.2] can be rewritten in conservation form as:

$$\frac{\partial \mathbf{U}}{\partial t} + \mathbf{A} \frac{\partial \mathbf{U}}{\partial x} = \mathbf{S}' \quad [2.5]$$

where \mathbf{A} is an $m \times m$ matrix and \mathbf{S}' is a source term, not necessarily equal to \mathbf{S} . The non-conservation form [2.5] is equivalent to the conservation form [2.2] provided that \mathbf{A} and \mathbf{S}' are defined as follows:

$$\left. \begin{array}{l} d\mathbf{F} = \mathbf{A} d\mathbf{U} \\ \mathbf{S}' = \mathbf{S} - \left(\frac{\partial \mathbf{F}}{\partial x} \right)_{\mathbf{U}=\text{Const}} \end{array} \right\} \quad [2.6]$$

By definition, A is the Jacobian matrix of F with respect to U. A is calculated by differentiating the components of F with respect to the components of U:

$$A = \begin{bmatrix} \partial F_1 / \partial U_1 & \cdots & \partial F_1 / \partial U_p & \cdots & \partial F_1 / \partial U_m \\ \vdots & \ddots & \vdots & & \vdots \\ \partial F_p / \partial U_1 & \cdots & \partial F_p / \partial U_p & \cdots & \partial F_p / \partial U_m \\ \vdots & & \vdots & \ddots & \vdots \\ \partial F_m / \partial U_m & \cdots & \partial F_m / \partial U_p & \cdots & \partial F_1 / \partial U_m \end{bmatrix} \quad [2.7]$$

It can be checked easily that definition [2.7] verifies the first equation [2.6]. The vector S' contains the components of the vector S and the derivatives of F that do not depend on U. S' and S are not identical in the general case.

2.1.2. Hyperbolic systems of conservation laws – examples

A hyperbolic system of conservation laws is a conservation law system with the following properties:

- the components of F and S are functions of the components of U and possibly of x and t , but F does not contain any derivative of U with respect to x or t ;
- the Jacobian matrix A has m real, distinct eigenvalues.

Example 1: consider a river reach, the cross-section of which is rectangular, where the cross-sectional area A and the liquid discharge Q verify the kinematic wave equation [1.84] seen in section 1.5. A contaminant with a concentration C is subjected to pure advection, as described by equation [1.39] (see section 1.3). This 2×2 system can be rewritten in the vector conservation form [2.2] by defining U, F and S as follows:

$$U = \begin{bmatrix} A \\ AC \end{bmatrix}, \quad F = \begin{bmatrix} Q \\ QC \end{bmatrix} = \begin{bmatrix} K_{\text{Str}} b^{-2/3} S_0^{1/2} A^{5/3} \\ K_{\text{Str}} b^{-2/3} S_0^{1/2} A^{5/3} C \end{bmatrix}, \quad S = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad [2.8]$$

Equation [2.8] can be rewritten in the form [2.3] by defining the components of U and F as:

$$\left. \begin{array}{l} U_1 = A \\ U_2 = AC \\ F_1 = K_{\text{Str}} b^{-2/3} S_0^{-1/2} U_1^{5/3} \\ F_2 = K_{\text{Str}} b^{-2/3} S_0^{-1/2} U_1^{2/3} U_2 \end{array} \right\} \quad [2.9]$$

The Jacobian matrix A of F with respect to U is given by:

$$A = \begin{bmatrix} \frac{5K_{\text{Str}}S_0^{-1/2}}{3b^{2/3}}U_1^{2/3} & 0 \\ \frac{2K_{\text{Str}}S_0^{-1/2}}{3b^{2/3}}U_1^{-1/3}U_2 & \frac{K_{\text{Str}}S_0^{-1/2}}{b^{2/3}}U_1^{2/3} \end{bmatrix} \quad [2.10]$$

It is easy to check that A has the following eigenvalues:

$$\left. \begin{aligned} \lambda^{(1)} &= \frac{K_{\text{Str}}S_0^{-1/2}}{b^{2/3}}U_1^{2/3} = \frac{K_{\text{Str}}S_0^{-1/2}}{b^{2/3}}A^{2/3} \\ \lambda^{(2)} &= \frac{5K_{\text{Str}}S_0^{-1/2}}{3b^{2/3}}U_1^{2/3} = \frac{5}{3}\frac{K_{\text{Str}}S_0^{-1/2}}{b^{2/3}}A^{2/3} \end{aligned} \right\} \quad [2.11]$$

Note that the first eigenvalue is the flow velocity u (see equation [1.103]) and the second eigenvalue is the wave speed λ of the kinematic wave (see equation [1.106]). The eigenvalues $\lambda^{(1)}$ and $\lambda^{(2)}$ are real and distinct, therefore the system is a 2×2 system of hyperbolic conservation laws.

Example 2: consider the system formed by the continuity equation [1.62] and the momentum equation [1.63]. Remember that these equations form the basis for the derivation of the inviscid Burgers equation in section 1.4.1. System [1.62–63] can be rewritten in the form [2.2] by defining the components of U and F as:

$$\left. \begin{aligned} U_1 &= \rho \\ U_2 &= \rho u \\ F_1 &= \rho u = U_1 \\ F_2 &= \rho u^2 = U_2^2 / U_1 \end{aligned} \right\} \quad [2.12]$$

The Jacobian matrix of F with respect to U is given by:

$$A = \begin{bmatrix} 0 & 1 \\ -\left(\frac{U_2}{U_1}\right)^2 & 2\frac{U_2}{U_1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -u^2 & 2u \end{bmatrix} \quad [2.13]$$

This matrix has the following double eigenvalue:

$$\lambda^{(1)} = \lambda^{(2)} = u \quad [2.14]$$

System [1.62–63] is a system of conservation laws but it is not hyperbolic because its two eigenvalues are not distinct. Such a system is said to be “linearly degenerate”.

2.1.3. Characteristic form – Riemann invariants

The purpose is to find trajectories (called the characteristics) along which some quantities (the Riemann invariants) are constant. The present section outlines a general method for the derivation of the invariants from the Jacobian matrix A.

The starting point is the non-conservation form [2.5], recalled here:

$$\frac{\partial \mathbf{U}}{\partial t} + \mathbf{A} \frac{\partial \mathbf{U}}{\partial x} = \mathbf{S}'$$

Since the system is hyperbolic, the matrix A has m real, distinct eigenvalues with the corresponding m eigenvectors. The p th eigenvector (i.e. the eigenvector associated with the p th eigenvalue $\lambda^{(p)}$ of A) is denoted by $\mathbf{K}^{(p)}$ hereafter. The eigenvectors $\mathbf{K}^{(1)}, \mathbf{K}^{(2)}, \dots, \mathbf{K}^{(m)}$ can be seen as the columns of the $m \times m$ matrix of eigenvectors of A. This matrix is denoted by K.

$$\mathbf{K} = [\mathbf{K}^{(1)}, \mathbf{K}^{(2)}, \dots, \mathbf{K}^{(m)}] = \begin{bmatrix} \mathbf{K}_1^{(1)} & \dots & \mathbf{K}_1^{(m)} \\ \vdots & & \vdots \\ \mathbf{K}_m^{(1)} & \dots & \mathbf{K}_m^{(m)} \end{bmatrix} \quad [2.15]$$

where $K_i^{(p)}$ is the i th component of the p th eigenvector of A. An interesting property of K is that it allows A to be transformed into a diagonal matrix using the following matrix product

$$\mathbf{K}^{-1} \mathbf{A} \mathbf{K} = \Lambda \quad [2.16]$$

where \mathbf{K}^{-1} is the inverse of K and Λ is the diagonal matrix, the elements of which are the eigenvalues of A:

$$\Lambda = \begin{bmatrix} \lambda^{(1)} & & & \\ & \ddots & & 0 \\ & & \lambda^{(p)} & \\ 0 & & & \ddots \\ & & & & \lambda^{(m)} \end{bmatrix} \quad [2.17]$$

The Riemann invariants are introduced by left-multiplying the non-conservation form [2.5] by K^{-1} :

$$K^{-1} \frac{\partial U}{\partial t} + K^{-1} A \frac{\partial U}{\partial x} = K^{-1} S' \quad [2.18]$$

Noting that $K K^{-1}$ is equal to the identity matrix, equation [2.18] can be rewritten as:

$$K^{-1} \frac{\partial U}{\partial t} + K^{-1} A K K^{-1} \frac{\partial U}{\partial x} = K^{-1} S' \quad [2.19]$$

Substituting equation [2.16] into equation [2.19] leads to:

$$K^{-1} \frac{\partial U}{\partial t} + \Lambda K^{-1} \frac{\partial U}{\partial x} = K^{-1} S' \quad [2.20]$$

The vectors W and S'' are defined as:

$$\left. \begin{aligned} dW &= K^{-1} dU \\ S'' &= K^{-1} S \end{aligned} \right\} \quad [2.21]$$

The definitions [2.21] allow equation [2.20] to be rewritten as:

$$\frac{\partial W}{\partial t} + \Lambda \frac{\partial W}{\partial x} = S'' \quad [2.22]$$

an alternative writing for which is:

$$\frac{\partial}{\partial t} \begin{bmatrix} W_1 \\ \vdots \\ W_p \\ \vdots \\ W_m \end{bmatrix} + \begin{bmatrix} \lambda^{(1)} & & & & \\ & \ddots & & & \\ & & \lambda^{(p)} & 0 & \\ & & 0 & \ddots & \\ & & & & \lambda^{(m)} \end{bmatrix} \frac{\partial}{\partial x} \begin{bmatrix} W_1 \\ \vdots \\ W_p \\ \vdots \\ W_m \end{bmatrix} = \begin{bmatrix} S'_1 \\ \vdots \\ S'_p \\ \vdots \\ S'_m \end{bmatrix} \quad [2.23]$$

where W_p is the p th component of the vector W . Equation [2.23] is equivalent to the following system of independent equations in non-conservation form:

$$\frac{\partial W_p}{\partial t} + \lambda^{(p)} \frac{\partial W_p}{\partial x} = S''_p \quad \forall p = 1, 2, \dots, m \quad [2.24]$$

As shown in section 1.1.4, equation [2.24] is equivalent to the following ordinary differential equation

$$\frac{dW_p}{dt} = S_p'' \quad \text{for } \frac{dx}{dt} = \lambda^{(p)} \quad \forall p = 1, 2, \dots, m \quad [2.25]$$

The m equations [2.25] form the characteristic form of equation [2.2]. When $S'' = 0$, system [2.25] simplifies into:

$$W_p = \text{Const} \quad \text{for } \frac{dx}{dt} = \lambda^{(p)} \quad \forall p = 1, 2, \dots, m \quad [2.26]$$

The quantity W_p is called the p th Riemann invariant. Note that if U is known at a given point (x, t) in the phase space, the m Riemann invariants can be calculated by integrating the first equation [2.21]. Conversely, if the m Riemann invariants are known, U can be determined by integrating the reciprocal relationship:

$$dU = K \ dW \quad [2.27]$$

Equation [2.27] can also be rewritten as:

$$dU_p = \sum_{k=1}^m K_{p,k} \ dW_k = \sum_{k=1}^m K_p^{(k)} \ dW_k \quad [2.28]$$

where $K_{p,k} = K_k^{(p)}$ denotes the element on the p th row and the k th column of K , that is, the p th component of the k th eigenvector of A . The set of equivalent relationships [2.21] and [2.27–28] indicate that the variation dW is nothing but the expression of the variation dU in the base of eigenvectors of A .

2.2. Determination of the solution

2.2.1. Domain of influence, domain of dependence

The characteristic relationships [2.25] naturally lead to the notions of domain of influence and domain of dependence. Consider a point $A(x_0, t_0)$ in the phase space at which U is known. Along each of the m characteristic lines that pass through A , a relationship [2.25] is verified (Figure 2.1). U being known at A , each of the invariants W_p , defined as in equation [2.21], are also known at A . If the ordinary differential equation [2.25] can be solved for a given p , the value of W_p can be determined at any time $t_1 > t_0$ along the p th characteristic.

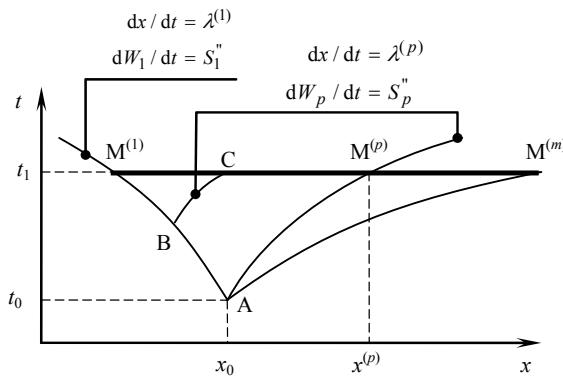


Figure 2.1. Domain of influence (bold line). Representation in the phase space

In what follows, $M^{(p)}$ denotes the intersection between the p th characteristic and the line $t = t_1$ in the phase space (Figure 2.1). The wave speeds $\lambda^{(p)}$ being ranked in ascending order, the points $M^{(1)}$ and $M^{(m)}$ are respectively the leftmost and rightmost points on the line $t = t_1$. The following reasoning is made:

- The values of all the invariants W_p at the point A are determined directly from the value of U at A because the variations in U and W obey equation [2.21]. Therefore the value of W_1 stems directly from that of U .
- The value of W_p at the point $M^{(1)}$ is a direct function of its value at A via equation [2.25].
- The value of W_p at $M^{(1)}$ influences that of U via equation [2.27].

Consequently, the value of U at the point A influences the value of U at the point $M^{(1)}$ via the first Riemann invariant. Reproducing this reasoning for any characteristic $dx/dt = \lambda^{(p)}$ leads to the conclusion that the value of U at A influences the value of U at all the points $M^{(p)}$.

Consider now a point B located on the first characteristic passing at A . The point B is located at a time between t_0 and t_1 . m characteristics can be drawn from B . The p th characteristic issued from B intersects the line $t = t_1$ at the point C . Reproducing the reasoning above along (AB) and (BC) successively leads to the conclusion that the value of U at A influences the value of U at C . This reasoning can be generalized to all the points B located on all the characteristics issued from A and to all the points C located on all the characteristics issued from all the possible locations of B . Since the possible locations of C span the segment $[M^{(1)}M^{(m)}]$, the point A influences the value of U over this segment. The segment $[M^{(1)}M^{(m)}]$ is called the domain of influence of point A .

In most analytical and numerical solution methods, the important issue is to determine the region of space over which U should be known at a given time t_0 in order to be computable at a later time t_1 . The set of points that influence point A is called the domain of dependence (Figure 2.2).

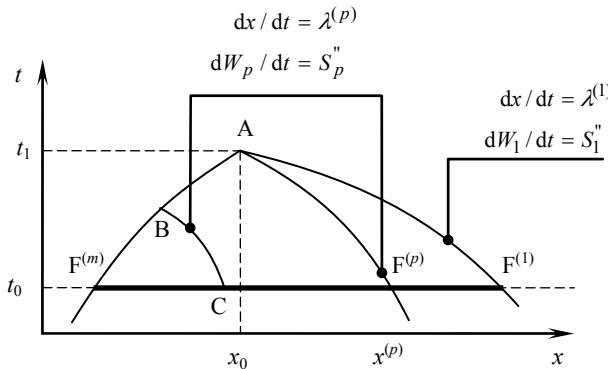


Figure 2.2. Domain of dependence (bold line). Representation in the phase space

The domain of dependence is determined in the same way as the domain of influence, except that the characteristics are followed backward in time. Let $F^{(p)}$ be the point at the intersection between the p th characteristic that passes through A and the line $t = t_0$. The point $F^{(p)}$ is called the foot of the p th characteristic. Using the same reasoning as in the previous two sections leads to the conclusion that the value of U at point A is influenced by the value of U at all the points between $F^{(1)}$ and $F^{(m)}$. The segment $[F^{(m)}F^{(1)}]$ is called the domain of dependence of A.

2.2.2. Existence and uniqueness of solutions – initial and boundary conditions

This section deals with the existence and uniqueness of the solutions of hyperbolic systems of conservation laws. Assume that U is to be determined at the point $A(x_0, t_1)$ in Figure 2.2. To do so, the m components U_1, U_2, \dots, U_m must be determined uniquely. Therefore, m independent equations in the form [2.25] must be written for these m components.

Most practical applications deal with a finite domain of length L , called the computational domain. For the sake of clarity, the point $x = 0$ is located at the left-hand boundary of the domain, the right-hand boundary of which is located at $x = L$ (Figure 2.3). Two configurations may occur:

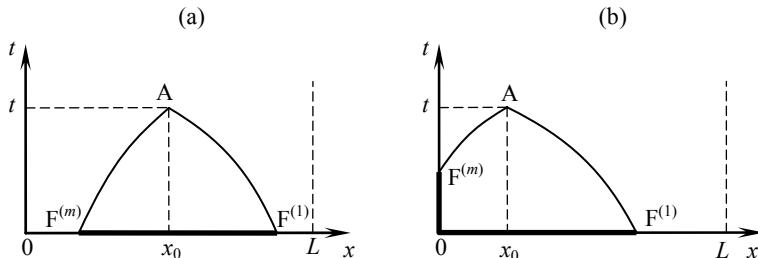


Figure 2.3. Determination of the solutions when the domain of dependence is included entirely in the computational domain (a) and when the domain of dependence includes the domain boundary (b)

– The computational domain $[0, L]$ contains entirely the domain of dependence of A (Figure 2.3a). In this case, the knowledge of U over the domain of dependence at $t = 0$ allows the value of U to be determined uniquely at A by writing the m characteristic equations [2.25] between the feet $F^{(p)}$ and A. The m equations [2.25] allow the m Riemann invariants to be determined at A. The knowledge of the m Riemann invariants leads to that of U via relationships [2.27].

– There exists points (in particular near the boundaries of the computational domain), the domain of dependence of which is not entirely included in the segment $[0, L]$ (see Figure 2.3b). In such a case, knowing the value of U over the segment $[0, L]$ is insufficient because there is at least one (and possibly more than one) characteristic entering the domain across a boundary. In Figure 2.3, the m th invariant can be determined at A only if its value is known at the left-hand boundary. The value of the invariant is known at $F^{(m)}$ provided that U is known at $F^{(m)}$. U is known at $F^{(m)}$ only if the m Riemann invariants are known. Let n_e denote the number of Riemann invariants that enter the domain at the boundary. In order to determine U uniquely at the boundary, n_e conditions must be supplied on the components of U , the remaining $m - n_e$ conditions being supplied by the invariants that travel to $F^{(m)}$ from within the domain.

To summarize, the following conditions must be prescribed for the existence and uniqueness of the solution to be guaranteed over the domain:

– the initial condition $U(x, 0)$ (that is, all the components U_p of U) must be known over the entire domain $[0, L]$ at $t = 0$;

– boundary conditions must be prescribed at the boundaries of the computational domain. The number of boundary conditions needed at a given boundary is equal to the number of characteristics that enter the domain.

2.3. A particular case: compressible flows

2.3.1. Definition

Compressible flows represent a vast majority of the flows that can be described using hyperbolic systems of conservation laws. The term “compressible flow” refers to flows with the following characteristics:

- the governing equations contain at least the continuity and the momentum equations (there may be additional equations in the system);
- the momentum equation includes a term that accounts for the pressure forces within the fluid;
- the pressure in the fluid is a function of at least the fluid density (or the mass of fluid per unit length of volume if a one-dimensional configuration is dealt with; the mass per unit surface if a two-dimensional configuration is dealt with). The law that relates the pressure to the density (and possibly to other variables) is called an equation of state;
- the system of conservation laws is hyperbolic.

2.3.2. Conservation form

The simplest possible system for compressible flows is a 2×2 system, that is, a system with two equations (the continuity equation and the momentum equation) in two independent variables. The equations for such a system are written by defining a control volume between the abscissas x_0 and $x_0 + \delta x$ and carrying out a mass and momentum balance between the times t_0 and $t_0 + \delta t$ (Figure 2.4).

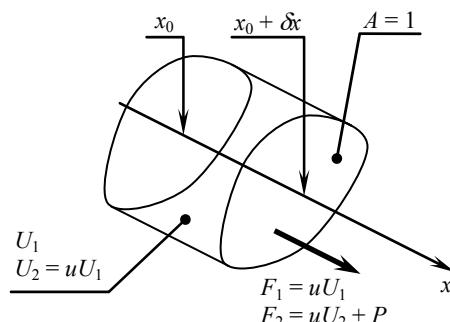


Figure 2.4. Mass and momentum balance over a control volume

The continuity equation is derived as in Chapter 1 by noting that the mass per unit length of domain (denoted by U_1 hereafter) is transported at the flow velocity u , yielding a momentum flux F_1 :

$$F_1 = uU_1 \quad [2.29]$$

As shown in Chapter 1, the continuity equation is written in conservation form as:

$$\frac{\partial U_1}{\partial t} + \frac{\partial F_1}{\partial x} = 0 \quad [2.30]$$

The momentum U_2 per unit length in the domain is defined as:

$$U_2 = uU_1 \quad [2.31]$$

Consequently the total amount of momentum in the control volume at time t is given by:

$$\delta U_2(t) = \int_{x_0}^{x_0 + \delta x} U_2(x, t) dx \quad [2.32]$$

Between the times t_0 and $t_0 + \delta t$ an amount $\delta F_{QdM}(x_0)$ of momentum enters the control volume at the left-hand boundary:

$$\delta F_{QdM}(x_0) = \int_{t_0}^{t_0 + \delta t} uU_2(x_0, t) dt \quad [2.33]$$

During the same time interval, an amount $\delta F_M(x_0 + \delta x)$ of momentum leaves the control volume at the right-hand boundary:

$$\delta F_{QdM}(x_0 + \delta x) = \int_{t_0}^{t_0 + \delta t} uU_2(x_0 + \delta x, t) dt \quad [2.34]$$

Denoting by $P(x_0)$ the pressure force exerted on the left-hand boundary ($x = x_0$), the pressure force exerted on the right-hand boundary ($x = x_0 + \delta x$) is $-P(x_0 + \delta x)$. The fundamental principle of dynamics can be written as:

$$\begin{aligned} \delta U_2(t_0 + \delta t) - \delta U_2(t_0) &= \delta F_{QdM}(x_0) - \delta F_{QdM}(x_0 + \delta x) \\ &\quad + \int_{t_0}^{t_0 + \delta t} [P(x_0, t) - P(x_0 + \delta x, t)] dt \\ &\quad + \int_{t_0}^{t_0 + \delta t} \int_{x_0}^{x_0 + \delta x} S_2(x, t) dx dt \end{aligned} \quad [2.35]$$

where S_2 is a source term due to external forces such as volume or wall forces. When δt and δx tend to zero the following equivalences hold:

$$\left. \begin{aligned} \delta U_2(t_0 + \delta t) - \delta U_2(t_0) &\underset{\substack{\delta x \rightarrow 0 \\ \delta t \rightarrow 0}}{\approx} \delta x \delta t \frac{\partial U_2}{\partial t} \\ \delta F_{QdM}(x_0) - \delta F_{QdM}(x_0 + \delta x) &\underset{\substack{\delta x \rightarrow 0 \\ \delta t \rightarrow 0}}{\approx} -\delta x \delta t \frac{\partial}{\partial x}(uU_2) \\ \int_{t_0}^{t_0 + \delta t} [P(x_0, t) - P(x_0 + \delta x, t)] dt &\underset{\substack{\delta x \rightarrow 0 \\ \delta t \rightarrow 0}}{\approx} -\delta x \delta t \frac{\partial P}{\partial x} \\ \int_{t_0}^{t_0 + \delta t} \int_{x_0}^{x_0 + \delta x} S_2(x, t) dx dt &\underset{\substack{\delta x \rightarrow 0 \\ \delta t \rightarrow 0}}{\approx} \delta x \delta t S_2 \end{aligned} \right\} [2.36]$$

Substituting equations [2.36] into equation [2.35] and simplifying by δt and δx yields:

$$\frac{\partial U_2}{\partial t} + \frac{\partial F_{QdM}}{\partial x} + \frac{\partial P}{\partial x} = S_2 \quad [2.37]$$

The flux F_2 (also known as the “specific force” [CHO 73]) is defined as:

$$F_2 = F_{QdM} + P = uU_2 + P \quad [2.38]$$

Equation [2.38] allows equation [2.37] to be written in conservation form as:

$$\frac{\partial U_2}{\partial t} + \frac{\partial F_2}{\partial x} = S_2 \quad [2.39]$$

Equations [2.30] and [2.39] can be written in the vector form [2.1], recalled here:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} = \mathbf{S}$$

by defining \mathbf{U} , \mathbf{F} and \mathbf{S} as:

$$\mathbf{U} = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} uU_1 \\ uU_2 + P \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 0 \\ S_2 \end{bmatrix} \quad [2.40]$$

with the additional relationship [2.29], that can be rewritten as:

$$\frac{U_2}{U_1} = u \quad [2.41]$$

2.3.3. Characteristic form

The characteristic form is derived as follows. In a first step the expression of the Jacobian matrix of F with respect to U is determined. To do this, the components of F must be rewritten as functions of the components of U . Substituting equation [2.41] into definition [2.40] leads to the following expression:

$$F = \begin{bmatrix} U_2 \\ U_2^2/U_1 + P \end{bmatrix} \quad [2.42]$$

Under the assumption of a compressible flow, the pressure force P is a function of U_1 . The equation of state $P = P(U)$ is assumed to be known. The conservation form [2.2] is equivalent to the non-conservation form [2.5], recalled here:

$$\frac{\partial U}{\partial t} + A \frac{\partial U}{\partial x} = S'$$

provided that A is the Jacobian matrix of F with respect to U . The expression of A stems directly from equation [2.42]:

$$A = \begin{bmatrix} 0 & 1 \\ \frac{\partial P}{\partial U_1} - \left(\frac{U_2}{U_1}\right)^2 & 2\frac{U_2}{U_1} \end{bmatrix} \quad [2.43]$$

The quantity c is defined as:

$$c = \left(\frac{\partial P}{\partial U_1} \right)^{1/2} \quad [2.44]$$

Substituting equations [2.41] and [2.44] into equation [2.43] leads to:

$$A = \begin{bmatrix} 0 & 1 \\ c^2 - u^2 & 2u \end{bmatrix} \quad [2.45]$$

The eigenvalues of A are:

$$\left. \begin{array}{l} \lambda^{(1)} = u - c \\ \lambda^{(2)} = u + c \end{array} \right\} \quad [2.46]$$

with corresponding eigenvectors $\mathbf{K}^{(1)}$ and $\mathbf{K}^{(2)}$ given by:

$$\mathbf{K}^{(1)} = \begin{bmatrix} 1 \\ u - c \end{bmatrix}, \quad \mathbf{K}^{(2)} = \begin{bmatrix} 1 \\ u + c \end{bmatrix} \quad [2.47]$$

Therefore the matrix K and its inverse are:

$$\mathbf{K} = \begin{bmatrix} 1 & 1 \\ u - c & u + c \end{bmatrix}, \quad \mathbf{K}^{-1} = \frac{1}{2c} \begin{bmatrix} c + u & -1 \\ c - u & 1 \end{bmatrix} \quad [2.48]$$

The Riemann invariants are given by the first relationship [2.21]:

$$\left. \begin{array}{l} dW_1 = \left(\frac{1}{2} + \frac{u}{2c} \right) dU_1 - \frac{1}{2} dU_2 \\ dW_2 = \left(\frac{1}{2} - \frac{u}{2c} \right) dU_1 + \frac{1}{2} dU_2 \end{array} \right\} \quad [2.49]$$

2.3.4. Physical interpretation

Consider a fluid initially at rest ($u = 0$). Due to some reason (for example, a transient originating from an external forcing), a perturbation ΔU_1 appears in the fluid at the abscissa x_0 . This perturbation triggers a perturbation in the Riemann invariants W_1 and W_2 . The perturbation in W_1 propagates at a speed $\lambda^{(1)}$, while the perturbation in W_2 propagates at a speed $\lambda^{(2)}$. If ΔU_1 is small, u remains negligible compared to c , and $\lambda^{(1)}$ and $\lambda^{(2)}$ are approximately equal to $-c$ and $+c$ respectively. Consequently, the initial perturbation in U_1 triggers two perturbations that propagate at the same speed in opposite directions (Figure 2.5). Since the pressure is a function of P , c is also called the speed of the pressure waves. In the case of a compressible gas or liquid, c is referred to as the speed of sound.

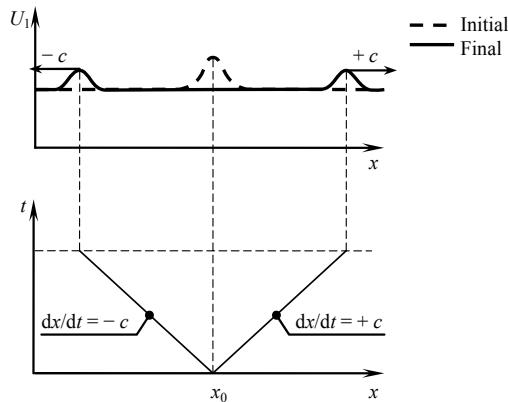


Figure 2.5. Compressible flow. Wave propagation in a fluid initially at rest

2.4. A linear 2×2 system: the water hammer equations

2.4.1. Physical context – assumptions

The water hammer equations are the simplest existing equations for compressible systems. Indeed, these equations are linear and the momentum equation is simplified by neglecting inertia. Although the first version of the complete equations is usually attributed to Joukowski [JOU 98], they are best known as the Allievi equations [ALL 03]. The derivation of these equations has been widely addressed in the literature [JAE 33, JAE 77, WYL 77].

Water hammer is a wave propagation phenomenon that occurs in systems of pressurized conduits (e.g. a drinking water supply network, pipes in hydropower plants) under rapid changes in the flow conditions. Such changes may result from the sudden opening (or closing) of a valve, pipe failure, etc., that leads to rapid variations in the discharge or pressure (or both) at a given point in the network. When the flow conditions are modified almost instantaneously, dynamical equilibrium is no longer ensured. The fluid is subjected to strong accelerations within a few hundredths or tenths of a second. Such accelerations result in local mass accumulation or deficit, depending on the case. Owing to the small compressibility of the fluid and the rigidity of the pipe material, the locally accumulating water is subjected to a local compression, while deficit regions are characterized by a pressure drop. The pressure fluctuations propagate in the pipe to yield pressure waves. The pressure variations may easily reach several atmospheres or tens of atmospheres, which may damage the pipe network severely. The simulation of water hammer episodes resulting from pipe, pump or valve failure is an important task in the design of protection devices for pressurized pipe networks.

The following example illustrates the water hammer phenomenon in a schematic way. Consider steady state water flow in a pipe at a uniform pressure p_0 and a constant discharge Q_0 . The water is pumped into the pipe at the left-hand end and leaves the pipe at the right-hand end (Figure 2.6a). At $t = 0$ the right-hand side of the pipe is closed instantaneously. The water is still being pumped at a discharge Q_0 at the left-hand end of the pipe and the discharge is still equal to Q_0 everywhere in the pipe, including immediately upstream of the downstream valve. Continuity imposes that the water that reaches the valve accumulates in the pipe next to the valve, thus creating a local pressure rise and the inflation of the conduit (Figure 2.6b). The pressure in the inflated zone is higher than the initial pressure p_0 , while continuity with the downstream section of the pipe imposes that the discharge be zero. Since the water keeps flowing to the upstream end of the inflated zone, it accumulates at its boundary. The boundary of the inflated zone where the water is immobile moves to the left to accommodate the incoming water (Figure 2.6b). The inflated zone eventually reaches the left-hand end of the pipe (Figure 2.6c). The pressure wave propagates at the speed $-c$, the typical order of magnitude of which is a kilometer per second.

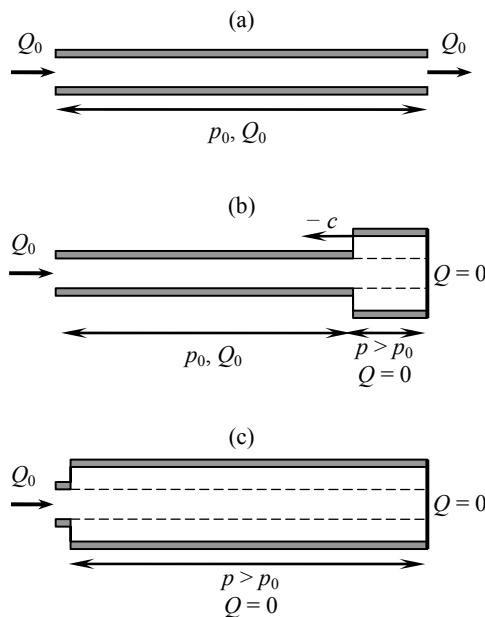


Figure 2.6. Sudden closure of a valve at the downstream end of a pipe and resulting water hammer phenomenon. Initial state (a), next to the closure of the valve (b), after the wave reaches the upstream end of the pipe (c). The dashed line indicates the initial size of the pipe

The governing assumptions of the water hammer phenomenon are the following:

– Assumption (A1). The pressure variations in a pipe may reach several millions of Pa, while the typical variations in the flow speed may be 1 m/s. An order of magnitude analysis indicates that the inertial term is negligible compared to the pressure term in the momentum equation [2.37].

– Assumption (A2). The fluid is (slightly) compressible and the conduit is (slightly) elastic. This yields a linear dependence between the pressure and the mass of fluid per unit length. The corresponding relationship, expressed as in equation [2.44], leads to the conclusion that the speed of sound c is a constant. Note that a typical order of magnitude for c is 1 km/s.

– Assumption (A3). Owing to high rigidity, the conduit is only slightly deformable, even under sever transients. The variations in the pressure force P are mainly due to the variations in the pressure, not to the variations in the cross-sectional area of the pipe.

– Assumption (A4). Owing to the weak compressibility of the fluid, the relative variations in the fluid density are very small. Consequently, the variations in the mass discharge are mainly due to the variations in the flow speed.

2.4.2. Conservation form

2.4.2.1. Notation

The water hammer phenomenon can be described using the continuity and momentum equations in one dimension of space. The assumption of a one-dimensional flow is valid because of the high contrast between the length (typically, tens to hundreds of meters, sometimes kilometers) and the diameter of the pipes (typically a few centimeters to a meter). The equations are derived in conservation form for pipes of arbitrary shape, with a variable cross-section and a non-horizontal axis. Figure 2.7 illustrates the notation and the coordinate system used.

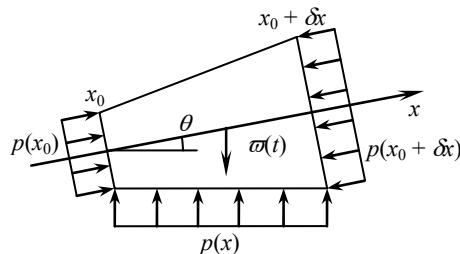


Figure 2.7. Derivation of the water hammer equations. Definition sketch for the geometry

The x -axis is that of the pipe. The angle between the x -axis and the horizontal is denoted by θ , with the convention that θ is positive when the elevation increases with x . The cross-section of the pipe is denoted by A . A mass and momentum balance is carried out between the times t_0 and $t_0 + \delta t$ over a slice of pipe delineated by $x = x_0$ and $x = x_0 + \delta x$.

2.4.2.2. Continuity equation

The mass balance can be written as follows:

$$\delta U_1(t_0 + \delta t) - \delta U_1(t_0) = \delta F_1(x_0) - \delta F_1(x_0 + \delta x) \quad [2.50]$$

where $\delta U_1(t)$ is the mass of fluid contained in the control volume at the time t and $\delta F_1(x)$ is the mass of fluid that passes at the abscissa x over the time interval δt . Equation [2.50] expresses that the variation in the mass of the control volume is due to the difference between the mass that enters and the mass that leaves the control volume. By definition, $\delta U_1(t)$ and $\delta F_1(x)$ are given by:

$$\left. \begin{aligned} \delta U_1(t) &= \int_{x_0}^{x_0 + \delta x} (\rho A)(x, t) dx \\ \delta F_1(x) &= \int_{t_0}^{t_0 + \delta t} (\rho u A)(x, t) dx \end{aligned} \right\} \quad [2.51]$$

where u is the fluid velocity and ρ is the fluid density. Substituting equation [2.51] into equation [2.50], introducing the derivatives with respect to time and space leads, in the limit of vanishing δt and δx , to the so-called continuity equation (see equations [1.12–16] for details of the proof):

$$\frac{\partial}{\partial t}(\rho A) + \frac{\partial}{\partial x}(\rho u A) = 0 \quad [2.52]$$

Defining the first component U_1 of the conserved variable as $U_1 = \rho A$, equation [2.52] can be rewritten as

$$\left. \begin{aligned} \frac{\partial U_1}{\partial t} + \frac{\partial F_1}{\partial x} &= 0 \\ U_1 &= \rho A \\ F_1 &= u U_1 \end{aligned} \right\} \quad [2.53]$$

2.4.2.3. Momentum equation

The momentum balance is given by the fundamental principle of dynamics:

$$\delta U_2(t_0 + \delta t) - \delta U_2(t_0) = \delta F_2(x_0) - \delta F_2(x_0 + \delta x) + \delta P(x_0) - \delta P(x_0 + \delta x) + \delta F_V + \delta F_P \quad [2.54]$$

where $\delta U_2(t)$ is the total momentum of the fluid over the control volume at the time t and $\delta F_2(x)$ is the momentum of the volume of fluid that passes at x over the time interval δt . $\delta P(x)$ is the integral of the pressure force exerted on the pipe cross-section at x over the time interval δt . δF_V and δF_W are respectively the integrals of the volume forces and the forces exerted by the walls on the control volume over the time interval δt . By definition, δU_2 and δF_2 are given by:

$$\left. \begin{aligned} \delta U_2(t) &= \int_{x_0}^{x_0 + \delta x} (\rho u A)(x, t) dx \\ \delta F_2(x) &= \int_{t_0}^{t_0 + \delta t} (\rho u^2 A)(x, t) dx \end{aligned} \right\} \quad [2.55]$$

The pressure force $P(x)$ being defined as the product of the pipe cross-section and the average pressure over the cross-section (Figure 2.7), δP is given by:

$$\delta P(t) = \int_{t_0}^{t_0 + \delta t} (Ap)(x, t) dt \quad [2.56]$$

The only volume force is the weight of the control volume:

$$\varpi(t) = \int_{x_0}^{x_0 + \delta x} g(\rho A)(x, t) dx \quad [2.57]$$

The weight of the control volume is a vector collinear with the vertical. Since the momentum balance [2.54] is written in the x -direction, δF_V is the projection of the integral of $\varpi(t)$ between t_0 and $t_0 + \delta t$ onto the x -axis:

$$\delta F_V(t) = \int_{t_0}^{t_0 + \delta t} -\varpi(t) \sin \theta dt = -g \sin \theta \int_{t_0}^{t_0 + \delta t} \int_{x_0}^{x_0 + \delta x} (\rho A)(x, t) dx dt \quad [2.58]$$

The integral of the forces exerted by the walls can be broken into two terms:

$$\delta F_W(t) = \int_{t_0}^{t_0 + \delta t} \int_{x_0}^{x_0 + \delta x} (R_p + R_f)(x, t) dx dt \quad [2.59]$$

where R_f and R_p are respectively the forces due to friction against the pipe wall and the projection of the reaction of the wall onto the x -axis. Under the assumption of turbulent flow, R_f is classically assumed to be proportional to the square of the flow velocity. The friction force is exerted in the opposite direction to that of the flow, which leads to an expression in the form:

$$R_f = -k|u|u \quad [2.60]$$

The term R_p is equal to the projection of the reaction of the wall to the pressure force onto the x -axis. The reaction of the wall stems from the pressure exerted by the fluid in the direction of the normal unit vector to the wall. According to the theorem of action and reaction, the reaction exerted by the wall on the fluid is the opposite of the force exerted by the fluid on the wall (Figure 2.8).

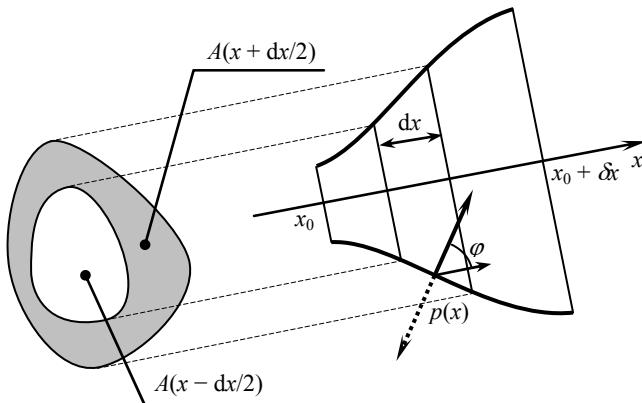


Figure 2.8. Pressure force exerted by the fluid on the wall (dashed arrow) and reaction of the wall on the fluid (solid arrow)

The projection of R_p onto the x -axis is determined by multiplying its norm by the cosine of the angle φ between the wall and the x -axis (Figure 2.8). The expression of the integral of the projected vector over the wall of the entire control volume may become rather complex when the shape of the pipe is complex. However, the developments are kept simple by noting that the only quantity of interest is the

projection of the reaction onto the x -axis. The x -component of the pressure force exerted on a slice of pipe of infinitesimal length dx centered around the abscissa x is the product of the pressure $p(x)$ over an infinitesimal area dA in the plane normal to the x -axis. dA is the variation of the cross-sectional area between x and $x + dx$:

$$dA = A(x + dx/2) - A(x - dx/2) = \frac{\partial A}{\partial x} dx \quad [2.61]$$

Integrating the pressure force over the control volume leads to:

$$\int_{x_0}^{x_0 + \delta x} R_p(x, t) dx = \int_{x_0}^{x_0 + \delta x} \frac{\partial A}{\partial x} p(x, t) dx \quad [2.62]$$

Substituting equations [2.60] and [2.62] into equation [2.59] leads to:

$$\delta F_P(t) = \int_{t_0}^{t_0 + \delta t} \int_{x_0}^{x_0 + \delta x} \left(-k|u|u + \frac{\partial A}{\partial x} p \right) (x, t) dx dt \quad [2.63]$$

Substituting equations [2.55], [2.56], [2.58] and [2.53] into [2.54] and reasoning as in equations [1.12–16], the momentum equation is written as:

$$\frac{\partial}{\partial t} (\rho A u) + \frac{\partial}{\partial x} (\rho u^2 A + P) = p \frac{\partial A}{\partial x} - g \rho A \sin \theta - k|u|u \quad [2.64]$$

where the pressure force P is equal to the product Ap . Recalling definition [2.53], the expression of the second component of the conserved variable becomes:

$$\left. \begin{aligned} U_2 &= \rho Q = F_1 \\ \rho u^2 A &= \rho \frac{Q^2}{A} = \frac{U_2^2}{U_1} \end{aligned} \right\} \quad [2.65]$$

2.4.2.4. Simplification – vector form

The expression of the second component F_2 of the flux vector can be simplified by recalling Assumption (A1) stated in section 2.4.1. Neglecting the momentum flux $\rho Q^2/A$ with respect to the pressure terms in the momentum equation leads to:

$$\frac{\partial}{\partial x} \left(\rho \frac{Q^2}{A} \right) \ll \frac{\partial P}{\partial x} \quad [2.66]$$

and equation [2.64] is simplified into:

$$\frac{\partial}{\partial t}(\rho A u) + \frac{\partial P}{\partial x} = p \frac{\partial A}{\partial x} - g \rho A \sin \theta - k|u|u \quad [2.67]$$

Equations [2.52] and [2.67] can be written in the vector form [2.2], recalled here:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} = \mathbf{S}$$

provided that \mathbf{U} , \mathbf{F} and \mathbf{S} are defined as:

$$\mathbf{U} = \begin{bmatrix} \rho A \\ \rho Q \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} \rho Q \\ Ap \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 0 \\ p \frac{\partial A}{\partial x} - \rho g A \sin \theta - k|u|u \end{bmatrix} \quad [2.68]$$

2.4.3. Characteristic form – Riemann invariants

System [2.2] can be rewritten in the non-conservation form [2.5], recalled here:

$$\frac{\partial \mathbf{U}}{\partial t} + \mathbf{A} \frac{\partial \mathbf{U}}{\partial x} = \mathbf{S}'$$

where \mathbf{A} is the Jacobian matrix of \mathbf{F} with respect to \mathbf{U} and $\mathbf{S}' = \mathbf{S}$. The definitions [2.68] lead to the following expression for \mathbf{A} :

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ c^2 & 0 \end{bmatrix} \quad [2.69]$$

where c is defined as in equation [2.44], recalled here:

$$c = \left(\frac{\partial P}{\partial U_1} \right)^{1/2} = \left(\frac{\partial (Ap)}{\partial (\rho A)} \right)^{1/2}$$

From Assumption (A2) stated in section 2.4.1, the variations in the pressure force Ap are proportional to those in the mass per unit length ρA . Consequently, c does not depend on the flow variables, it is a constant that depends only on the local properties of the pipe, such as its diameter, thickness and material.

The matrix A has the following eigenvalues and eigenvectors:

$$\lambda^{(1)} = -c, \quad \lambda^{(2)} = c$$

$$K^{(1)} = \begin{bmatrix} 1 \\ -c \end{bmatrix}, \quad K^{(2)} = \begin{bmatrix} 1 \\ c \end{bmatrix} \quad [2.70]$$

The eigenvalues being real and distinct, system [2.2, 2.68] is hyperbolic. The variations in the pressure force being related to those of the mass per unit length, the system is compressible.

The characteristic form is obtained by multiplying equation [2.5] by the matrix K^{-1} and noting that A is transformed into the diagonal matrix Λ in the base formed by its eigenvectors (see equations [2.18–22] in section 2.1.3). This leads to equation [2.22], recalled here:

$$\frac{\partial W}{\partial t} + \Lambda \frac{\partial W}{\partial x} = S'$$

where the vector dW is defined as in the first equation [2.21], recalled here:

$$dW = K^{-1} dU$$

Substituting equations [2.68] and [2.70] into equation [2.21] and multiplying by $2c$, the following expression is obtained for dW and S":

$$dW = 2c \begin{bmatrix} \frac{1}{2} & -\frac{1}{2c} \\ \frac{1}{2} & \frac{1}{2c} \end{bmatrix} \begin{bmatrix} dU_1 \\ dU_2 \end{bmatrix} = \begin{bmatrix} c d(\rho A) - d(\rho Q) \\ c d(\rho A) + d(\rho Q) \end{bmatrix} = \begin{bmatrix} dW_1 \\ dW_2 \end{bmatrix}$$

$$S'' = 2c K^{-1} S = \frac{1}{2c} \begin{bmatrix} k|u|u + \rho g A \sin \theta - \frac{\partial A}{\partial x} p \\ -k|u|u - \rho g A \sin \theta + \frac{\partial A}{\partial x} p \end{bmatrix} \quad \left. \right\} \quad [2.71]$$

Equation [2.22] can be rewritten as:

$$\frac{\partial W_p}{\partial t} + \lambda^{(p)} \frac{\partial W_p}{\partial x} = S''_p, \quad p = 1, 2 \quad [2.72]$$

The mass per unit length and the pressure force are not directly measurable in practice. Engineers usually deal with the pressure and the liquid discharge that can be easily measured using manometers and flow meters. The Riemann invariants are

rewritten so as to involve p and Q . Equations [2.44] allow the following relationships to be written:

$$\left. \begin{aligned} \frac{\partial}{\partial t}(\rho A) &= \frac{1}{c^2} \frac{\partial}{\partial t}(Ap) \\ \frac{\partial}{\partial x}(\rho A) &= \frac{1}{c^2} \frac{\partial}{\partial x}(Ap) = \frac{A}{c^2} \frac{\partial p}{\partial x} + \frac{p}{c^2} \frac{\partial A}{\partial x} \end{aligned} \right\} [2.73]$$

From Assumption (A3), the variations in the pressure force (Ap) are mainly due to the variations in p . Equation [2.73] simplifies into:

$$\left. \begin{aligned} \frac{\partial}{\partial t}(\rho A) &\approx \frac{A}{c^2} \frac{\partial p}{\partial t} \\ \frac{\partial}{\partial x}(\rho A) &= \frac{A}{c^2} \frac{\partial p}{\partial x} + \frac{p}{c^2} \frac{\partial A}{\partial x} \end{aligned} \right\} [2.74]$$

Assumption (A4) allows the following expression to be derived for the derivatives of the mass discharge:

$$\left. \begin{aligned} \frac{\partial}{\partial t}(\rho Q) &\approx \rho \frac{\partial Q}{\partial t} \\ \frac{\partial}{\partial x}(\rho Q) &\approx \rho \frac{\partial Q}{\partial x} \end{aligned} \right\} [2.75]$$

Substituting equations [2.74] and [2.75] into equations [2.71] and [2.71], introducing equation [2.70] yields the following expressions:

$$\left. \begin{aligned} \frac{A}{c} \frac{\partial p}{\partial t} - \rho \frac{\partial Q}{\partial t} - \left(\frac{A}{c} \frac{\partial p}{\partial x} + \frac{p}{c} \frac{\partial A}{\partial x} - \rho \frac{\partial Q}{\partial x} \right) c &= k|u|u + \rho g A \sin \theta - p \frac{\partial A}{\partial x} \\ \frac{A}{c} \frac{\partial p}{\partial t} + \rho \frac{\partial Q}{\partial t} + \left(\frac{A}{c} \frac{\partial p}{\partial x} + \frac{p}{c} \frac{\partial A}{\partial x} + \rho \frac{\partial Q}{\partial x} \right) c &= -k|u|u - \rho g A \sin \theta + p \frac{\partial A}{\partial x} \end{aligned} \right\} [2.76]$$

Note that the terms $p \frac{\partial A}{\partial x}$ on both sides of the equal sign are canceled out. Simplifying by A/c leads to the following equations:

$$\left. \begin{aligned} \frac{\partial p}{\partial t} - \frac{\rho c}{A} \frac{\partial Q}{\partial t} - \left(\frac{\partial p}{\partial x} - \frac{\rho c}{A} \frac{\partial Q}{\partial x} \right) c &= \frac{c}{A} k|u|u + \rho g c \sin \theta \\ \frac{\partial p}{\partial t} + \frac{\rho c}{A} \frac{\partial Q}{\partial t} + \left(\frac{\partial p}{\partial x} + \frac{\rho c}{A} \frac{\partial Q}{\partial x} \right) c &= -\frac{c}{A} k|u|u - \rho g c \sin \theta \end{aligned} \right\} [2.77]$$

that can be rewritten as:

$$\left. \begin{aligned} \frac{\partial p}{\partial t} - c \frac{\partial p}{\partial x} - \frac{\rho c}{A} \left(\frac{\partial Q}{\partial t} - c \frac{\partial Q}{\partial x} \right) &= \frac{c}{A} k|u|u + \rho g c \sin \theta \\ \frac{\partial p}{\partial t} + c \frac{\partial p}{\partial x} + \frac{\rho c}{A} \left(\frac{\partial Q}{\partial t} + c \frac{\partial Q}{\partial x} \right) &= -\frac{c}{A} k|u|u - \rho g c \sin \theta \end{aligned} \right\} [2.78]$$

The following characteristic equations are obtained:

$$\left. \begin{aligned} \frac{dp}{dt} - \frac{\rho c}{A} \frac{dQ}{dt} &= (k|u|u + \rho g A \sin \theta) \frac{c}{A} & \text{for } \frac{dx}{dt} = -c \\ \frac{dp}{dt} + \frac{\rho c}{A} \frac{dQ}{dt} &= (-k|u|u - \rho g A \sin \theta) \frac{c}{A} & \text{for } \frac{dx}{dt} = c \end{aligned} \right\} [2.79]$$

In the particular case where the cross-sectional area and the wave speed are constant in space, $\partial A / \partial x = 0$ and c can be taken out of the derivative. This remark also holds for A and ρ because of Assumption (A4) that allows the following approximation to be made:

$$d\left(\frac{\rho c}{A} Q\right) = \frac{\rho c}{A} \left(\frac{d\rho}{\rho} + \frac{dc}{c} - \frac{dA}{A} \right) Q + \frac{\rho c}{A} dQ \approx \frac{\rho c}{A} dQ [2.80]$$

Introducing the above simplifications, equation [2.79] can be rewritten as:

$$\left. \begin{aligned} \frac{d}{dt} \left(p - \frac{\rho c}{A} Q \right) &= \frac{kc}{A} |u|u + \rho g c \sin \theta & \text{for } \frac{dx}{dt} = -c \\ \frac{d}{dt} \left(p - \frac{\rho c}{A} Q \right) &= -\frac{kc}{A} |u|u - \rho g c \sin \theta & \text{for } \frac{dx}{dt} = c \end{aligned} \right\} [2.81]$$

Noting that $u = Q/A$, equation [2.81] becomes:

$$\left. \begin{aligned} \frac{d}{dt} (p - \rho c u) &= \frac{kc}{A} |u|u + \rho g c \sin \theta & \text{for } \frac{dx}{dt} = -c \\ \frac{d}{dt} (p + \rho c u) &= -\frac{kc}{A} |u|u - \rho g c \sin \theta & \text{for } \frac{dx}{dt} = c \end{aligned} \right\} [2.82]$$

Consequently, the vector of Riemann invariants and the source term S'' are defined as:

$$W = \begin{bmatrix} p - \rho c u \\ p + \rho c u \end{bmatrix}, \quad S'' = \frac{kc}{A} |u| u \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad [2.83]$$

For the sake of clarity (or to make the derivation of analytical solutions easier), the slope of the pipe and the friction term are sometimes assumed to be zero. In this case the source term S'' becomes zero. Note that equation [2.83] leads to a very simple relationship between W and the state vector (p, u) :

$$\left. \begin{aligned} p &= \frac{1}{2}(W_1 + W_2) \\ u &= \frac{1}{2\rho c}(W_2 - W_1) \end{aligned} \right\} \quad [2.84]$$

NOTE. – The assumption of a constant area and speed of sound is essential in the derivation of the Riemann invariants [2.83]. When A and c are not constant in space (that is, when the properties of the pipe are not homogeneous), an analytical expression must be provided for their variations for equation [2.79] to be integrated.

2.4.4. Calculation of the solution

2.4.4.1. Treatment of internal points

For the sake of simplicity, the pipe is assumed to be horizontal and friction is assumed to be negligible. The properties of the material and the cross-sectional area are assumed to be constant in space. The pipe extends from $x = 0$ to $x = L$. The present section focuses on the calculation of the flow variables p and u at any internal point M of the domain, to the exclusion of the boundaries (Figure 2.9). The issue of boundaries is discussed in the next section. Denoting by A and B the feet of the characteristics $dx/dt = -c$ and $dx/dt = +c$ passing at M respectively, the knowledge of the invariant W_1 at B and that of the invariant W_2 at A allows the solution to be determined uniquely at M. Indeed, using equation [2.82] along the characteristics (AM) and (BM) under the assumption of zero friction and slope ($k = 0$, $\theta = 0$) leads to:

$$\left. \begin{aligned} W_1(M) &= W_1(B) \\ W_2(M) &= W_2(A) \end{aligned} \right\} \quad [2.85]$$

If p and u are known at A and B, equation [2.83] allows $W_1(B)$ and $W_2(A)$ to be determined. $W_1(M)$ and $W_2(M)$ are derived using equation [2.85]. The pressure p_M and the velocity u_M at the point M are determined using relationships [2.84].

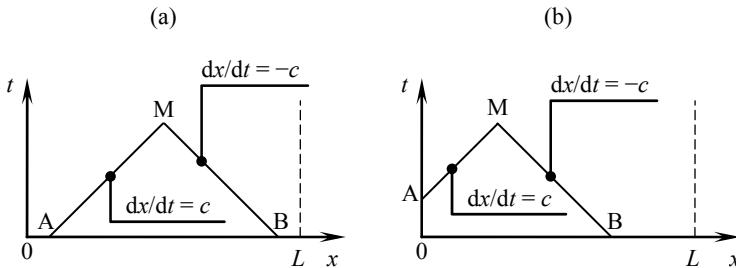


Figure 2.9. Determination of the solution at internal points when the domain of dependence is included entirely in the computational domain (a) and the domain of dependence contains a boundary (b)

$$\left. \begin{aligned} p_M &= \frac{1}{2}[W_1(B) + W_2(A)] = \frac{p_A + p_B}{2} + \rho c \frac{u_A - u_B}{2} \\ u_M &= \frac{1}{2\rho c}[W_2(A) - W_1(B)] = \frac{p_A - p_B}{2\rho c} + \frac{u_A + u_B}{2} \end{aligned} \right\} \quad [2.86]$$

Note that if u_A is larger than u_B , the water accumulates between A and B, which causes an increase in the pressure at the intermediate point M, hence the term $(u_A - u_B)$ in the first equation [2.86]. Conversely, if p_A is larger than p_B , the points located between A and B are subjected to a positive acceleration and u_M increases, hence the term $(p_A - p_B)$ in the second equation [2.86].

The coordinates of points A and B do not need to be identical for relationships [2.86] to be applicable. Two cases may be encountered depending on the value of t_M :

- if the domain of dependence of the point M is included in the segment $[0, L]$ (Figure 2.9a), the knowledge of the initial condition is sufficient for the determination of p_M and u_M ;

- if the domain of dependence of M contains a boundary (Figure 2.9b), only one invariant (here, the invariant W_1) reaches the point M from inside the computational domain. The second invariant (here W_2) travels along the characteristic (AM) issued from the left-hand boundary and requires that a boundary condition be prescribed at $x = 0$ for the solution to be determined uniquely. Boundary conditions are dealt with in section 2.4.4.2.

2.4.4.2. Treatment of boundary conditions

As shown in the previous section, only one Riemann invariant is known at the boundary of the domain because there is only one characteristic coming from the domain. For the solution to be unique at the boundary, additional information should be supplied in the form of a boundary condition.

Consider first the left-hand boundary (Figure 2.10a). The pressure p_A and the flow velocity u_A are determined uniquely at A provided that the invariants W_1 and W_2 are known at A. Then, equation [2.84] can be applied. $W_1(A)$ can be determined from the initial condition at the internal point A' , that is the foot of the characteristic $dx/dt = -c$ that passes at A. The following equality holds:

$$W_1(A) = W_1(A') \quad [2.87]$$

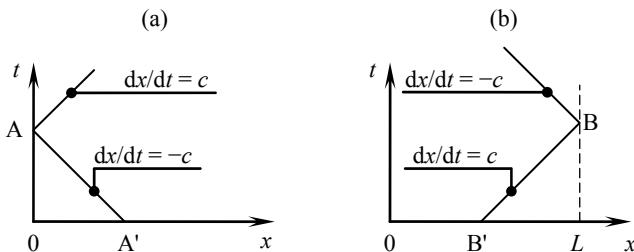


Figure 2.10. Determination of the solution at the boundaries:
left-hand boundary (a); right-hand boundary (b)

However, $W_2(A)$ cannot be computed from the initial condition because the characteristic $dx/dt = +c$ does not reach point A from inside the domain. The missing information must be supplied in the form of an additional boundary condition. The general expression for such a condition is:

$$f(p_A, u_A, t) = 0 \quad [2.88]$$

where the function f is known *a priori*. Equations [2.87–88] form a 2×2 system, the solution of which is unique.

The main three types of boundary condition met in engineering practice are the following:

- prescribed pressure p_b at the left-hand boundary. System [2.87–88] can be rewritten as:

$$\left. \begin{aligned} p_A - \rho c u_A &= p_{A'} - \rho c u_{A'} \\ p_A &= p_b(t) \end{aligned} \right\} \quad [2.89]$$

The pressure p_A is known and the flow velocity u_A is obtained from the first equation [2.89]:

$$u_A = \frac{p_b(t) - p_{A'}}{\rho c} + u_{A'} \quad [2.90]$$

– prescribed velocity u_b at the left-hand boundary (this condition is equivalent to prescribing a discharge because the cross-sectional area A of the pipe is known). System [2.87–88] can be rewritten as:

$$\left. \begin{aligned} p_A - \rho c u_A &= p_{A'} - \rho c u_{A'} \\ u_A &= u_b(t) \end{aligned} \right\} \quad [2.91]$$

which leads to the following expression for p_A :

$$p_A = p_{A'} + [u_b(t) - u_{A'}]\rho c \quad [2.92]$$

– known relationship between the flow velocity u_b (or the discharge Q_b) and the pressure p_b . Such a relationship may express the head loss across singularities such as valves, bends, sudden changes in the cross-sectional area, as well as head gains across pumps, etc. In most cases, the function f is nonlinear and system [2.87–88] must be solved using iterative methods.

Reasoning by symmetry, the following formulations are obtained for a right-hand boundary condition:

– prescribed pressure p_b at the right-hand boundary. u_B is given by:

$$u_B = \frac{p_{B'} - p_b(t)}{\rho c} + u_{B'} \quad [2.93]$$

– prescribed velocity u_b at the right-hand boundary. The pressure p_B is given by:

$$p_B = p_{B'} + [u_{B'} - u_b(t)]\rho c \quad [2.94]$$

2.4.4.3. Bergeron's graphical method

Bergeron's graphical method uses a representation of the characteristic relationships in the (p, u) plane instead of the (x, t) phase space. The principle of the method is outlined for horizontal pipes with a constant cross-sectional area where

the wave speed is constant and friction is negligible. Under such assumptions, equations [2.85] can be rewritten as:

$$\left. \begin{array}{l} p_M - \rho c u_M = p_B - \rho c u_B \\ p_M + \rho c u_M = p_A + \rho c u_A \end{array} \right\} \quad [2.95]$$

These are the equations of straight lines in the (p, u) coordinate system (Figure 2.11). Indeed, system [2.95] can be rewritten as:

$$\left. \begin{array}{l} u_M = \frac{p_M - p_B}{\rho c} + u_B \\ u_M = \frac{p_A - p_M}{\rho c} + u_A \end{array} \right\} \quad [2.96]$$

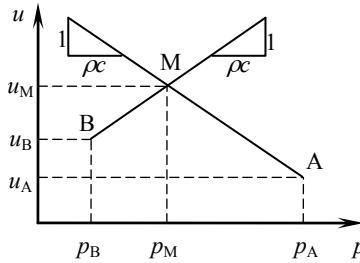


Figure 2.11. Principle of Bergeron's graphical method.

The slopes of the straight lines are $-1/(\rho c)$ and $1/(\rho c)$. The solution (p_M, u_M) , that satisfies equations [2.96], is therefore located at the intersection between the straight lines in the plane (p, u) . The solution is determined graphically by locating the points A and B in the (p, u) plane, drawing the straight lines with slopes $-1/(\rho c)$ and $1/(\rho c)$ that pass at A and B respectively and finding their intersection M (Figure 2.11). Note that the method can be applied to more complex physical situations, in particular when friction is not negligible or when a boundary condition is supplied in the form of a relationship between the pressure and the flow velocity. Application examples can be found in [CAR 72] and [LEN 96].

2.4.5. Summary

The water hammer equations are based on the assumptions of a weakly compressible fluid and a weakly deformable pipe, where inertia is negligible and the

relationship between the pressure and the mass per unit length is linear (see Assumptions (A1–4) in section 2.4.1).

The water hammer equations are written in conservation form [2.2] by defining the variable vector, the flux and the source term as in equation [2.68].

The characteristic form of the water hammer equations is given by equation [2.79]. When the physical properties of the pipe are homogeneous (constant wave speed and cross-sectional area) equation [2.79] can be simplified into equation [2.82]. The differential dW can then be integrated and W and S are given by equation [2.83].

The solution is computed at internal points using equation [2.86]. The existence and uniqueness of the solution requires that a boundary condition should be prescribed at each end of the computational domain. Boundary conditions in the form of prescribed pressure or velocity are given by equations [2.89–94].

2.5. A nonlinear 2×2 system: the Saint Venant equations

2.5.1. Physical context – assumptions

The open channel flow equations, also known as the Saint Venant equations, describe the behavior of flows in open channels such as rivers, canals or pipes where the water flows freely (in other words, the flow is not pressurized). Such equations are one-dimensional because the transverse dimensions of open channels are often very small compared to their longitudinal dimension. The Saint Venant equations are based on the following set of assumptions:

- Assumption (A1). The water is assumed to be incompressible within the usual range of pressure. Its density ρ is constant.
- Assumption (A2). The transverse and vertical acceleration of the water particles can be neglected compared to the longitudinal component of the acceleration. This is equivalent to assuming that the streamlines are only weakly curved. Consequently the pressure field is hydrostatic in a given channel cross-section.
- Assumption (A3). The flow regime is turbulent. The head loss, mainly due to friction against the channel walls, is proportional to the square of the velocity.
- Assumption (A4). The slope of the channel is small enough for the longitudinal coordinate to coincide with the horizontal axis.

If we compare the Saint Venant equations to the water hammer equations studied in section 2.4, several differences arise:

- the water in an open channel is not assumed to be compressible because the range of the pressure variations in an open channel is much smaller than in pressurized pipes;
- the inertial terms cannot be neglected in the momentum equation. They may even become locally predominant and trigger discontinuities in the flow, such as hydraulic jumps, that appear when a rapid flow enters a zone where the flow is slower.

A detailed analysis of the Saint Venant equations, as well as their numerical solution by standard, commercially available numerical models, can be found in [CUN 80].

2.5.2. Conservation form

2.5.2.1. Notation

The following notation is used (Figure 2.12). $A(x)$ is the cross-sectional area of the channel at the abscissa x , $b(x)$ is the width of the free surface, $h(x)$ is the water depth, that is, the vertical distance between the free surface and the lowest point of the section (called the bottom of the channel), $W(x, z)$ is the width of the channel at the abscissa x and the elevation z , $z_b(x)$ is the elevation of the bottom, $\zeta(x)$ is the elevation of the free surface and χ is the wetted perimeter.

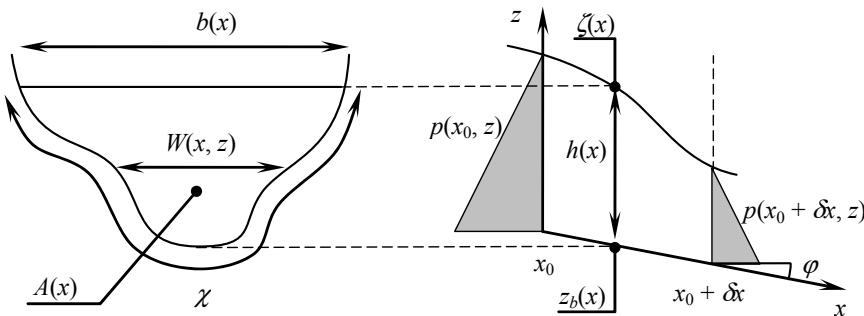


Figure 2.12. Definition sketch for the Saint Venant equations. The pressure forces exerted on the control volume are given by the areas of the gray-shaded triangles

Note that by definition, the following relationship holds between A and b :

$$dA = b \, dz \quad [2.97]$$

This relationship is used in section 2.5.3.3 to derive the expression of the wave speed.

2.5.2.2. Continuity equation

A mass balance is carried out between the times t_0 and $t_0 + \delta t$ over an elementary control volume extending from x_0 to $x_0 + \delta x$:

$$\delta U_1(t_0 + \delta t) - \delta U_1(t_0) = \delta F_1(x_0) - \delta F_1(x_0 + \delta x) \quad [2.98]$$

where $\delta U_1(t)$ is the mass of the fluid contained in the control volume at the time t and $\delta F_1(x)$ is the mass of fluid that passes at the abscissa x during the time interval δt . Equation [2.98] states that the variation of the amount of water contained in the control volume is equal to the difference between the amount that flows in and the amount that flows out. By definition, δU_1 and δF_1 are given by:

$$\left. \begin{aligned} \delta U_1(t) &= \int_{x_0}^{x_0 + \delta x} (\rho A)(x, t) dx \\ \delta F_1(x) &= \int_{t_0}^{t_0 + \delta t} (\rho u A)(x, t) dx \end{aligned} \right\} \quad [2.99]$$

where A is the cross-sectional area, u is the fluid velocity and ρ is the density of the fluid. Substituting equation [2.99] into equation [2.98] in the limit of small δt and δx leads to the following equation (see equations [1.12–16] for a detailed proof):

$$\frac{\partial}{\partial t}(\rho A) + \frac{\partial}{\partial x}(\rho u A) = 0 \quad [2.100]$$

Using Assumption (A2) of an incompressible fluid allows equation [2.100] to be divided by the (constant) density ρ . Noting that $Q = Au$, equation [2.100] is simplified into:

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0 \quad [2.101]$$

Defining the first component of the vector variable as A , equation [2.101] can be written in conservation form as:

$$\left. \begin{aligned} \frac{\partial U_1}{\partial t} + \frac{\partial F_1}{\partial x} &= 0 \\ U_1 &= A \\ F_1 &= Q \end{aligned} \right\} \quad [2.102]$$

2.5.2.3. Momentum equation

The momentum balance is obtained by applying the fundamental principle of dynamics to a control volume of length δx over a time interval δt :

$$\begin{aligned} \delta U_2(t_0 + \delta t) - \delta U_2(t_0) &= \delta F_2(x_0) - \delta F_2(x_0 + \delta x) \\ &\quad + \delta P(x_0) - \delta P(x_0 + \delta x) + \delta F_W \end{aligned} \quad [2.103]$$

where $\delta P(x)$ is the integral between t and $t + \delta t$ of the pressure force exerted on the cross-section at the abscissa x , $\delta U_2(t)$ is the momentum of the fluid contained in the slice of channel of length δx , and δF_W is the x -component of the reaction exerted by the channel walls on the control volume. By definition, δU_2 and δF_2 are given by equation [2.55], recalled here:

$$\left. \begin{aligned} \delta U_2(t) &= \int_{x_0}^{x_0 + \delta x} (\rho u A)(x, t) dx \\ \delta F_2(x) &= \int_{t_0}^{t_0 + \delta t} (\rho u^2 A)(x, t) dx \end{aligned} \right\}$$

The pressure force $P(x)$ is equal to the integral of the pressure over the cross-section. From Assumption (A2), the pressure field is hydrostatic, that is, it is proportional to the distance between the point under consideration and the free surface. The pressure $p(x, z)$ at the abscissa x and at the elevation z is therefore:

$$p(x, z) = (\zeta - z)\rho g \quad [2.104]$$

where both z and ζ depend on x (this is omitted in the notation for the sake of clarity). The pressure $p(x, z)$ is exerted on an elementary slice of height δz and width $W(x, z)$. The pressure force $P(x)$ is therefore given by:

$$P(x) = \int_{z_b(x)}^{\zeta(x)} p(x, z) W(x, z) dz = \int_{z_b(x)}^{\zeta(x)} [\zeta(x) - z] \rho g W(x, z) dz \quad [2.105]$$

The force δF_W exerted by the walls of the channel on the control volume is expressed as the sum of three forces:

$$\delta F_P(t) = \int_{x_0}^{x_0 + \delta x} (R_p + R_f + R_b)(x, t) dx \quad [2.106]$$

where R_b , R_f and R_w are respectively the reaction of the bottom on the control volume in the vertical plane, the friction force and the reaction of the walls on the control volume in the horizontal plane. The friction force R_f is usually expressed in terms of the “energy slope”, or “energy grade line” S_f . S_f is positive when energy is lost in the direction of positive x . The following equivalence holds between R_f and S_f :

$$R_f = -\rho g h S_f \quad [2.107]$$

Note that R_f is exerted in a direction that is parallel to the channel bottom and not along the horizontal. In addition, owing to the non-zero bottom slope, R_f is exerted over a length that is slightly larger than δx . However, Assumption (A4) allows R_f to be approximated as its projection on the x -axis, while the length over which the force is exerted is approximated with δx . Several formulae are available for S_f . All of them use Assumption (A3) of a turbulent flow regime, hence the assumption that the slope of the energy line is proportional to the square of the flow velocity u . The most frequently used laws are:

$$\begin{aligned} S_f &= \frac{u^2}{C^2 R_H} && \text{(Chezy)} \\ S_f &= \frac{u^2}{K_{\text{Str}}^2 R_H^{4/3}} && \text{(Strickler)} \\ S_f &= n_M^2 \frac{u^2}{R_H^{4/3}} && \text{(Manning)} \end{aligned} \quad [2.108]$$

where C , K_{Str} and n_M are respectively the Chezy, Strickler and Manning friction coefficients. R_H is the hydraulic radius, defined as the ratio of the cross-sectional area to the wetted perimeter:

$$R_H = \frac{A}{\chi} \quad [2.109]$$

The Chezy coefficient is often used by coastal engineers, while river engineers usually prefer the Strickler and Manning coefficients. A large value of the Chezy or Strickler coefficients, or a small value of the Manning coefficient, indicate that friction is small. The three coefficients can be related to each other as follows:

$$K_{\text{Str}} = \frac{1}{n_M} = \frac{C}{R_H^{1/6}} \quad [2.110]$$

As in the water hammer equation, the force R_p is given by the projection of the reaction of the walls of the channel onto the x -axis. In what follows, the reaction of the walls is understood as the horizontal component of the reaction of the walls. The reaction of the walls also has a vertical component that is usually referred to as the reaction R_b of the bottom of the channel. The force R_p is considered first. Consider an elementary volume of channel of height dz , the distance of which to the free surface is denoted by d . The domain extends from one channel wall to the other between the abscissas $x - dx/2$ and $x + dx/2$ (Figure 2.13). From Assumption (A2), the (hydrostatic) pressure p is given by equation [2.104]. The x -component of the reaction of the wall onto the elementary volume is therefore:

$$\begin{aligned} dR_p &= [W(x + dx/2) - W(x - dx/2)] p(x, z) dz \\ &\approx \left(\frac{\partial W}{\partial x} \right)_{\zeta-z=\text{Const}} (\zeta - z) \rho g dx dz \end{aligned} \quad [2.111]$$

where the notation $(\partial W / \partial x)_{\zeta-z=\text{Const}}$ indicates that the partial derivative of the width with respect to z is estimated keeping a constant distance below the free surface. The force R_p is obtained by integrating dR_p from z_b to ζ and from x_0 to $x_0 + \delta x$:

$$R_p = \rho g \int_{x_0}^{x_0 + \delta x} \int_{z_b(x)}^{\zeta(x)} (\zeta - z) \left(\frac{\partial W}{\partial x} \right)_{\zeta-z=\text{Const}} dz dx \quad [2.112]$$

The projection R_b of the reaction of the bottom onto the x -axis is determined by carrying out a balance between the forces exerted onto a slice of channel delineated by $x - dx/2$ and $x + dx/2$. The reaction of the bottom onto the control volume is exerted in a vertical plane, in the direction of the normal unit vector to the bottom of the channel (see Figure 2.13, top). The projection of the reaction of the channel onto the z -axis is equal to the opposite of the weight $\rho g A dx$ of the control volume. Since the bottom of the channel makes an angle φ with the horizontal, the horizontal component of the reaction of the bottom of the channel onto the fluid is:

$$R_b = \rho g A(x) \operatorname{tg} \varphi dx \quad [2.113]$$

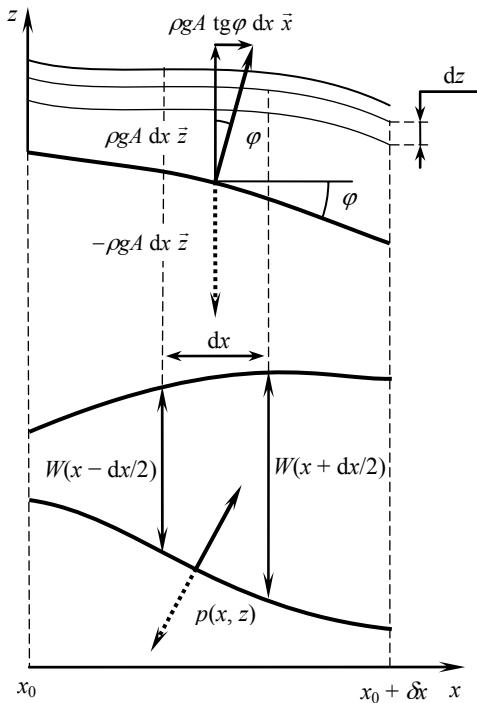


Figure 2.13. Force exerted by the fluid onto the wall (dashed arrows), reaction of the wall onto the fluid (solid arrows). Vertical projection (top), horizontal projection (bottom)

Note that the tangent of the angle φ is the slope S_0 of the channel:

$$S_0 = \tan \varphi = -\frac{\partial z_b}{\partial x} \quad [2.114]$$

Substituting equations [2.55, 2.105, 2.107, 2.112–114] into equation [2.103] in the limit of zero δt and δx , dividing by the constant density ρ leads to:

$$\frac{\partial}{\partial t}(uA) + \frac{\partial}{\partial x}(u^2 A + P / \rho) = (S_0 - S_f)gA + I_p \quad [2.115]$$

where the integral I_p is given by:

$$I_p = \frac{1}{\rho} \frac{\partial R_p}{\partial x} = g \int_{z_b(x)}^{\zeta(x)} (\zeta - z) \left(\frac{\partial W}{\partial x} \right)_{\zeta - z = \text{Const}} dz \quad [2.116]$$

Equation [2.115] can be written in conservation form as:

$$\left. \begin{aligned} \frac{\partial U_2}{\partial t} + \frac{\partial F_2}{\partial x} &= S_2 \\ U_2 &= Q \\ F_2 &= M = Q^2 / A + P / \rho \\ S_2 &= (S_0 - S_f)gA + I_p \end{aligned} \right\} \quad [2.117]$$

2.5.2.4. Vector form

The system formed by equations [2.102] and [2.117] can be written in the vector conservation form [2.2], recalled here:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} = \mathbf{S}$$

by defining \mathbf{U} , \mathbf{F} and \mathbf{S} as:

$$\mathbf{U} = \begin{bmatrix} A \\ Q \end{bmatrix}, \mathbf{F} = \begin{bmatrix} Q \\ M \end{bmatrix} = \begin{bmatrix} Q \\ \frac{Q^2}{A} + \frac{P}{\rho} \end{bmatrix}, \mathbf{S} = \begin{bmatrix} 0 \\ (S_0 - S_f)gA + I_p \end{bmatrix} \quad [2.118]$$

where $M = Q^2/A + P/\rho$ is the specific force and Q^2/A is the momentum flux.

2.5.3. Characteristic form – Riemann invariants

2.5.3.1. Non-conservation form

Equation [2.2] is first rewritten in the non-conservation form [2.5], recalled hereafter:

$$\frac{\partial \mathbf{U}}{\partial t} + \mathbf{A} \frac{\partial \mathbf{U}}{\partial x} = \mathbf{S}'$$

where \mathbf{A} and \mathbf{S}' are given as in equation [2.6], recalled here:

$$\left. \begin{aligned} d\mathbf{F} &= \mathbf{A} d\mathbf{U} \\ \mathbf{S}' &= \mathbf{S} - \left(\frac{\partial \mathbf{F}}{\partial x} \right)_{U=\text{Const}} \end{aligned} \right\}$$

Note from equation [2.118] that A is given by:

$$A = \begin{bmatrix} 0 & 1 \\ c^2 - u^2 & 2u \end{bmatrix} \quad [2.119]$$

where c is defined as:

$$c = \left[\frac{\partial(P/\rho)}{\partial A} \right]^{1/2} \quad [2.120]$$

The expression for c is given in section 2.5.3.3. The source term S' is derived by noting that $(\partial F/\partial U)_{U=\text{Const}}$ is to be computed for a fixed value of U , that is, for constant A and Q . Consequently:

- the first component of S' is zero because both A and Q are assumed constant;
- the term Q^2/A in the second component M of F is also assumed constant. The only term in M that changes with x for a constant U is the pressure force P . Variations in P arise either from variations in the water depth h for a given channel shape, or from variations in the shape of the channel cross-section for a given h . Differentiating P/ρ with respect to x gives:

$$\frac{\partial}{\partial x} \left(\frac{P}{\rho} \right) = \frac{1}{\rho} \frac{\partial}{\partial x} \int_{z_b}^{\zeta} p(z) W(z) dz = \frac{1}{\rho} \frac{\partial}{\partial x} \int_0^h p(z') W(z') dz' \quad [2.121]$$

where $z' = z - z_b$ is the elevation above the bed level. Using Leibniz' differentiation rule for the integral, noting that $p = 0$ for $z' = h$ and substituting equation [2.104] into equation [2.121] leads to:

$$\begin{aligned} \frac{\partial}{\partial x} \left(\frac{P}{\rho} \right) &= \frac{1}{\rho} \int_0^h \frac{\partial}{\partial x} [p(z') W(z')] dz' = g \int_0^h \frac{\partial}{\partial x} [(h - z') W(z')] dz' \\ &= g \int_0^h \frac{\partial h}{\partial x} W(z') dz' + g \int_0^h \frac{\partial W(z')}{\partial x} (h - z') dz' \end{aligned} \quad [2.122]$$

Note that the second term in equation [2.122] is nothing but the integral I_p as defined in equation [2.116]. Moreover, it is noted that in the case of a prismatic channel, $I_p = 0$. Consequently, the first integral in equation [2.122] is part of the

second component of the vector term $(\partial F / \partial U)_{x=\text{Const}} \partial U / \partial x$ in equation [2.6], while I_p represents the second component of the term $(\partial F / \partial x)_{U=\text{Const}}$:

$$\left. \begin{aligned} \left[\frac{\partial}{\partial x} \left(\frac{P}{\rho} \right) \right]_{x=\text{Const}} &= g \int_0^h \frac{\partial h}{\partial x} W(z') dz' \\ \left(\frac{\partial M}{\partial x} \right)_{\substack{A=\text{Const} \\ Q=\text{Const}}} &= \left[\frac{\partial}{\partial x} \left(\frac{P}{\rho} \right) \right]_{A=\text{Const}} = g \int_0^h \frac{\partial W(z')}{\partial x} (h - z') dz' = I_p \end{aligned} \right\} \quad [2.123]$$

Consequently, we have:

$$S' = S - \left(\frac{\partial F}{\partial x} \right)_{U=\text{Const}} = \begin{bmatrix} 0 \\ (S_0 - S_f)gA \end{bmatrix} \quad [2.124]$$

2.5.3.2. Characteristic form

The eigenvalues and eigenvectors of A are:

$$\begin{aligned} \lambda^{(1)} &= u - c, & \lambda^{(2)} &= u + c \\ K^{(1)} &= \begin{bmatrix} 1 \\ u - c \end{bmatrix}, & K^{(2)} &= \begin{bmatrix} 1 \\ u + c \end{bmatrix} \end{aligned} \quad [2.125]$$

Since the Jacobian matrix A has two real and distinct eigenvalues, system [2.2, 2.118] is hyperbolic. The variations of the pressure force are related to those of the mass per unit length; therefore the system describes a compressible behavior.

Recall that the characteristic form is obtained by multiplying the non-conservation form by K^{-1} so as to involve the diagonal matrix Λ formed by the eigenvalues of A:

$$\frac{\partial W}{\partial t} + \Lambda \frac{\partial W}{\partial x} = S''$$

The differential of the vector Riemann invariant dW is defined by the first relationship [2.21]:

$$dW = K^{-1} dU$$

Matrices K and K^{-1} are derived from equation [2.125]:

$$K = \begin{bmatrix} 1 & 1 \\ u-c & u+c \end{bmatrix}, \quad K^{-1} = \frac{1}{2c} \begin{bmatrix} u+c & -1 \\ c-u & 1 \end{bmatrix} \quad [2.126]$$

Substituting equations [2.119] and [2.126] into equation [2.21] with the definition [2.124] leads to:

$$\left. \begin{aligned} dW &= \frac{1}{2c} \begin{bmatrix} (u+c)dU_1 - dU_2 \\ (c-u)dU_1 + dU_2 \end{bmatrix} = \frac{1}{2c} \begin{bmatrix} (u+c)dA - d(Au) \\ (c-u)dA + d(Au) \end{bmatrix} \\ S'' &= \frac{1}{2c} \begin{bmatrix} -(S_0 - S_f)gA \\ (S_0 - S_f)gA \end{bmatrix} \end{aligned} \right\} \quad [2.127]$$

The differential dW can be integrated only if the expression for c is known. This is the subject of section 2.5.3.2. The formula for c is then used to derive that of the Riemann invariants in a number of particular cases examined in section 2.5.3.3.

2.5.3.3. Expression of the speed of the waves in still water

Definition [2.120] is recalled for c :

$$c = \left[\frac{\partial(P/\rho)}{\partial A} \right]^{1/2}$$

where P is given by equation [2.105]. Dividing equation [2.105] by ρ leads to:

$$\frac{P}{\rho}(x) = \int_{z_b(x)}^{\zeta(x)} (\zeta - z) g W(x, z) dz \quad [2.128]$$

The pressure force P is first shown to be related to the cross-sectional area by a one-to-one relationship. Both the cross-sectional area A and the pressure force P are strictly increasing functions of the elevation ζ of the free surface, provided that b is non-zero. Consequently there is a one-to-one relationship between P and ζ , and there is a one-to-one relationship between A and ζ . Therefore, there exists a one-to-one relationship between A and P and the derivative $\partial P / \partial A$ can be computed as:

$$\frac{\partial(P/\rho)}{\partial A} = \frac{\partial(P/\rho)}{\partial \zeta} \frac{\partial \zeta}{\partial A} = \frac{\partial(P/\rho)}{\partial \zeta} \left(\frac{\partial A}{\partial \zeta} \right)^{-1} \quad [2.129]$$

The expression of $\partial A / \partial \zeta$ is derived assuming that the elevation of the free surface is subjected to an infinitesimal variation $d\zeta$. The corresponding variation dA in the cross-sectional area is:

$$dA = b d\zeta \quad [2.130]$$

Consequently:

$$\frac{\partial A}{\partial \zeta} = b \quad [2.131]$$

The derivative $\partial P / \partial \zeta$ is derived as follows. Consider an infinitesimal variation $d\zeta$ in the elevation of the free surface, that moves from the point A to the point A' (Figure 2.14).

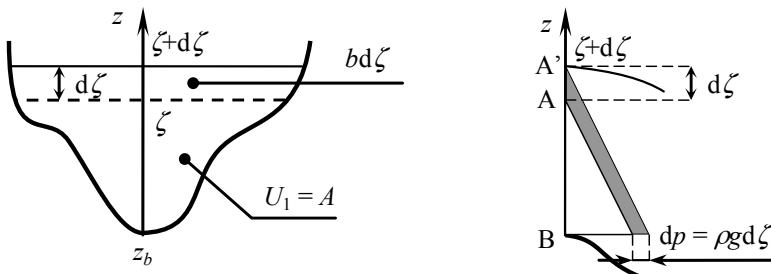


Figure 2.14. Variation in the pressure force caused by a variation $d\zeta$ in the water level. The variation dP_1 in the pressure force triggered by the variation $d\zeta$ is the area of the gray-shaded areas between the triangles that illustrate the pressure field

Since the pressure is hydrostatic, it increases uniformly by a quantity $\rho g dz$ over the entire section. The resulting increase in the pressure field is therefore:

$$dP_1 = Adp = A\rho g d\zeta \quad [2.132]$$

The additional force brought by the infinitesimal area between the points A and A' in the side view (Figure 2.14, right) is given by:

$$dP_2 = b\rho g \frac{(d\zeta)^2}{2} \quad [2.133]$$

The total variation in the pressure force is given by $dP_1 + dP_2$:

$$dP = \left(A + \frac{b d\zeta}{2} \right) \rho g d\zeta \quad [2.134]$$

In the limit of small $d\zeta$, dP becomes equivalent to $\rho g A d\zeta$ and:

$$\frac{\partial(P/\rho)}{\partial\zeta} = \rho g A \quad [2.135]$$

Substituting equations [2.131] and [2.135] into equation [2.129] leads to:

$$\frac{\partial(P/\rho)}{\partial A} = \frac{gA}{b} \quad [2.136]$$

Hence the expression of c :

$$c = \left(\frac{gA}{b} \right)^{1/2} \quad [2.137]$$

In the particular case of a rectangular channel, equation [2.137] simplifies into:

$$c = (gh)^{1/2} \quad [2.138]$$

This expression is to be connected to a very remarkable behavior. Free-surface waves propagate more rapidly in deep water than in shallow water. This explains in particular the steepening and the breaking of sea waves traveling to the shore. In a more dramatic fashion, equation [2.138] also accounts for the development of tsunamis. The behavior of the solutions of hyperbolic conservation laws in the presence of steep fronts and discontinuities is covered in detail in Chapter 3.

2.5.3.4. Riemann invariants

The purpose is to derive a simplified formulation for the Riemann invariants as given by equations [2.127]. Substituting equations [2.135] into equation [2.22] and multiplying by $2c$ leads to the following system:

$$\begin{aligned} c \frac{dA}{dt} - A \frac{du}{dt} &= -(S_0 - S_f) g A & \text{for } \frac{dx}{dt} = u - c \\ c \frac{dA}{dt} + A \frac{du}{dt} &= (S_0 - S_f) g A & \text{for } \frac{dx}{dt} = u + c \end{aligned} \quad [2.139]$$

Dividing by A leads to:

$$\left. \begin{aligned} \frac{du}{dt} - \frac{c}{A} \frac{dA}{dt} &= (S_0 - S_f)g & \text{for } \frac{dx}{dt} = u - c \\ \frac{du}{dt} + \frac{c}{A} \frac{dA}{dt} &= (S_0 - S_f)g & \text{for } \frac{dx}{dt} = u + c \end{aligned} \right\} \quad [2.140]$$

The above equations can be integrated easily in a number of particular cases. The following three configurations are examined for prismatic channels ($h_{x,A} = 0$):

– *Rectangular channel.* b is constant, A and c are given by:

$$\left. \begin{aligned} c &= (gh)^{1/2} \\ A = bh &= \frac{bc^2}{g} \end{aligned} \right\} \quad [2.141]$$

The following relationship holds:

$$dA = bdh = \frac{2c}{g} dc \quad [2.142]$$

Substituting equation [2.142] into equation [2.140] and noting that $c^2 = gh$, the following system is obtained:

$$\left. \begin{aligned} \frac{d}{dt}(u - 2c) &= (S_0 - S_f)g & \text{for } \frac{dx}{dt} = u - c \\ \frac{d}{dt}(u + 2c) &= (S_0 - S_f)g & \text{for } \frac{dx}{dt} = u + c \end{aligned} \right\} \quad [2.143]$$

Note that the source term on the right-hand side of the equations is canceled out in horizontal channels with negligible friction or when the flow is uniform ($S_0 = S_f$ by definition). The quantity $u - 2c$ is then constant along the characteristics, the speed of which is $u - c$. The quantity $u + 2c$ is constant along the characteristics of speed $u + c$. The Riemann invariants and the source term in equation [2.22] can be redefined as:

$$W = \begin{bmatrix} u - 2c \\ u + 2c \end{bmatrix}, \quad S'' = \begin{bmatrix} (S_0 - S_f)g \\ (S_0 - S_f)g \end{bmatrix} \quad [2.144]$$

– *Triangular channel.* The width b is proportional to the water depth h , A is proportional to the square of b and h :

$$\left. \begin{array}{l} b = 2h \operatorname{tg} \theta \\ A = h^2 \operatorname{tg} \theta \\ c = (gh/2)^{1/2} \end{array} \right\} [2.145]$$

where θ is the angle of the embankments with the vertical (the channel is assumed to be symmetric). Equations [2.140] can be rewritten as:

$$\left. \begin{array}{l} \frac{d}{dt}(u - 4c) = (S_0 - S_f)g \quad \text{for } \frac{dx}{dt} = u - c \\ \frac{d}{dt}(u + 4c) = (S_0 - S_f)g \quad \text{for } \frac{dx}{dt} = u + c \end{array} \right\} [2.146]$$

Under uniform conditions, or when the channel is horizontal and friction is negligible, the quantity $u - 4c$ is invariant along the characteristic of speed $u - c$. The quantity $u + 4c$ is invariant along the characteristic of speed $u + c$. The Riemann invariants and the source term in equation [2.22] may be redefined as:

$$W = \begin{bmatrix} u - 4c \\ u + 4c \end{bmatrix}, \quad S'' = \begin{bmatrix} (S_0 - S_f)g \\ (S_0 - S_f)g \end{bmatrix} [2.147]$$

– Approximation for arbitrary-shaped channels. The following relationships hold:

$$\left. \begin{array}{l} A = \frac{bc^2}{g} \\ dA = \frac{c^2}{g} db + \frac{2b}{g} c dc = \left(\frac{c^2}{g} \frac{db}{dc} + \frac{2b}{g} c \right) dc \end{array} \right\} [2.148]$$

Substituting equation [2.148] into equation [2.140] leads to the following system:

$$\left. \begin{array}{l} \frac{du}{dt} - \left(2 + \frac{c}{b} \frac{db}{dc} \right) \frac{dc}{dt} = (S_0 - S_f)g \quad \text{for } \frac{dx}{dt} = u - c \\ \frac{du}{dt} + \left(2 + \frac{c}{b} \frac{db}{dc} \right) \frac{dc}{dt} = (S_0 - S_f)g \quad \text{for } \frac{dx}{dt} = u + c \end{array} \right\} [2.149]$$

If the quantity db/dc can be assumed to be proportional to b/c , then:

$$\frac{c}{b} \frac{db}{dc} = \text{Const} \quad [2.150]$$

and the term $(c/b db/dc)$ can be passed into the operator d/dt . Equations [2.149] then become:

$$\left. \begin{aligned} \frac{d}{dt} \left[u - \left(2 + \frac{c}{b} \frac{db}{dc} \right) c \right] &= (S_0 - S_f) g && \text{for } \frac{dx}{dt} = u - c \\ \frac{d}{dt} \left[u + \left(2 + \frac{c}{b} \frac{db}{dc} \right) c \right] &= (S_0 - S_f) g && \text{for } \frac{dx}{dt} = u + c \end{aligned} \right\} \quad [2.151]$$

If $(c/b db/dc)$ is not strictly constant, equation [2.151] is only an approximation. This approximation may however prove useful in a number of cases. Note that the rectangular channel ($db/dc = 0$) and the triangular channel ($db/dc = 2$) are particular cases of equations [2.151].

The quantity $(c/b db/dc)$ is difficult to estimate in practical applications. It is more conveniently related to the geometry of the channel via the general definition of the propagation speed of the waves in still water:

$$c^2 = \frac{gA}{b} \quad [2.152]$$

Differentiating equation [2.152] leads to:

$$2c \frac{dc}{db} = \left(\frac{dA}{b} - \frac{A}{b^2} db \right) g = \left(dh - \frac{A}{b^2} db \right) g = \left(\frac{dh}{db} - \frac{A}{b^2} \right) g \frac{db}{b} \quad [2.153]$$

Hence the expression of dc/db :

$$\frac{dc}{db} = \left(\frac{dh}{db} - \frac{A}{b^2} \right) \frac{g}{2c} \quad [2.154]$$

and:

$$\frac{b}{c} \frac{dc}{db} = \left(\frac{dh}{db} - \frac{A}{b^2} \right) \frac{gb}{2c^2} = \left(\frac{dh}{db} - \frac{A}{b^2} \right) \frac{gb^2}{2gA} = \frac{1}{2} \left(\frac{b^2}{A} \frac{dh}{db} - 1 \right) \quad [2.155]$$

The quantity $c/b \, db/dc$ is obtained directly from equation [2.155]. Note that dh/db can be estimated very easily from the geometry of the channel. The general formula [2.149] becomes:

$$\left. \begin{array}{l} \frac{du}{dt} - \beta \frac{dc}{dt} = (S_0 - S_f)g \\ \frac{du}{dt} + \beta \frac{dc}{dt} = (S_0 - S_f)g \end{array} \right\} \begin{array}{l} \text{for } \frac{dx}{dt} = u - c \\ \text{for } \frac{dx}{dt} = u + c \end{array} \quad [2.156]$$

where β is defined as:

$$\beta = 2 + 2 \frac{b^2}{A} \left(\frac{dh}{db} - 1 \right)^{-1} \quad [2.157]$$

If dh/db can be assumed to be proportional to A/b^2 , equation [2.151] becomes:

$$\left. \begin{array}{l} \frac{d}{dt}(u - \beta c) = (S_0 - S_f)g \quad \text{for } \frac{dx}{dt} = u - c \\ \frac{d}{dt}(u + \beta c) = (S_0 - S_f)g \quad \text{for } \frac{dx}{dt} = u + c \end{array} \right\} \quad [2.158]$$

The Riemann invariants and the source term in equation [2.22] can be redefined as:

$$W = \begin{bmatrix} u - \beta c \\ u + \beta c \end{bmatrix}, \quad S'' = \begin{bmatrix} (S_0 - S_f)g \\ (S_0 - S_f)g \end{bmatrix} \quad [2.159]$$

NOTE.— Equation [2.158] leads to equation [2.147]. Equation [2.158] does not lead to equation [2.145] for a rectangular channel because the assumption $db \neq 0$ is needed in equation [2.154] to proceed from equation [2.149] to equation [2.158]. This assumption is wrong in the case of a rectangular channel because b is constant.

2.5.4. Calculation of solutions

2.5.4.1. The various possible flow regimes

As seen in section 2.5.3.1, the propagation speeds of the waves are $u - c$ and $u + c$. In contrast with the water hammer phenomenon, the speeds of the waves are not independent of the flow variables. They can be seen as the result of the superimposition of two phenomena:

– The propagation of the pressure waves is accounted for the terms $-c$ and $+c$. As in the water hammer phenomenon, a perturbation in the flow gives rise to two pressure waves traveling in opposite directions.

– The pressure waves travel in the fluid (here, water) that moves at the speed u . Therefore, the flow velocity u must be added to the speed of the pressure waves to account for the movement of the water molecules in the channel.

In a Lagrangian coordinate system, that is, a coordinate system that moves at the speed of the flow, the speeds of the waves become $-c$ and $+c$ respectively.

The so-called Froude number is commonly used to characterize the flow regime. This dimensionless number is defined as:

$$\text{Fr} = \frac{|u|}{c} \quad [2.160]$$

Depending on the value of Fr, the flow regime is said to be subcritical, critical or supercritical:

– For $\text{Fr} < 1$, the flow is said to be subcritical. This corresponds to the condition $|u| < c$. In this case, $u - c$ is negative and $u + c$ is positive. The first wave travels upstream, the second wave travels downstream. As a consequence, the behavior of the flow at a given point in the channel is influenced by the flow conditions downstream of it.

– For $\text{Fr} > 1$, the flow is said to be supercritical. This corresponds to the condition $|u| > c$. If u is positive, both wave speeds are positive. If u is negative, both wave speeds are negative. Both waves propagate downstream and the local behavior of the flow is influenced only by the flow conditions upstream.

– For $\text{Fr} = 1$, the flow is said to be critical. This situation corresponds to the transition between subcritical and supercritical conditions. It is usually restricted to a very small region of the flow. A critical point indicates the limit point in the channel where the flow conditions cease to be influenced by the conditions downstream. Critical conditions are usually encountered at singularities such as sills, weirs or next to bridge piers.

The various possible flow regimes can be illustrated in the phase space as in Figure 2.15. If the flow is subcritical, the slopes of the characteristics have opposite signs (Figure 2.15a). When the flow is critical, one of the characteristics is parallel to the time axis (Figure 2.15b). When the flow is supercritical, the slopes of the two characteristics have the same sign (Figure 2.15c).

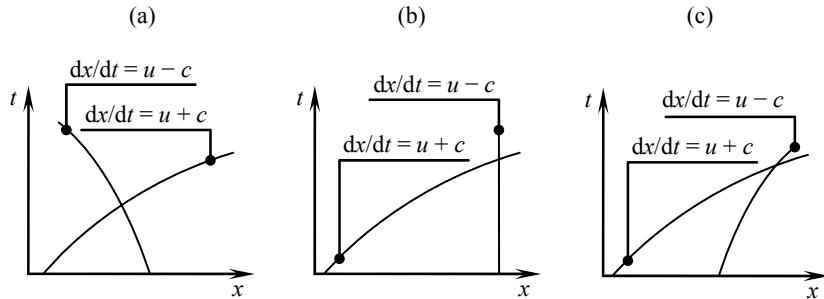


Figure 2.15. Definition sketch for the various possible flow regimes in the phase space: subcritical (a), critical (b), supercritical (c)

2.5.4.2. Treatment of internal points

Assume first that $S'' = 0$ and that the Riemann invariants can be integrated in the general form [2.159] with the coefficient β defined as in equation [2.157]. The channel reach over which the solution (that is, the couple (A, Q)) is to be determined extends from $x = 0$ to $x = L$ (Figure 2.16). This section focuses on the calculation of the solution at internal points.

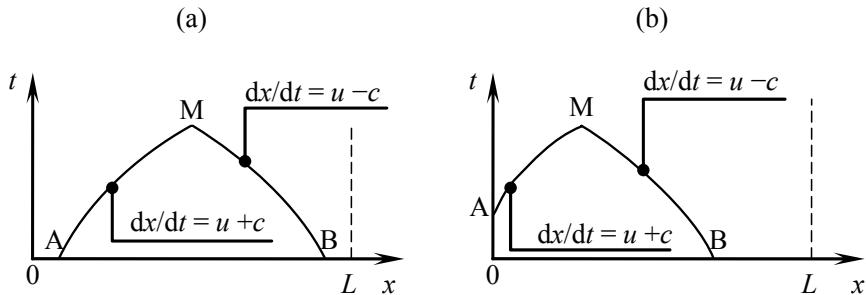


Figure 2.16. Calculation of the solution at internal points when the domain of dependence is included in the computational domain (a) and when it includes a boundary (b)

Let A and B denote the feet of the characteristics $dx/dt = u + c$ and $dx/dt = u - c$ respectively passing at the point M where the solution is sought. A and B may or may not be located at a boundary. A and Q are assumed to be known at both A and B.

As mentioned in section 2.2.2, the solution can be determined uniquely at M provided that the invariant W_1 at B and the invariant W_2 at A are known. Applying relationships [2.158] under the assumption of a zero source term gives:

$$\left. \begin{array}{l} \frac{d}{dt}(u - \beta c) = 0 \text{ for } \frac{dx}{dt} = u - c \\ \frac{d}{dt}(u + \beta c) = 0 \text{ for } \frac{dx}{dt} = u + c \end{array} \right\} \quad [2.161]$$

that is:

$$\left. \begin{array}{l} u_M - \beta c_M = u_B - \beta c_B \\ u_M + \beta c_M = u_A + \beta c_A \end{array} \right\} \quad [2.162]$$

Solving equations [2.162] for u_M and c_M leads to:

$$\left. \begin{array}{l} u_M = \frac{u_A + u_B}{2} - \frac{c_A - c_B}{2} \beta \\ c_M = \frac{u_A - u_B}{2\beta} + \frac{c_A + c_B}{2} \end{array} \right\} \quad [2.163]$$

The speed c of the waves in still water is in general related to the cross-sectional area A by a one-to-one relationship. The knowledge of c_M allows A_M to be determined uniquely.

When the source term S'' is non-zero, equations [2.158] are integrated into:

$$\left. \begin{array}{l} u_M - \beta c_M = u_B - \beta c_B + (t_M - t_B)S''_{BM} \\ u_M + \beta c_M = u_A + \beta c_A + (t_M - t_A)S''_{AM} \end{array} \right\} \quad [2.164]$$

where t_A , t_B and t_M are the time coordinates of the points A, B and M in the phase space respectively and S''_{AM} and S''_{BM} are the average values of the source term S'' along the characteristics [AM] and [BM] respectively:

$$\left. \begin{array}{l} S''_{AM} = \frac{1}{t_M - t_A} \int_A^M [(S_0 - S_f)g] dt \\ S''_{BM} = \frac{1}{t_M - t_B} \int_B^M [(S_0 - S_f)g] dt \end{array} \right\} \quad [2.165]$$

The solution of system [2.165] is unique:

$$\left. \begin{aligned} u_M &= \frac{u_A + u_B}{2} + \beta \frac{c_A - c_B}{2} + \frac{S''_{AM} + (t_M - t_B)S''_{BM}}{2} \\ c_M &= \frac{u_A - u_B}{2\beta} + \frac{c_A + c_B}{2} + \frac{(t_M - t_A)S''_{AM} - (t_M - t_B)S''_{BM}}{2\beta} \end{aligned} \right\} [2.166]$$

Despite its apparent simplicity, the calculation of the solution as in equation [2.166] is not straightforward. Indeed, it requires that the average values of S'' be estimated along the characteristics [AM] and [BM]. From a theoretical point of view, this requires that u and c be known exactly at all points along these characteristics. In arbitrarily-shaped channels, u and c vary along the characteristics. Moreover, the variations in the bed slope S_0 and the friction slope S_f cannot be described analytically. In practical applications, the source term S'' must be approximated.

2.5.4.3. Treatment of boundary conditions

The treatment of boundary points depends on the flow regime. Three typical situations are examined (Figure 2.17): (1) the flow is subcritical at the boundary, (2) the flow is supercritical, entering the domain, (3) the flow is supercritical, leaving the domain. These cases are detailed hereafter for the treatment of the left-hand boundary of the domain:

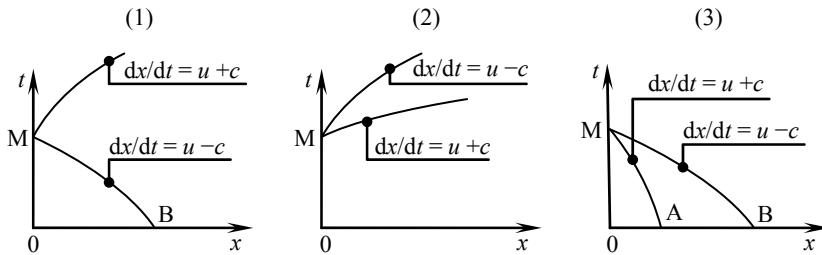


Figure 2.17. Calculation of the solution at the left-hand boundary: subcritical flow (1), inflowing supercritical flow (2), outflowing supercritical flow (3)

(1) Subcritical flow. The characteristic $dx/dt = u - c$ leaves the domain, the characteristic $dx/dt = u + c$ enters the domain. The invariant W_1 at any point M on the boundary is known from its value at the internal point B. In contrast, the unknown invariant W_2 cannot be determined from internal points. Additional information must be supplied in the form of a boundary condition. Two equations may be written, the first for the Riemann invariant along the characteristic [BM], the

second for the boundary condition. Except for very simple geometries and boundary conditions, these equations are nonlinear. They can be written in general form as:

$$\left. \begin{aligned} u_M - \beta c_M &= u_B - \beta c_B + (t_M - t_B) S''_{BM} \\ f_b(u_M, c_M, t) &= 0 \end{aligned} \right\} [2.167]$$

where the (assumed known) function f_b expresses the boundary condition in the form of a time-dependent relationship between u and c . Three main types of boundary conditions are used in practice: prescribed water level (or depth), prescribed discharge, stage-discharge relationship. Their expression for a rectangular channel can be found in section 2.5.4.4. In the general case, system [2.167] must be solved using iterative techniques.

(2) Inflowing supercritical flow. The two characteristics $dx/dt = u - c$ and $dx/dt = u + c$ enter the domain. The solution at the boundary point M is not influenced by the solution at internal points. Two boundary conditions must be supplied. This is usually done by prescribing the water level and the discharge as functions of time.

(3) Outflowing supercritical flow. The two characteristics $dx/dt = u - c$ and $dx/dt = u + c$ leave the domain. The solution at the boundary point M is uniquely determined from the internal points. It can be calculated by applying relationships [2.166] between M and the internal points A and B, M being treated exactly in the same way as an internal point (see section 2.5.4.2).

2.5.4.4. Boundary conditions for a rectangular channel

The full expression of the main three types of boundary conditions in a rectangular channel where the flow is subcritical is provided hereafter. The channel is assumed to be prismatic, the slope and the friction term are neglected, which leads to a zero source term S'' . Remember that for a rectangular channel, the speed of the waves in still water is given by $c = (gh)^{1/2}$.

– Prescribed water level h_b at the left-hand boundary. The invariant W_1 is used together with the boundary condition:

$$\left. \begin{aligned} u_M - 2c_M &= u_B - 2c_B \\ c_M &= (gh_b)^{1/2} \end{aligned} \right\} [2.168]$$

Solving system [2.168] for u_M yields:

$$u_M = u_B - 2c_B + 2(gh_b)^{1/2} [2.169]$$

– Prescribed discharge Q_b at the left-hand boundary. The invariant W_1 is used again:

$$\left. \begin{aligned} u_M - 2c_M &= u_B - 2c_B \\ u_M A_M &= \frac{bc_M^2 u_M}{g} = Q_b \end{aligned} \right\} [2.170]$$

Note that the relationship $A = bh = bc^2/g$ is used in the second equation. The nonlinear system [2.170] must be solved iteratively for u_M and c_M .

– Prescribed stage-discharge relationship at the left-hand boundary. The invariant W_1 and the boundary condition are used:

$$\left. \begin{aligned} u_M - 2c_M &= u_B - 2c_B \\ f_b(h_M, u_M A_M, t) &= 0 \end{aligned} \right\} [2.171]$$

Using the relationship $A = bc^2/g$ leads to:

$$\left. \begin{aligned} u_M - 2c_M &= u_B - 2c_B \\ f_b(c_M, u_M, t) &= 0 \end{aligned} \right\} [2.172]$$

In the general case system [2.172] is nonlinear and must be solved iteratively.

– Prescribed water depth h_b at the right-hand boundary. The invariant W_2 is used together with the boundary condition:

$$\left. \begin{aligned} u_M + 2c_M &= u_A + 2c_A \\ c_M &= (gh_b)^{1/2} \end{aligned} \right\} [2.173]$$

Solving system [2.173] for u_M yields:

$$u_M = u_A + 2c_A - 2(gh_b)^{1/2} [2.174]$$

– Prescribed discharge Q_b at the right-hand boundary. The following system is obtained:

$$\left. \begin{aligned} u_M + 2c_M &= u_A + 2c_A \\ \frac{bc_M^2 u_M}{g} &= Q_b \end{aligned} \right\} [2.175]$$

The nonlinear system [2.175] must be solved iteratively.

– Prescribed stage-discharge relationship at the right-hand boundary. The invariant W_2 is used together with the boundary condition that is rewritten in the form of a time-dependent relationship between u and c :

$$\left. \begin{aligned} u_M + 2c_M &= u_A + 2c_A \\ f_b(c_M, u_M, t) &= 0 \end{aligned} \right\} \quad [2.176]$$

In this case again, an iterative solution technique is needed.

2.5.5. Summary

The Saint Venant equations are derived on the assumption of incompressible water, of a hydrostatic pressure field, for turbulent flow in nearly horizontal channels.

The structure of the Saint Venant equations is that of a 2×2 compressible system. The wave speeds are $u - c$ and $u + c$. The general formula for the speed c of the waves in still water is given by equation [2.137]. It is given by equation [2.138] for a rectangular channel and by equation [2.145] for a triangular channel.

The general form of the Riemann invariants is defined as in equation [2.140]. Equations [2.149] and equations [2.156–157] provide an alternative writing. These differential relationships may be integrated easily for simple geometries such as rectangular prismatic channels (see equations [2.143–144]) or triangular prismatic channels (see equations [2.146–147]). In prismatic channels with arbitrarily-shaped cross-sections, the (equivalent) formulations [2.151] and [2.158] are applicable, with Riemann invariants defined as in equation [2.159] when dh/db can be assumed to be proportional to A/b^2 . If this is not the case, equations [2.151], [2.158] and [2.159] are not exact formulae but approximations of the Riemann invariants.

The flow regime can be characterized using the so-called Froude number Fr , defined as the ratio of the flow velocity u to the speed c of the waves in still water (see equation [2.160]). The flow is said to be subcritical if Fr is smaller than one, supercritical if Fr is larger than one, and critical if Fr is equal to one. The Froude number has the same meaning as the Mach number used for the Euler equations of gas dynamics (see section 2.6).

The solution is unique provided that the initial condition is known at the internal points of the computational domain and that a boundary condition is specified for each characteristic that enters the domain at the boundary of the domain. When the

flow is subcritical at a given boundary, only one condition is required. When the flow is supercritical, entering the domain, two boundary conditions are needed. When the flow is supercritical, leaving the domain, no boundary condition is required.

2.6. A nonlinear 3×3 system: the Euler equations

2.6.1. Physical context – assumptions

The Euler equations are among the simplest possible systems of governing equations for compressible gas dynamics. They can be seen as a simplified form of the Navier-Stokes equations in the absence of diffusion terms, with the addition of an equation for energy. The Euler equations are derived under the following assumptions:

- Assumption (A1). The gas is compressible. It obeys the equation of state for a perfect gas, whereby the gas internal energy (thus the gas temperature) is a function of the gas pressure and density.
- Assumption (A2). The density, the momentum and the total energy of the gas are assumed to be conserved. The total energy of the gas is formed by the kinetic energy, the internal energy and the energy per unit volume, i.e. the pressure. The total energy should be distinguished from the mechanical energy of the gas, which does not account for thermal effects.
- Assumption (A3). The effects of viscosity, turbulence and heat conduction are neglected. Consequently, the Euler equations do not account for momentum or heat diffusion.
- Assumption (A4). The effects of volume forces such as gravity are negligible.

The Euler equations are used in aerodynamics for far field flow simulations. In near field simulations, a turbulent boundary layer appears in the neighborhood of walls and obstacles that is not accounted for by the Euler equations. Since the continuity and momentum equations in the Euler equations account for the hyperbolic part of the Navier-Stokes equations, their understanding and the understanding of the properties of their solution appears as an important prerequisite in the study of the complete Navier-Stokes equations with energy transport. In fact, the Euler equations allow a number of fundamental phenomena of aerodynamics to be accounted for:

- As shown in sections 2.6.3 and 2.6.4, Assumption (A1) accounts for the fact that the Euler equations describe a compressible flow system. The pressure waves (also called the sound waves) propagate at a finite speed. The Euler equations allow acoustic phenomena to be accounted for.

– The Euler equations allow the occurrence of subsonic, sonic and supersonic flow regimes to be explained.

– The Euler equations allow thermodynamic effects associated with the sudden expansion or compression of gases to be accounted for (deflagrations, etc.).

The study of the Euler equation therefore appears as an indispensable step in the study and practice of aerodynamic modeling.

2.6.2. Conservation form

2.6.2.1. Definitions – notation

The following notation is used. The gas density, velocity and pressure are denoted by ρ , u and p respectively. The total energy per unit volume, denoted by E , is defined as follows:

$$E = \left(\frac{u^2}{2} + e \right) \rho \quad [2.177]$$

where e is the internal energy of the gas. From Assumption (A1), e is a function of p and ρ :

$$e = e(p, \rho) \quad [2.178]$$

The internal energy of a perfect gas is expressed as:

$$e(p, \rho) = \frac{P}{(\gamma - 1)\rho} \quad [2.179]$$

where $\gamma = 7/5$. The entropy s is defined in differential form as:

$$ds = \frac{de + pd(1/\rho)}{T} \quad [2.180]$$

where T is the temperature, assumed to satisfy the equation of state for a perfect gas:

$$RT = \frac{P}{\rho} \quad [2.181]$$

Substituting equations [2.181] and [2.179] into equation [2.180] yields the following expression:

$$ds = \frac{R}{\gamma - 1} \left(\frac{dp}{p} - \gamma \frac{d\rho}{\rho} \right) = \frac{R}{\gamma - 1} d \left[\ln \left(\frac{p}{\rho^\gamma} \right) \right] \quad [2.182]$$

The conservation form of the one-dimensional continuity, momentum and energy equations is obtained from a balance over a control volume of unit cross-sectional area extending from $x = x_0$ to $x = x_0 + \delta x$ (Figure 2.18).

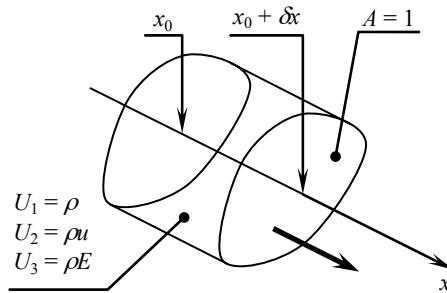


Figure 2.18. Mass, momentum and energy balance over a control volume

2.6.2.2. Continuity equation

The mass balance can be written as follows:

$$\delta U_1(t_0 + \delta t) - \delta U_1(t_0) = \delta F_1(x_0) - \delta F_1(x_0 + \delta x) \quad [2.183]$$

where $\delta U_1(t)$ is the mass of fluid contained in the control volume at the time t and $\delta F_1(x)$ is the mass of fluid that passes at the abscissa x during the time interval δt . Equation [2.183] expresses the fact that the variation of the mass contained in the control volume is due to the difference between the inflowing and outflowing mass fluxes across the control sections located at x_0 and $x_0 + \delta x$. δU_1 and δF_1 are defined as:

$$\left. \begin{aligned} \delta U_1(t) &= \int_{x_0}^{x_0 + \delta x} (\rho A)(x, t) dx = A \int_{x_0}^{x_0 + \delta x} \rho(x, t) dx \\ \delta F_1(x) &= \int_{t_0}^{t_0 + \delta t} (\rho u A)(x, t) dx = A \int_{t_0}^{t_0 + \delta t} (\rho u)(x, t) dx \end{aligned} \right\} \quad [2.184]$$

Substituting equation [2.184] into equation [2.183], dividing by A in the limit of small δt and δx (see equations [1.12–16] for a detailed proof) leads to the following equation for continuity:

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho u) = 0 \quad [2.185]$$

Equation [2.185] can be written in the form [2.1] as:

$$\left. \begin{aligned} \frac{\partial U_1}{\partial t} + \frac{\partial F_1}{\partial x} &= 0 \\ U_1 &= \rho \\ F_1 &= \rho u \end{aligned} \right\} \quad [2.186]$$

2.6.2.3. Momentum equation

A momentum balance over the control volume between t_0 and $t_0 + \delta t$ yields:

$$\delta U_2(t_0 + \delta t) - \delta U_2(t_0) = \delta F_2(x_0) - \delta F_2(x_0 + \delta x) + \delta P(x_0) - \delta P(x_0 + \delta x) \quad [2.187]$$

where $\delta U_2(t)$ is the momentum contained in the control volume at the time t and $\delta F_2(x)$ is the momentum attached to the volume of fluid that crosses the abscissa x between t_0 and $t_0 + \delta t$. $\delta P(x)$ represents the integral of the pressure force with respect to time between t_0 and $t_0 + \delta t$. δU_2 and δF_2 are defined as:

$$\left. \begin{aligned} \delta U_2(t) &= \int_{x_0}^{x_0 + \delta x} (\rho u A)(x, t) dx = A \int_{x_0}^{x_0 + \delta x} (\rho u)(x, t) dx \\ \delta F_2(x) &= \int_{t_0}^{t_0 + \delta t} (\rho u^2 A)(x, t) dx = A \int_{t_0}^{t_0 + \delta t} (\rho u^2)(x, t) dx \end{aligned} \right\} \quad [2.188]$$

The pressure force P is the product of the pressure p and the cross-sectional area A of the control volume:

$$\delta P(t) = \int_{t_0}^{t_0 + \delta t} (Ap)(x, t) dt = A \int_{t_0}^{t_0 + \delta t} p(x, t) dt \quad [2.189]$$

Substituting equations [2.188–189] into equation [2.187], dividing by A in the limit of small δt and δx (see equations [1.12–16] for details of the proof) leads to the following equation:

$$\frac{\partial}{\partial t}(\rho u) + \frac{\partial}{\partial x}(\rho u^2 + p) = 0 \quad [2.190]$$

that can be written in the form [2.1] as:

$$\left. \begin{aligned} \frac{\partial U_2}{\partial t} + \frac{\partial F_2}{\partial x} &= 0 \\ U_2 = \rho u = F_1 \\ \rho u^2 = \frac{(\rho u)^2}{\rho} &= \frac{U_2^2}{U_1} \end{aligned} \right\} \quad [2.191]$$

2.6.2.4. Energy equation

The energy balance over the control volume can be written as:

$$\delta U_3(t_0 + \delta t) - \delta U_3(t_0) = \delta F_3(x_0) - \delta F_3(x_0 + \delta x) + \delta W_p(x_0) - \delta W_p(x_0 + \delta x) \quad [2.192]$$

where $\delta U_3(t)$ is the total energy contained in the control volume at the time t , $\delta F_3(x)$ is the total energy attached to the volume of fluid that crosses the abscissa x during the time interval δt and $\delta W_p(x)$ is the work of the pressure force exerted onto the section at the abscissa x during the time interval δt . δU_3 and δF_3 are defined as:

$$\left. \begin{aligned} \delta U_3(t) &= \int_{x_0}^{x_0 + \delta x} (AE)(x, t) dx = A \int_{x_0}^{x_0 + \delta x} E(x, t) dx \\ \delta F_3(x) &= \int_{t_0}^{t_0 + \delta t} (AuE)(x, t) dx = A \int_{t_0}^{t_0 + \delta t} (uE)(x, t) dx \end{aligned} \right\} \quad [2.193]$$

The work $dW_p(x)$ of the pressure force over an infinitesimal time interval dt is defined as the product of the pressure force Ap and the displacement $u dt$ of the fluid. Consequently:

$$\delta W_p(t) = A \int_{t_0}^{t_0 + \delta t} up(x, t) dt \quad [2.194]$$

Substituting equations [2.193–194] into equation [2.192], dividing by A in the limit of small δt and δx leads to (see equations [1.12–16] for the details of the reasoning):

$$\frac{\partial E}{\partial t} + \frac{\partial}{\partial x}(uE + up) = 0 \quad [2.195]$$

The energy equation can be rewritten in the conservation form [2.1] as:

$$\left. \begin{aligned} \frac{\partial U_3}{\partial t} + \frac{\partial F_3}{\partial x} &= 0 \\ U_3 &= E \\ F_3 &= (E + p)u = (U_3 + p)\frac{U_2}{U_1} \end{aligned} \right\} \quad [2.196]$$

2.6.2.5. Vector conservation form

Equations [2.186], [2.191] and [2.196] can be rewritten in the vector conservation form [2.2] by defining U and F as:

$$U = \begin{bmatrix} \rho \\ \rho u \\ E \end{bmatrix}, \quad F = \begin{bmatrix} \rho u \\ \rho u^2 + p \\ (E + p)u \end{bmatrix} = \begin{bmatrix} U_2 \\ U_2^2 / U_1 + p \\ (U_3 + p)U_2 / U_1 \end{bmatrix} \quad [2.197]$$

2.6.3. Characteristic form – Riemann invariants

System [2.2] can be rewritten in the non-conservation form [2.5] by setting the source term to zero:

$$\frac{\partial U}{\partial t} + A \frac{\partial U}{\partial x} = 0 \quad [2.198]$$

where A is defined as $\partial F / \partial U$. Defining the sound speed c as:

$$c^2 = \left(\frac{\partial p}{\partial \rho} \right)_{ds=0} \quad [2.199]$$

leads to the following expression for A :

$$A = \begin{bmatrix} 0 & 1 & 0 \\ c^2 - u^2 & 2u & 0 \\ uc^2 - (E + p)u / \rho & (E + p) / \rho & u \end{bmatrix} \quad [2.200]$$

A has the following eigenvalues and eigenvectors:

$$\lambda^{(1)} = u - c \quad \lambda^{(2)} = u \quad \lambda^{(3)} = u + c$$

$$K^{(1)} = \begin{bmatrix} 1 \\ u - c \\ (E + p) / \rho - uc \end{bmatrix} \quad K^{(2)} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad K^{(3)} = \begin{bmatrix} 1 \\ u + c \\ (E + p) / \rho + uc \end{bmatrix} \quad [2.201]$$

The vector of the Riemann invariants is defined in differential form as $dW = K^{-1}dU \cdot K^{-1}$ is given by:

$$K^{-1} = \frac{1}{2c} \begin{bmatrix} u + c & -1 & 0 \\ 2[u^2 - (E + p) / \rho]c & -2uc & 2c \\ c - u & 1 & 0 \end{bmatrix} \quad [2.202]$$

hence the expression of dW :

$$\left. \begin{aligned} (u + c) \frac{d\rho}{dt} - \frac{d}{dt}(\rho u) &= 0 & \text{for } \frac{dx}{dt} = u - c \\ \left(u^2 - \frac{E + p}{\rho} \right) \frac{d\rho}{dt} - u \frac{d}{dt}(\rho u) + \frac{dE}{dt} &= 0 & \text{for } \frac{dx}{dt} = u \\ (c - u) \frac{d\rho}{dt} + \frac{d}{dt}(\rho u) &= 0 & \text{for } \frac{dx}{dt} = u + c \end{aligned} \right\} \quad [2.203]$$

Noting that $d(\rho u) = \rho du + u d\rho$, equations [2.203] become:

$$\left. \begin{aligned} c \frac{d\rho}{dt} - \rho \frac{du}{dt} &= 0 & \text{for } \frac{dx}{dt} = u - c \\ \left(-\frac{E + p}{\rho} \right) \frac{d\rho}{dt} - \rho \frac{d}{dt} \left(\frac{u^2}{2} \right) + \frac{dE}{dt} &= 0 & \text{for } \frac{dx}{dt} = u \\ c \frac{d\rho}{dt} + \rho \frac{du}{dt} &= 0 & \text{for } \frac{dx}{dt} = u + c \end{aligned} \right\} \quad [2.204]$$

The term $E/\rho d\rho$ is removed from the second equation [2.204] by noting that:

$$dE = d\left(\rho \frac{E}{\rho}\right) = \frac{E}{\rho} d\rho + \rho d\left(\frac{E}{\rho}\right) \quad [2.205]$$

Substituting equation [2.205] into the second equation [2.204] yields:

$$-\frac{p}{\rho^2} \frac{d\rho}{dt} + \frac{d}{dt} \left(\frac{E}{\rho} - \frac{u^2}{2} \right) = 0 \quad \text{for } \frac{dx}{dt} = u \quad [2.206]$$

Using definition [2.177], equation [2.206] is transformed into:

$$p \frac{d}{dt} \left(\frac{1}{\rho} \right) + \frac{de}{dt} = 0 \quad \text{for } \frac{dx}{dt} = u \quad [2.207]$$

Using equation [2.180] leads to the following expression:

$$\frac{ds}{dt} = 0 \quad \text{for } \frac{dx}{dt} = u \quad [2.208]$$

The density ρ not being measurable directly, the first and third characteristic equations are rewritten so as to involve the pressure p that can be measured directly. This is done by substituting relationship [2.199] into equations [2.204]:

$$\left. \begin{array}{l} \frac{dp}{dt} - \rho c \frac{du}{dt} = 0 \quad \text{for } \frac{dx}{dt} = u - c \\ \frac{ds}{dt} = 0 \quad \text{for } \frac{dx}{dt} = u \\ \frac{dp}{dt} + \rho c \frac{du}{dt} = 0 \quad \text{for } \frac{dx}{dt} = u + c \end{array} \right\} \quad [2.209]$$

Note that definition [2.199] allows for an explicit definition of the speed of sound c . This definition can be used to integrate the Riemann invariants. Indeed, c is defined for a constant entropy, that is, for $ds = 0$. Using the differential form [2.184], it is easy to check that assuming $ds = 0$ is equivalent to assuming a constant ratio p^γ/ρ . Consequently, the pressure and the density are related by a law of the type:

$$p = p_0 \left(\frac{\rho}{\rho_0} \right)^\gamma \quad [2.210]$$

where p_0 is a reference pressure for which the density is equal to ρ_0 . Substituting equation [2.210] into equation [2.199] yields:

$$c = \left(\frac{dp}{d\rho} \right)_{ds=0}^{1/2} = \left(\gamma \frac{p_0}{\rho_0^\gamma} \rho^{\gamma-1} \right)^{1/2} = \left(\gamma \frac{p}{\rho} \right)^{1/2} \quad [2.211]$$

Using [2.210] and [2.211], the following expression is obtained for ρc :

$$\rho c = (\gamma \rho p)^{1/2} = \left[\gamma \rho_0 \left(\frac{p}{p_0} \right)^{1/\gamma} p \right]^{1/2} = \left(\frac{\gamma \rho_0}{p_0^{1/\gamma}} \right)^{1/2} p^{\frac{\gamma+1}{2\gamma}} \quad [2.212]$$

The first Riemann invariant becomes:

$$\left(\frac{\gamma \rho_0}{p_0^{1/\gamma}} \right)^{1/2} p^{\frac{\gamma+1}{2\gamma}} dp - du = 0 \quad \text{for } \frac{dx}{dt} = u - c \quad [2.213]$$

Equation [2.213] can be integrated into:

$$\beta_1 p^{\beta_2} - u = \text{Const} \quad \text{for } \frac{dx}{dt} = u - c \quad [2.214]$$

where the coefficients β_1 and β_2 are defined as:

$$\left. \begin{aligned} \beta_1 &= \frac{2\gamma}{\gamma+1} \left(\frac{\gamma \rho_0}{p_0^{1/\gamma}} \right)^{1/2} \\ \beta_2 &= \frac{3\gamma+1}{2\gamma} \end{aligned} \right\} \quad [2.215]$$

Applying the same reasoning to the third Riemann invariant, the vector \mathbf{W} is eventually defined as:

$$\mathbf{W} = \begin{bmatrix} \beta_1 p^{\beta_2} - u \\ s \\ \beta_1 p^{\beta_2} + u \end{bmatrix} \quad [2.216]$$

The characteristic equations are:

$$\left. \begin{array}{l} \frac{d}{dt}(u - \beta_1 p^{\beta_2}) = 0 \quad \text{for } \frac{dx}{dt} = u - c \\ \frac{ds}{dt} = 0 \quad \text{for } \frac{dx}{dt} = u - c \\ \frac{d}{dt}(u + \beta_1 p^{\beta_2}) = 0 \quad \text{for } \frac{dx}{dt} = u + c \end{array} \right\} \quad [2.217]$$

Note that this expression is valid only because the equation of state [2.179] allows the system to be closed, allowing equation [2.210] to be derived from equation [2.182].

2.6.4. Calculation of the solution

2.6.4.1. The various possible flow regimes

As seen in section 2.6.3, the three wave speeds are $u - c$, u and $u + c$. They reflect the combination of two physical processes:

- the variables that reflect the local state of the flow (the pressure p , the entropy s , the density ρ) are transported with the local flow velocity u ;
- the propagation of the pressure waves is “superimposed” onto the transport phenomenon. The pressure waves, also called the sound waves, arise from the equation of state that provides a relationship between the pressure and the density. From the point of view of an observer moving at the fluid velocity u , the pressure waves propagate in opposite directions at speeds $-c$ and $+c$.

Note that the equation of state between the pressure and the density is essential to the hyperbolic character of the Euler equations. Indeed, the inviscid Burgers equation seen in section 1.4 is derived on the basis of equations [1.62] and [1.63] that do not form a hyperbolic system. In contrast, the continuity and momentum equations [2.186] and [2.191] in the Euler equations form a hyperbolic system (adding the energy equation [2.196] is not necessary for the system to be hyperbolic). Equations [2.186, 2.191] differ from system [1.62–63] only by the pressure term p in equation [2.191]. Therefore the presence of the pressure term is the necessary condition for the system to be hyperbolic.

The flow regime may be characterized by the dimensionless Mach number M , defined as the ratio of the flow velocity to the speed of sound:

$$M = \frac{|u|}{c} \quad [2.218]$$

The Mach number is the equivalent of the Froude number used for open channel flows (see section 2.5). Three types of flow regimes are distinguished:

- If M is smaller than one, the flow is said to be subsonic. The fluid velocity is smaller than the speed of sound. The characteristic $dx/dt = u - c$ influences the points located upstream, while the characteristics $dx/dt = u$ and $dx/dt = u + c$ travel downstream, thus influencing the points located downstream.
- If M is larger than one, the flow is said to be supersonic. The flow velocity being larger than the speed of sound, the three waves travel in the downstream direction. A perturbation arising in the flow cannot influence the points located upstream of its original location.
- If M is strictly equal to one, the flow is said to be sonic, or transonic.

The subsonic, transonic and supersonic flow regimes are illustrated in Figure 2.19 for a flow directed from left to right.

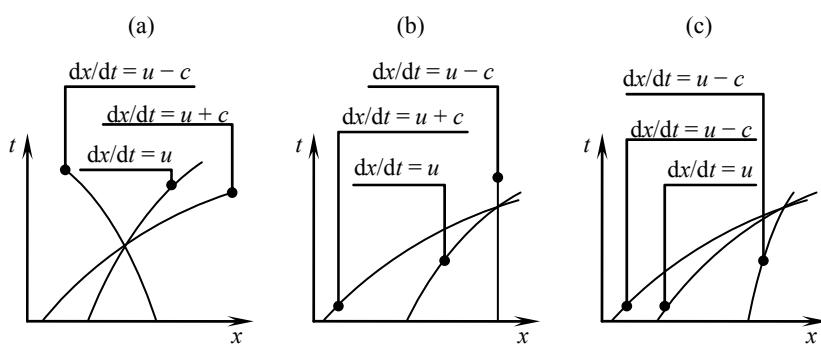


Figure 2.19. Representation of the various possible flow regimes in the phase space: subsonic (a), transonic (b), supersonic (c). Sketches for a positive flow velocity

2.6.4.2. Treatment of internal points

This section focuses on the solution of the Euler equations at the internal points M of a computational domain that is assumed to extend from $x = 0$ to $x = L$ (Figure 2.20). The feet of the characteristics $u + c$, u and $u - c$ passing at M are denoted by A , B and C respectively. A , B and C may be located on a domain boundary as well as be internal points. The flow variables ρ , u and E are assumed to be known at A , B and C .

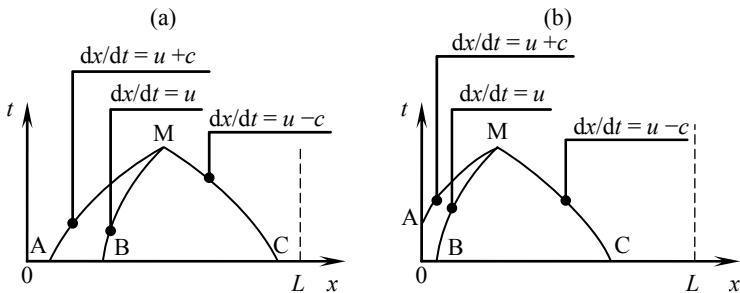


Figure 2.20. Calculation of the solution at internal points.
The domain of dependence may be entirely included in the computational domain (a) or include a boundary (b)

The Riemann invariants [2.224] allow the following relationships to be written:

$$\left. \begin{aligned} \beta_1 p_M^{\beta_2} - u_M &= \beta_1 p_C^{\beta_2} - u_C \\ s_M &= s_B \\ \beta_1 p_M^{\beta_2} + u_M &= \beta_1 p_A^{\beta_2} + u_A \end{aligned} \right\} \quad [2.219]$$

Owing to the presence of the nonlinear pressure terms in equations [2.219] and to the dependence of the sound speed c on the pressure, the location of A , B and C cannot be determined analytically in the general case. The calculation of an analytical solution at M is therefore impossible in most practical applications where the initial and boundary conditions are arbitrary functions of space and time. However, analytical or semi-analytical solutions can be found for problems based on simple initial conditions. Such problems include the Riemann problem covered in detail in Chapter 4. The Riemann problem is used in a number of numerical techniques for hyperbolic systems of conservation laws.

Equations [2.219] involve the initial values of p , u and s over the domain of dependence of the solution. The solution can be calculated at internal points only if the initial condition is known over the entire domain $[0, L]$. The knowledge of the initial condition is a necessary condition. It is not a sufficient condition because the knowledge of the solution at the boundaries of the domain is necessary when the domain of dependence of the point M includes the boundaries, as illustrated in Figure 2.20b. The determination of the solution at the boundaries is detailed in the next section.

2.6.4.3. Treatment of boundary points

The number of boundary conditions to be supplied depends on the regime and direction of the flow. The following four configurations may occur (Figure 2.21): the flow is subsonic, leaving the domain (Figure 2.21a); the flow is subsonic, entering the domain (Figure 2.21b); the flow is supersonic, entering the domain (Figure 2.21c); or the flow is supersonic, leaving the domain (Figure 2.21d). These configurations are examined for a left-hand boundary, the treatment of a right-hand boundary being deduced by symmetry.

– *Subsonic flow leaving the domain* (Figure 2.21a). The characteristics $dx/dt = u - c$ and $dx/dt = u$ leave the domain. The flow at the boundary point M is influenced by the Riemann invariants W_2 and W_1 coming respectively from the feet B and C of the characteristics. The missing information on the Riemann invariant W_3 must be supplied in the form of a boundary condition, that is, a possibly time-dependent relationship between the pressure, the flow velocity and the entropy (or the density). This leads to the following system of equations:

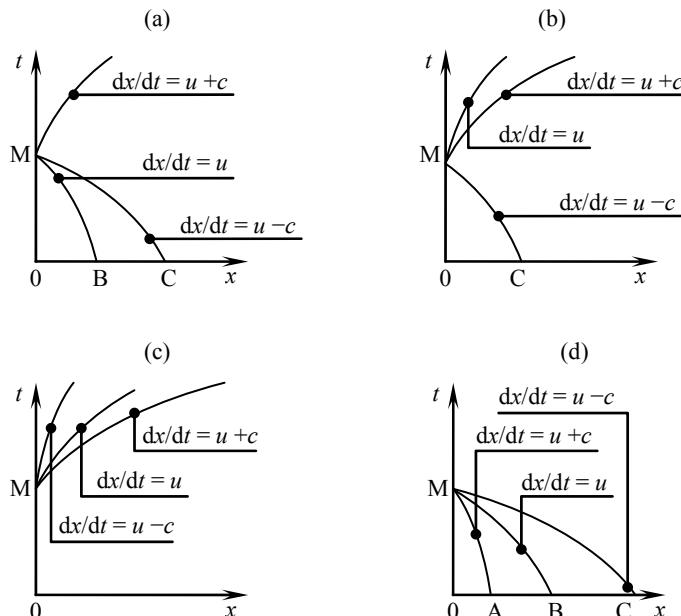


Figure 2.21. Treatment of boundary points (here on the left-hand boundary).
Subsonic flow leaving the domain (a), subsonic flow entering the domain (b), supersonic flow entering the domain (c), supersonic flow leaving the domain (d)

$$\left. \begin{array}{l} f(p_M, u_M, s_M, t) = 0 \\ \beta_1 p_M^{\beta_2} - u_M = \beta_1 p_C^{\beta_2} - u_C \\ s_M = s_B \end{array} \right\} [2.220]$$

– *Subsonic flow entering the domain* (Figure 2.21b). The characteristic $dx/dt = u - c$ leaves the domain, the remaining two characteristics enter the domain. The invariant W_3 is known from the initial condition at the foot C of the characteristic. The missing information about the remaining two invariants must be specified in the form of boundary conditions. This results in the following system:

$$\left. \begin{array}{l} f_1(p_M, u_M, s_M, t) = 0 \\ f_2(p_M, u_M, s_M, t) = 0 \\ \beta_1 p_M^{\beta_2} - u_M = \beta_1 p_C^{\beta_2} - u_C \end{array} \right\} [2.221]$$

– *Supersonic flow entering the domain* (Figure 2.21c). The three characteristics enter the domain. The behavior of the flow at the boundary is not influenced by the flow conditions inside the domain. Three boundary conditions must be specified:

$$\left. \begin{array}{l} p_M = f_1(t) \\ u_M = f_2(t) \\ s_M = f_3(t) \end{array} \right\} [2.222]$$

– *Supersonic flow leaving the domain* (Figure 2.21d). The three characteristics leave the domain. The flow conditions at M are entirely determined by the flow conditions at points A, B and C. System [2.219] is solved exactly as if M was an internal point. No boundary condition is required.

2.6.5. Summary

The Euler equations form a 3×3 hyperbolic system of conservation laws. The governing equations are derived from the assumptions of a compressible flow with negligible volume forces, momentum diffusion and heat diffusion.

The Euler equations verify the definitions of a compressible flow system. The speeds of the waves are $u - c$, u and $u + c$. The speed of sound c is given by equations [2.199] and [2.211]. The existence of a finite sound speed reflects the existence of a relationship between the pressure and the density. The hyperbolic character of the Euler equations stems directly from this relationship. If the pressure was independent of the density, the continuity and the momentum equation would

be degraded into a simpler formulation of the type [1.62–63] that leads to the inviscid Burgers equation. System [1.62–63] is not hyperbolic.

The Riemann invariants are defined in differential form by equations [2.209]. They can be integrated into equations [2.216]. Their analytical determination is in general impossible because the speed of the characteristics depends on the solution itself, which makes the analytical determination of the feet of the characteristics very difficult, if not impossible.

The flow regime is characterized by the dimensionless Mach number M , defined as in equation [2.218]. The Mach number is the ratio of the flow velocity to the speed of sound. The flow is said to be subsonic when M is smaller than one, sonic or transonic if M is strictly equal to one, and supersonic when M is larger than one.

The solution is determined uniquely over the computational domain provided that (i) the initial condition is known over all the domain and (ii) a boundary condition is supplied for each characteristic that enters the domain. No condition is needed at boundaries where the flow is supersonic, leaving the domain. A boundary with a supersonic inflow requires three conditions. A boundary where the flow is subsonic requires one or two conditions depending on whether the flow leaves or enters the domain.

2.7. Summary of Chapter 2

2.7.1. *What you should remember*

Hyperbolic systems of conservation laws may be expressed in conservation, non-conservation or characteristic form. They reflect the propagation of several waves at different finite speeds.

Coupling several scalar hyperbolic conservation laws does not necessarily yield a hyperbolic system. As shown in section 2.1.2, coupling the scalar hyperbolic equations for continuity and momentum conservation does not lead to a hyperbolic system.

A flow system is said to be compressible if its governing equations form a hyperbolic system of conservation laws with the additional criteria that (i) the system includes at least an equation for the conservation of mass and an equation for the conservation of momentum, (ii) the pressure force is related to the mass per unit volume via an equation of state. The water hammer equations, the Saint Venant equations and the Euler equations describe the behavior of compressible flow systems.

The wave speeds are the eigenvalues of the Jacobian matrix of the flux vector with respect to the conserved variable vector.

The water hammer equations presented in section 2.4 form a 2×2 hyperbolic system of conservation laws. The two waves propagate in opposite directions at the speed of sound. The wave speeds are functions of the local characteristics of the pipe and do not depend on the local flow conditions. The solution can be determined uniquely over a computational domain of finite length provided that the initial condition is known over the entire computational domain and that exactly one boundary condition is specified at each boundary of the domain.

The Saint Venant equations dealt with in section 2.5 form a 2×2 hyperbolic system of conservation laws where the wave speeds are the sum of the local flow velocity and the propagation speed of the waves in still water. The wave speeds are not constant in the general case because they depend on the local flow conditions. Their sign may change depending on whether the flow is subcritical, critical or supercritical. The solution of the Saint Venant equations is determined uniquely over a computational domain of finite length provided that the initial condition is known everywhere over the domain and that one boundary condition is specified for each characteristic that enters the domain at the boundaries.

The Euler equations dealt with in section 2.6 form a 3×3 hyperbolic system of conservation laws, the wave speeds of which are combinations of the local flow velocity and the sound speed. The hyperbolic character of the system stems directly from the presence of the pressure term in the momentum equation. The wave speeds are functions of the local characteristics of the flow. Their sign may change depending on the subsonic, sonic or supersonic nature of the flow. The solution of the Euler equations is determined uniquely over a computational domain of finite length provided that the initial condition is known over the entire domain and that one boundary condition is prescribed for each characteristic that enters the domain.

2.7.2. Application exercises

2.7.2.1. Exercise 2.1: the water hammer equations

Consider a horizontal pipe of cross-sectional area A , where the pressure waves propagate at the speed c . Friction is assumed to be negligible. The initial flow conditions are steady state conditions, with a pressure and velocity uniformly equal to p_0 and u_0 respectively.

A variation Δp in the pressure appears at the left-hand end of the pipe and propagates to the right at the speed c (Figure 2.22). As a consequence, a variation Δu appears in the flow velocity.

1) Show that Δp and Δu verify the following relationship:

$$\Delta p = \rho c \Delta u \quad [2.223]$$

2) Assume now that the wave propagates from right to left. Show that the following relationship holds between Δp and Δu :

$$\Delta p = -\rho c \Delta u \quad [2.224]$$

These equations are called Joukowski's relationships.

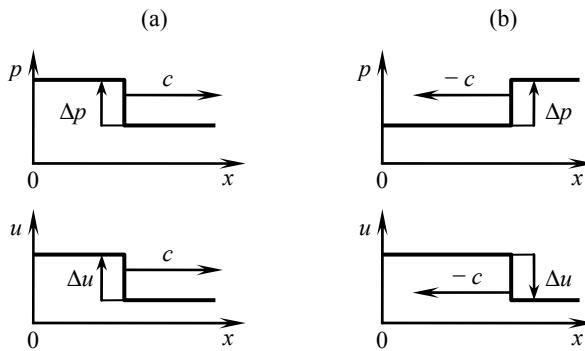


Figure 2.22. Propagation of a pressure and velocity variation in a pipe.
Propagation from left to right (a), from right to left (b)

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

2.7.2.2. Exercise 2.2: the water hammer equations

Consider a horizontal pipe of cross-sectional area A , where the speed of the pressure waves is piecewise constant. The wave speed to the left of the point $x = x_0$ is denoted by c_1 , the wave speed to the right of $x = x_0$ is denoted by c_2 . The fluid is initially at rest, the pressure is uniformly equal to p_0 . The influence of friction is assumed to be negligible.

At $t = 0$ the pressure at the left-hand end of the pipe rises instantaneously to the constant value p_1 . The resulting pressure discontinuity propagates to the right at a speed c_1 .

1) Derive the expression of the discharge Q_1 on the left-hand side of the pressure discontinuity.

2) The pressure wave reaches the abscissa x_0 where the speed of sound changes to c_2 . Considering that the pressure is continuous at $x = x_0$, show that the pressure and the discharge change to new values p_2 and Q_2 when the pressure wave reaches $x = x_0$. Provide the expression of p_2 and Q_2 as functions of p_0 , p_1 , c_1 and c_2 .

3) Show that the pressure surge is amplified if $c_1 < c_2$ (in other words, $|p_2 - p_0| > |p_1 - p_0|$). Conversely, show that the pressure surge is damped if $c_1 > c_2$. Provide a physical interpretation for such a behavior.

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

2.7.2.3. Exercise 2.3: the water hammer equations

Consider a horizontal pipe, the cross-sectional area is the following function of the longitudinal coordinate x (Figure 2.23):

- for $x < x_1$ the section is constant, equal to A_1 ;
- for $x > x_2 > x_1$ the section is constant, equal to A_2 ;
- for $x_1 < x < x_2$ the section varies continuously from A_1 to A_2 .

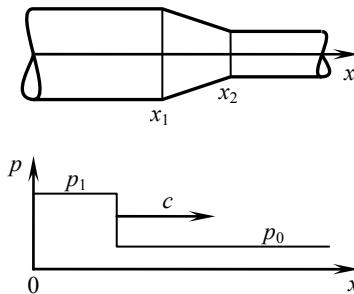


Figure 2.23. Propagation of a pressure wave in a pipe with variable cross-sectional area

A_2 may be larger or smaller than A_1 . The wave speed c is the same all along the pipe. The effects of friction are assumed to be negligible. The water is initially flowing with a uniform pressure p_0 and a uniform discharge Q_0 . At $t = 0$, the pressure at the left-hand end of the pipe changes instantaneously from p_0 to p_1 . The resulting pressure discontinuity propagates to the right at the constant wave speed c .

1) Provide the expression of the discharge Q_1 on the left-hand side of the pressure wave before the pressure wave reaches the abscissa x_1 .

- 2) Provide the expression for the pressure p_2 and the discharge Q_2 when the pressure wave reaches the abscissa x_2 . What is the effect of a narrowing on the pressure transient? What is the effect of a widening?

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

2.7.2.4. Exercise 2.4: the Saint Venant equations

Consider a channel, the width of which decreases from x_1 to x_2 and increases from x_2 to x_3 (Figure 2.24). Assuming steady state, negligible friction and bottom slope, show that:

- 1) if the flow is subcritical ($u < c$) upstream of the narrowing and supercritical ($u > c$) downstream of it, critical conditions ($u = c$) can be reached only at the narrowest point, $x = x_2$,
- 2) if the flow is subcritical everywhere in the channel, the water depth reaches a minimum value at $x = x_2$,
- 3) if the flow is supercritical everywhere in the channel, the water depth reaches a maximum value at $x = x_2$.

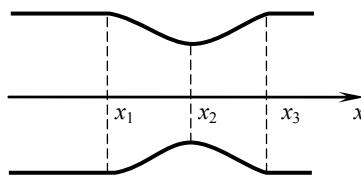


Figure 2.24. Free surface flow in a channel with a local section narrowing

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

2.7.2.5. Exercise 2.5: the Saint Venant equations

Consider a rectangular channel, the width and slope of which are denoted by b and S_0 respectively. The Strickler coefficient is assumed to be uniform. Steady, uniform flow is assumed, that is, the slope of the energy line is assumed to be identical to the slope of the bottom of the channel.

- 1) Provide the expression of the wave speed λ of the kinematic wave as a function of the water depth h . The wide channel approximation ($h \ll b$) will be assumed in the calculation of the hydraulic radius for the sake of simplicity.

2) Provide the expressions of the wave speeds $\lambda^{(1)}$ and $\lambda^{(2)}$ in the Saint Venant equations as a function of h , assuming that the assumption of a steady, uniform flow and the wide channel approximation remain valid (the assumption of a uniform flow allows the flow velocity u to be expressed as a function of h).

3) Compare the two expressions and plot the wave speeds as functions of h for the numerical values provided in Table 2.1. Conclude about the validity of the kinematic wave approximation in practical applications.

Symbol	Meaning	Value
b	Channel width	10 m
g	Gravitational acceleration	9.81 m/s ²
K_{Str}	Strickler coefficient	40
S_0	Channel bottom slope	0.1%, 1%, 5%

Table 2.1. Parameters for Exercise 2.5

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

2.7.2.6. Exercise 2.6: the Saint Venant equations

Consider a rectangular channel, the length and bottom slope of which are denoted by L and S_0 respectively. The elevation of the bottom at the left-hand end of the channel is denoted by z_L . The water is initially at rest. The elevation of the free surface is denoted by ζ_0 (Figure 2.25). The effect of friction is neglected and the perturbations in the free surface elevation are assumed to be small enough for the wave speeds to be considered independent of time. The numerical values of the physical parameters can be found in Table 2.2.

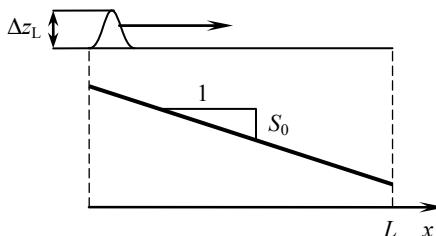


Figure 2.25. Propagation of a perturbation in a channel with constant bottom slope

At $t = 0$, a perturbation appears at the left-hand end of the channel. The height of the perturbation is denoted by Δz_L .

- 1) Provide the expressions of the resulting perturbations Δu_L and ΔQ_L in the velocity and in the discharge.
- 2) Provide a graphical representation of the characteristic along which the perturbation travels in the phase space. Provide the expression for the time T_R at which the perturbation reaches the right-hand end of the channel.
- 3) Compute the height Δz_R of the perturbation when it reaches the right-hand end of the channel, as well as the perturbations Δu_R and ΔQ_R in the velocity and in the discharge. What is your conclusion about the validity of the assumption that the wave speed does not depend on time?

Symbol	Meaning	Value
b	Channel width	10 m
g	Gravitational acceleration	9.81 m/s ²
L	Channel length	100 m
S_0	Channel bottom slope	10%
z_G	Elevation of the channel bottom at the left-hand end	0 m
Δz_G	Height of the perturbation at the left-hand end of the channel	0.1 m
ζ_0	Initial elevation of the free surface	1 m

Table 2.2. Parameters for Exercise 2.6

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

2.7.2.7. Exercise 2.7: the Euler equations

A loudspeaker may be schematized as a plane membrane of cross-sectional area A subjected to a displacement in the direction x normal to the plane (Figure 2.26).

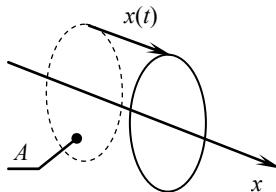


Figure 2.26. Definition sketch of a loudspeaker membrane

Both sides of the membrane are in contact with the ambient air (orifices on the rear side of the cabinet to allow for such a contact with the back side of the membrane). Assume that the movement $x(t)$ of the membrane can be described by a periodic, sinusoidal function of time in the form:

$$x(t) = a \cos(2\pi N t) \quad [2.225]$$

where a is the (constant) amplitude of the movement and N is the (constant) frequency of the sound signal.

- 1) Assuming a constant speed of sound c , provide the expression for the pressure as a function of time on both sides of the membrane.
- 2) Determine the average mechanical power needed to move the membrane over a period. Show that the power is proportional to the square of the frequency.
- 3) Carry out the numerical application for the parameter values in Table 2.3.

Chapter 3

Weak Solutions and their Properties

3.1. Appearance of discontinuous solutions

3.1.1. Governing mechanisms

As shown in section 1.4.3 with the example of the inviscid Burgers equation, initially continuous solutions may evolve into discontinuous solutions. This section focuses on the mechanisms that lead to the formation of discontinuities. Discontinuous solutions are an inevitable consequence of the nonlinearity of a hyperbolic conservation law, as shown hereafter.

Consider a scalar hyperbolic conservation law expressed in conservation form as:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} = 0 \quad [3.1]$$

This is the general form [1.1], where the source term is assumed to be zero. For the sake of simplicity, F is assumed to be a function of U only. As shown in Chapter 1, equation [3.1] can be written in the non-conservation form [1.28], recalled here:

$$\frac{\partial U}{\partial t} + \lambda \frac{\partial U}{\partial x} = 0$$

where $\lambda = dF/dU$. As shown in Chapter 1, equation [1.28] can be rewritten in the characteristic form [1.30], recalled here:

$$U = \text{Const} \quad \text{for } \frac{dx}{dt} = \lambda$$

U is constant along the characteristic lines $dx/dt = \lambda$. If F is a nonlinear function of U , the wave speed λ depends on the value of U . U being constant along a given characteristic, λ is also a constant along the characteristic and the characteristics are straight lines in the phase space (Figure 3.1).

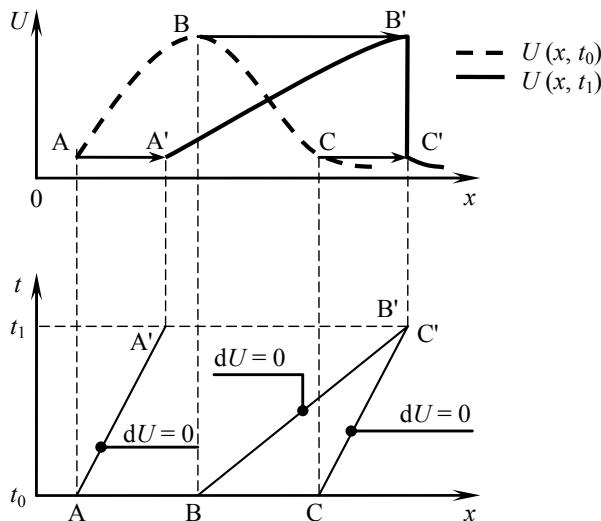


Figure 3.1. Continuous solution evolving into a discontinuous solution in the case of a convex flux function. Sketch in the physical space (top) and in the phase space (bottom)

Consider the case where the initial profile (ABC) at time t_0 is not monotonic. The maximum of U is reached at the point B.

If the flux function is convex, the characteristic issued from B moves faster than those issued from A and C because λ is an increasing function of U . The characteristic issued from B ‘catches up’ the characteristic issued from C at a time $t_1 > t_0$. At $t = t_1$, the points A, B and C move to A', B' and C'. U being constant along the characteristics, the values of U at A, B and C are identical to those at A', B' and C' respectively. Since B' and C' have the same abscissa, the profile of U is necessarily discontinuous because U simultaneously takes the value U_B and U_C at the same point.

If the flux function is concave, the characteristic issued from B is slower than those issued from A and C. Reasoning as in the paragraph above leads to the conclusion that a discontinuity appears at the point $A' = B'$.

A general formula can be derived for the time at which a discontinuity appears for the first time. This is achieved by deriving an expression for the space derivative $\partial U / \partial x$ from equation [1.28] and finding the date at which $\partial U / \partial x$ becomes infinite. Differentiating equation [1.28] with respect to x leads to:

$$\frac{\partial^2 U}{\partial x \partial t} + \frac{\partial}{\partial x} \left(\lambda \frac{\partial U}{\partial x} \right) = 0 \quad [3.2]$$

Expanding the space derivative and swapping the time and space differentials leads to:

$$\frac{\partial}{\partial t} \left(\frac{\partial U}{\partial x} \right) + \lambda \frac{\partial}{\partial x} \left(\frac{\partial U}{\partial x} \right) = - \frac{\partial \lambda}{\partial x} \frac{\partial U}{\partial x} \quad [3.3]$$

As shown in Chapter 1, equation [3.3] can be expressed in characteristic form as:

$$\frac{d}{dt} \left(\frac{\partial U}{\partial x} \right) = - \frac{\partial \lambda}{\partial x} \frac{\partial U}{\partial x} \quad \text{for } \frac{dx}{dt} = \lambda \quad [3.4]$$

Equation [3.4] is a first-order Ordinary Differential Equation (ODE) in $\partial U / \partial x$. Noting that $\partial \lambda / \partial x = \partial \lambda / \partial U \partial U / \partial x$, equation [3.4] is rewritten as:

$$\frac{d}{dt} \left(\frac{\partial U}{\partial x} \right) = - \frac{\partial \lambda}{\partial U} \left(\frac{\partial U}{\partial x} \right)^2 \quad \text{for } \frac{dx}{dt} = \lambda \quad [3.5]$$

Since $\partial \lambda / \partial U$ is a function of U only, U is constant along a characteristic line and $\partial \lambda / \partial U$ is also a constant along a characteristic line. ODE [3.5] has the following analytical solution:

$$\frac{\partial U}{\partial x}(t) = \left[\left(\frac{\partial U}{\partial x}(t_0) \right)^{-1} + (t - t_0) \frac{\partial \lambda}{\partial U} \right]^{-1} \quad \text{for } \frac{dx}{dt} = \lambda \quad [3.6]$$

Equation [3.6] describes the variations of $\partial U / \partial x$ as seen by an observer moving at a speed λ . $\partial U / \partial x$ becomes infinite if the quantity between the brackets in equation [3.6] becomes zero, which occurs at a time $t = t_1$ such that:

$$t_1 = t_0 - \left[\frac{\partial U}{\partial x}(t_0) \frac{d\lambda}{dU} \right]^{-1} = t_0 - \left[\frac{\partial \lambda}{\partial x}(t_0) \right]^{-1} \quad [3.7]$$

The time t_d at which the first discontinuity appears in the profile is given by the minimum of all the times t_1 associated with all the possible values of x :

$$t_d = \min_x \left\{ t_0 - \left[\frac{\partial \lambda}{\partial x}(x, t_0) \right]^{-1} \right\} = t_0 - \left\{ \min_x \left[\frac{\partial \lambda}{\partial x}(x, t_0) \right] \right\}^{-1} \quad [3.8]$$

Note that the profile can become discontinuous only if t_d is larger than t_0 , that is, if there exists at least one value of x for which the following condition is satisfied:

$$\frac{\partial \lambda}{\partial x}(x, t_0) \leq 0 \quad [3.9]$$

3.1.2. Local invalidity of the characteristic formulation – graphical approach

The conservation form and the characteristic formulation presented in Chapters 1 and 2 for scalar hyperbolic laws and hyperbolic systems of conservation laws were derived under the assumption that the derivatives of the variables and the fluxes are defined at all points of time and space. This assumption is not valid at discontinuities. The characteristic approach cannot be used in its classical form across discontinuities and a specific treatment must be applied. Such a treatment is detailed in section 3.4.

In the case of scalar hyperbolic conservation laws, the method of characteristics may be applied even to discontinuous solutions, provided that the method is modified using the so-called “equal area rule”. The equal area rule combines the properties of invariance and conservation with the necessary condition of solution uniqueness. It consists of the following two steps.

The first step consists of applying the original method of characteristics to the initial profile. The solution becomes discontinuous at the time t_1 when the characteristics issued from B and C intersect at $B' = C'$. Applying the method of characteristics at a time $t_2 > t_1$ leads to a multi-valued solution profile (A''B''C'') as sketched in Figure 3.2 because the point B' passes the point C'. Such a profile is not physically permissible in that U may take two or three different values at the abscissa x lying within the interval $[x_{B''}, x_{C''}]$.

The second step consists of correcting the profile [A"B"C"] so as to restore the uniqueness of the solution. The correction is made as follows:

- the corrected profile is discontinuous because the discontinuity appeared at $t_1 < t_2$;
- the correction should guarantee conservation. Consequently, the area under the corrected profile should be the same as the area under the profile before the correction.

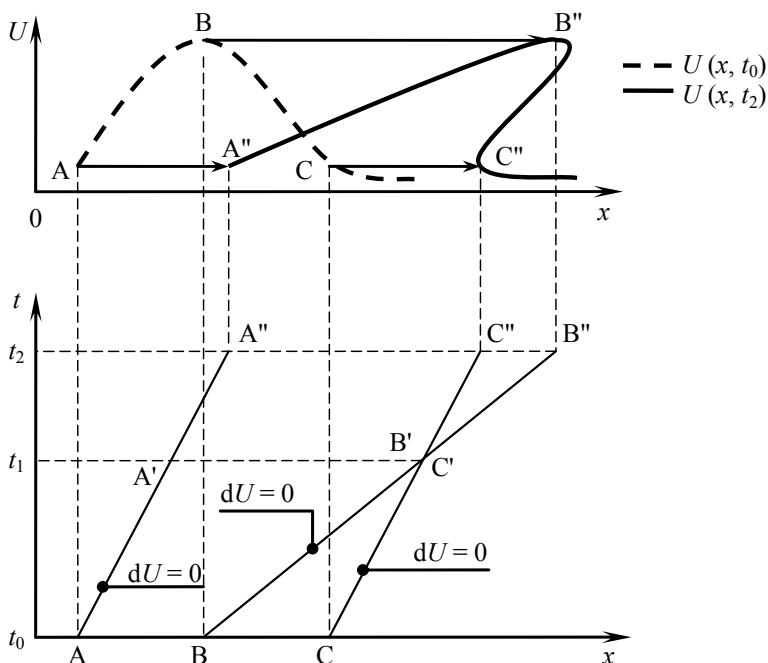


Figure 3.2. Using the method of characteristics beyond the time at which a discontinuity appears. Sketch in the physical space (top) and in the phase space (bottom)

The application of the correction to the profile [A"B"C"] in Figure 3.2 is illustrated by Figure 3.3. The gray-shaded areas on both sides of the discontinuity D are strictly equal, hence the term “equal area rule”. The correction is also reflected in the phase space (Figure 3.3, bottom), where the initial characteristics [B'C'] and [B'C''] are replaced with a bold line that represents the trajectory of the discontinuity.

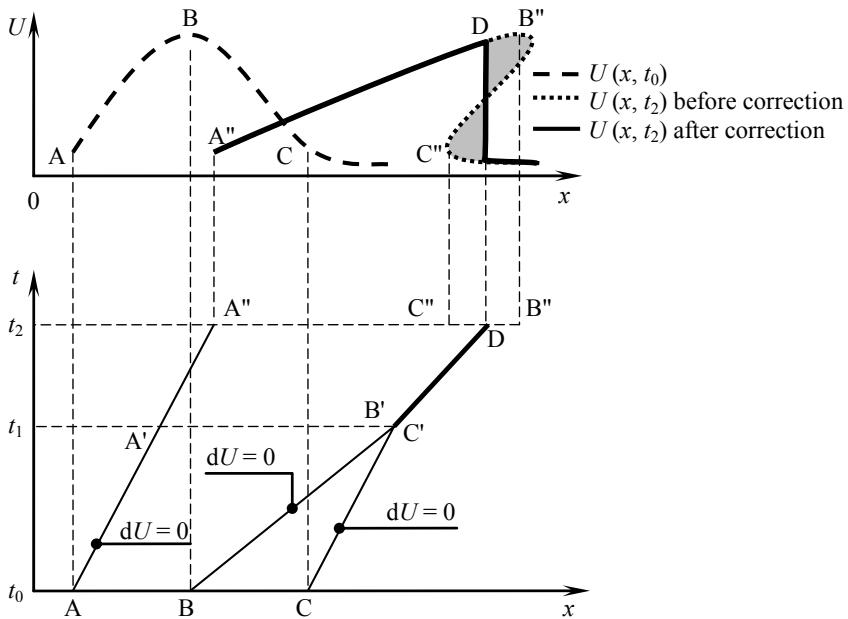


Figure 3.3. Application of the equal area rule to a multiple-valued solution. Representation in the physical space (top) and in the phase space (bottom)

3.1.3. Practical examples of discontinuous flows

3.1.3.1. Free surface flow: the breaking of a wave

The dependence of the wave speed on the flow variable is the main reason for the breaking of sea or ocean waves traveling to the shore. Figure 3.4 illustrates the behavior of a sea wave along a line drawn in the direction perpendicular to the shore. This direction may be seen as the longitudinal axis of a canal of infinite width, the bottom of which rises in the direction of positive x . Consider a wave traveling to the shore. The initial profile [ABC] of the free surface is continuous. The water depth h_B of the crest of the wave is larger than the depth h_C of the front. For a wave traveling over a mild beach slope, the depth h_A of the tail of the wave is larger than h_C and smaller than h_B . The average flow velocity u is small compared to the speed c of the waves in still water, with the consequence that the wave speed $u + c$ can be approximated reasonably with the speed $c = (gh)^{1/2}$. The wave speed being smaller in the regions where the flow is shallower, the characteristic lines $dx/dt = u + c$ are convex in the phase space (Figure 3.4, bottom).

The characteristic that passes at C is slower than the characteristic that passes at B, where the depth is larger. Consequently, the front side [BC] becomes steeper as the wave travels to the shore. Conversely, the characteristic issued from A is slower than that issued from B and the rear side [AB] of the wave becomes milder as time goes. After a certain time the point B' catches up the point C', the free surface becomes vertical and the wave breaks.

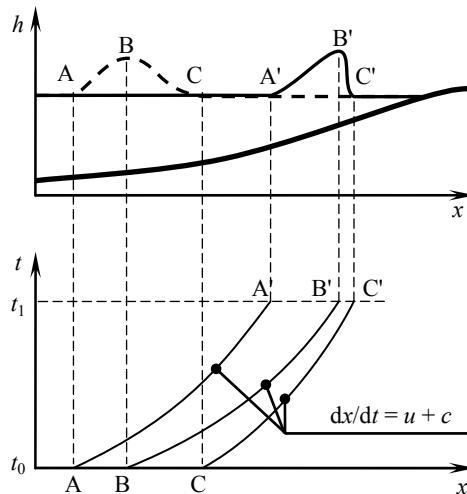


Figure 3.4. Breaking of a wave traveling to the shoreline.
Sketch in the physical space (top) and in the phase space (bottom)

NOTE.—Although the profile [A"B"C"] in Figure 3.2 is very similar to that of a breaking wave, the resemblance is purely coincidental. The profile [A"B"C"] is derived from a purely mathematical construction that does not account for the phenomena that govern the breaking of a wave. The breaking of the wave is a two-dimensional process in the vertical plane, while the construction in Figure 3.2 involves only one dimension of space.

3.1.3.2. Aerodynamics: supersonic flight

By definition, the speed of an airplane (or any other flying object) in supersonic flight is larger than the speed of sound. The gas molecules immediately in front of the airplane cannot move away fast enough and are “pushed” ahead and aside. The local accumulation of the gas molecules induces a rise in the pressure and in the density. The molecules travel to the zones of lower pressure and the thickness of the high pressure zone stabilizes to an equilibrium thickness (typically, millimeters to centimeters), adopting the shape of a V as does the wake of a ship. The gas ahead of

this V-shaped zone is undisturbed. The transition between the two zones is very thin (typically, millimeters). The thickness of the transition zone is negligible compared to the dimensions of the flying object and the pressure appears as discontinuous (Figure 3.5).

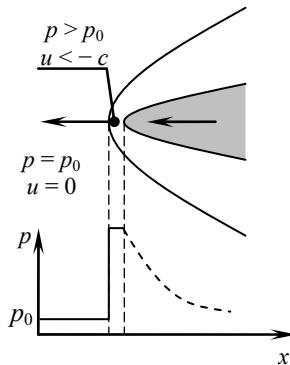


Figure 3.5. Shock wave created by a supersonic flying object.
Geometry of the wave in side view (top), pressure profile (bottom)

In practice, thermal diffusion and turbulence phenomena induce a widening of the transition zone. The pressure profile is not strictly discontinuous but may appear so at the metric scale. The pressure decreases gradually to the rear side of the plane and the initial pressure is recovered after a sufficiently long distance.

The well-known “supersonic bang” that can be heard when an airplane flies above the speed of sound is nothing but the consequence of the sudden pressure rise across the transition zone.

3.2. Classification of waves

3.2.1. Shock wave

A shock wave is characterized by a discontinuity in both the conserved variable and wave speed. It obeys the following criteria (Figure 3.6):

- Criterion (S1). The solution is discontinuous across the shock. The values U_L and U_R on the left- and right-hand sides of the shock are different.

- Criterion (S2). There is at least one wave (the p th wave), the speed of which is discontinuous across the shock, such that the wave speed on the left-hand side of the shock is larger than on the right-hand side and such that the propagation speed of the

discontinuity lies between these two wave speeds. The discontinuity is said to be a shock for the p th wave, or a p -shock.

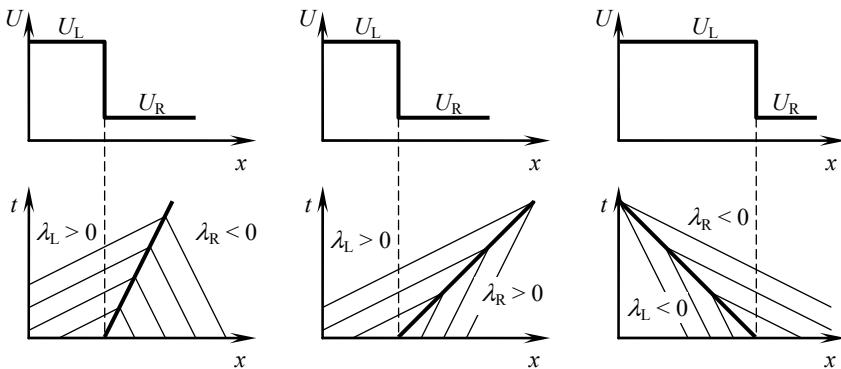


Figure 3.6. Definition sketch of a shock in the physical space (top) and in the phase space (bottom). Sketch for a scalar variable

When the law is scalar, there is only one wave. The criteria (S1–2) can be summarized as follows:

$$\left. \begin{array}{l} U_L \neq U_R \\ \lambda_L > c_s > \lambda_R \end{array} \right\} \quad [3.10]$$

where c_s is the speed of the shock and the subscripts L and R denote the values taken by U and λ on the left- and right-hand sides of the discontinuity respectively. When a hyperbolic system of conservation laws is dealt with, the wave speeds are numbered in ascending order. The discontinuity is a shock for the p th wave, or a p -shock, if the following conditions are satisfied:

$$\left. \begin{array}{l} U_L \neq U_R \\ \lambda_L^{(p-1)} < c_s < \lambda_L^{(p)} \\ \lambda_R^{(p)} < c_s < \lambda_R^{(p+1)} \end{array} \right\} \quad [3.11]$$

Note that the last two conditions [3.11] actually imply the inequality $\lambda_L^{(p)} > \lambda_R^{(p)}$. They also guarantee that the discontinuity is a shock neither for the wave $p - 1$ nor for the wave $p + 1$.

3.2.2. Rarefaction wave

A p -rarefaction wave (that is, a rarefaction for the p th wave) satisfies the following criteria (Figure 3.7):

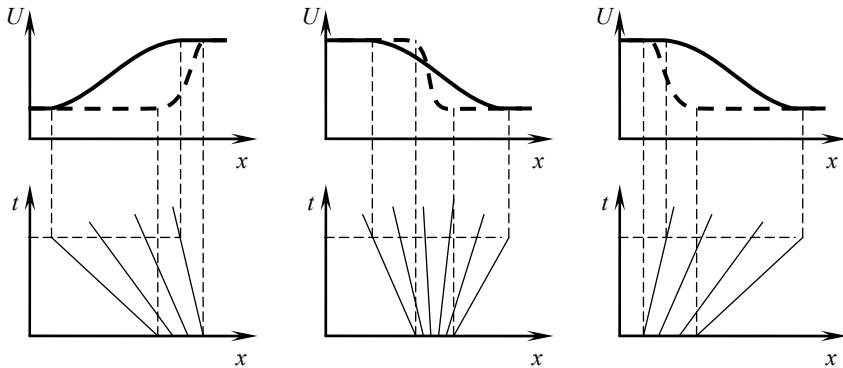


Figure 3.7. A rarefaction wave in the physical space (top) and in the phase space (bottom). Definition sketch for a scalar variable. Initial profile (dashed line), final profile (solid line)

- Criterion (R1). The variable U and the wave speeds vary continuously across the wave.
- Criterion (R2). The wave speed $\lambda^{(p)}$ increases from left to right across the wave

$$\frac{\partial \lambda^{(p)}}{\partial x} > 0 \quad [3.12]$$

Rarefaction waves cause front smearing and profile smoothing.

3.2.3. Contact discontinuity

The p th wave is a contact discontinuity if it satisfies the following criteria:

- Criterion (C1). The variable is discontinuous across the wave;
- Criterion (C2). The wave speed $\lambda^{(p)}$ is continuous across the wave;

$$\left. \begin{array}{l} U_L \neq U_R \\ \lambda_L^{(p)} = \lambda_R^{(p)} \end{array} \right\} \quad [3.13]$$

Figure 3.8 illustrates the behavior of a contact discontinuity in the physical space and in the phase space.

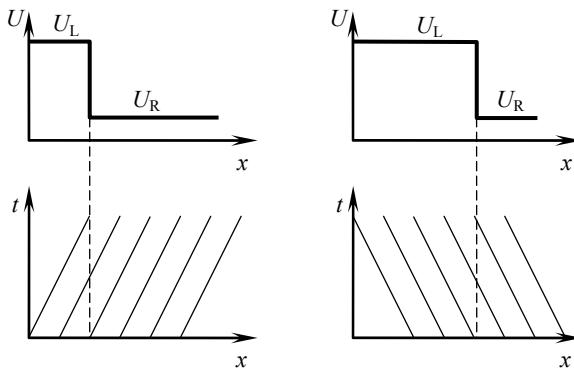


Figure 3.8. Definition sketch for a contact discontinuity in the physical space (top) and in the phase space (bottom)

3.2.4. Mixed/compound wave

Mixed waves, also known as compound waves, appear in very specific cases, such as non-convex flux functions. In such cases, a profile composed of a rarefaction wave and a shock may appear under certain combinations of initial and boundary conditions (see Chapter 4 for an example). A left-compound wave obeys the following criteria:

- Criterion (ML1). The celery $\lambda^{(p)}$ increases from left to right up to the abscissa x_s of the shock.
- Criterion (ML2). The conserved variable is discontinuous at $x = x_s$.
- Criterion (ML3). The wave speed on the right-hand side of the discontinuity is smaller than the wave speed on the left-hand side.

The criteria above can be summarized as:

$$\left. \begin{array}{l} U_L \neq U_R \\ \frac{\partial \lambda^{(p)}}{\partial x} > 0 \quad \text{for } x < x_s \\ \lambda_L^{(p)} > \lambda_R^{(p)} \end{array} \right\} [3.14]$$

Conversely, a right-compound wave obeys the following definitions:

- Criterion (MR1). The wave speed decreases from right to left down to the abscissa x_s of the shock.
- Criterion (MR2). The conserved variable is discontinuous across the shock.
- Criterion (MR3). The wave speed on the left-hand side of the discontinuity is larger than the wave speed on the right-hand side.

The criteria above can be summarized as:

$$\left. \begin{array}{l} U_L \neq U_R \\ \frac{\partial \lambda^{(p)}}{\partial x} > 0 \quad \text{for } x > x_s \\ \lambda_L^{(p)} > \lambda_R^{(p)} \end{array} \right\} [3.15]$$

3.3. Simple waves

3.3.1. Definition and properties

Consider an $m \times m$ hyperbolic system of conservation laws. By definition, its m wave speeds are all different. The p th wave is a simple wave if the conserved variable U is constant along the characteristics $dx/dt = \lambda^{(p)}$. By definition, the characteristic curve for a simple wave is a straight line in the phase space (see Figure 3.9) because the wave speed, that is a function of U , is also constant.

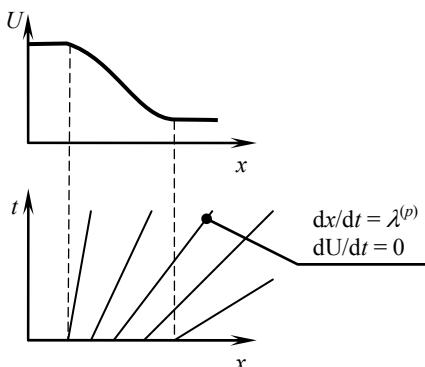


Figure 3.9. Definition sketch for a simple wave in the physical space (top) and in the phase space (bottom)

3.3.2. Generalized Riemann invariants

Generalized Riemann invariants are differential relationships that apply across simple waves. In contrast with Riemann invariants that may be used across any type of wave, generalized Riemann invariants can be used only across simple waves.

Assume that the p th wave is a simple wave (Figure 3.10). Then the characteristics $dx/dt = \lambda^{(p)}$ are straight lines in the phase space. Consider two such characteristics close to each other in the phase space. The leftmost characteristic is denoted by (A), the rightmost characteristic is denoted by B. In the general case, (A) and (B) are not parallel because the values of U along (A) and (B) are not identical. Since the system is hyperbolic, the wave speeds of all the remaining characteristics are different from $\lambda^{(p)}$. Consequently, the characteristics (A) and (B) can be connected using any of the remaining $m - 1$ characteristics.

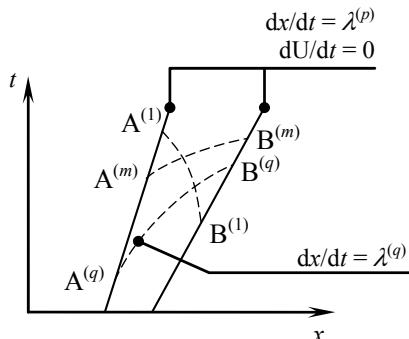


Figure 3.10. Two neighbor characteristics in the simple wave p .
Definition sketch in the phase space

The variation dU in U between the characteristics (A) and (B) is given by (see equation [2.27]):

$$dU = K \, dW = \begin{bmatrix} K_1^{(1)} & \dots & K_j^{(1)} & \dots & K_m^{(1)} \\ \vdots & & & & \vdots \\ K_i^{(1)} & \dots & K_j^{(1)} & \dots & K_i^{(m)} \\ \vdots & & & & \vdots \\ K_m^{(1)} & \dots & K_j^{(1)} & \dots & K_m^{(m)} \end{bmatrix} \begin{bmatrix} dW_1 \\ \vdots \\ dW_i \\ \vdots \\ dW_m \end{bmatrix} \quad [3.16]$$

By definition, all the Riemann invariants W_q , $q \neq p$, are constant between (A) and (B):

$$dW_q = 0 \quad \forall q \neq p \quad [3.17]$$

Consequently, the only non-constant invariant between (A) and (B) is the p th Riemann invariant. Substituting equation [3.17] into equation [3.16] leads to:

$$dU_k = K_k^{(p)} dW_p \quad [3.18]$$

In other words, the vector dU is collinear to the p th eigenvector $K^{(p)}$ across the p th wave.

$$\frac{dU_1}{K_1^{(p)}} = \frac{dU_2}{K_2^{(p)}} = \dots = \frac{dU_m}{K_m^{(p)}} \quad \text{across } \frac{dx}{dt} = \lambda^{(p)} \quad [3.19]$$

The relationships in equation [3.19] are called generalized Riemann invariants. They form a system of $m - 1$ equations that may be used across the p th wave to characterize the properties of the solution. They may be applied to specific problems such as the Riemann problem (see Chapter 4), the solution of which is made of simple waves.

Note however that the generalized Riemann invariants are meaningful only if the solution is continuous across the wave. The generalized Riemann invariants cannot be applied across shock waves. Jump relationships that are detailed in section 3.4.3 should be used instead.

3.4. Weak solutions and their properties

3.4.1. Definitions

Equation [1.1] can be rewritten as:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} - S = 0 \quad [3.20]$$

The weak form of equation [3.20] over a domain $[x_1, x_2] \times [t_1, t_2]$ is obtained by multiplying equation [3.20] by a function $w(x, t)$, also known as a weighting function, and by integrating the equation over the domain:

$$\int_{t_1}^{t_2} \int_{x_1}^{x_2} \left(\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} - S \right) w(x, t) dx dt = 0 \quad [3.21]$$

A solution of the weak form [3.21] is called a weak solution of [3.20]. The weak form of a vector equation in the form [2.2] is defined exactly in the same way as that of a scalar equation.

In the particular case where $w(x, t)$ is a constant, equation [3.21] can be simplified into:

$$\int_{t_1}^{t_2} \int_{x_1}^{x_2} \left(\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} - S \right) dx dt = 0 \quad [3.22]$$

Integrating $\partial U / \partial t$ with respect to time and $\partial F / \partial x$ with respect to x leads to:

$$\begin{aligned} & \int_{x_1}^{x_2} U(x, t_2) dx - \int_{x_1}^{x_2} U(x, t_1) dx + \int_{t_1}^{t_2} F(x_2, t) dt - \int_{t_1}^{t_2} F(x_1, t) dt \\ &= \int_{t_1}^{t_2} \int_{x_1}^{x_2} S(x, t) dx dt \end{aligned} \quad [3.23]$$

Note that equation [3.23] is strictly equivalent to the balance [1.11–12] over the control volume $[t_0, t_0 + \delta t] \times [x_0, x_0 + \delta x]$ if $x_1 = x_0$, $x_2 = x_0 + \delta x$, $t_1 = t_0$, $t_2 = t_0 + \delta t$.

3.4.2. Non-equivalence between the formulations

Although closely connected together, the forms [3.20] and [3.23] are not strictly equivalent. They differ by two important points:

- As shown in section 1.1.2, equation [3.20] is derived from equation [3.23] by assuming that the size of the integration domain tends to zero. This allows the derivatives $\partial U / \partial t$ and $\partial F / \partial x$ to be introduced. This implies that U (and therefore F) is continuous and differentiable with respect to time and space. The form [3.20] does not account for discontinuous solutions such as shocks and compound waves.

- The assumption of continuous and differentiable solutions is not needed in the form [3.23] because the integrals in equation [3.23] can be calculated even if the solution is discontinuous in time and/or space.

In other words, a “strong solution” of equation [3.20] (that is, a solution that verifies [3.20] for all x and t) is a particular case of a weak solution, while the reciprocal is not true. The “strong form” [3.20] and the weak form [3.23] are equivalent as long as the solution is continuous in time and space. If the solution is discontinuous, equations [3.20] and [3.23] cease to be equivalent. This is of primary

importance in the solution of hyperbolic PDEs with discontinuous solutions (see section 3.4.4).

3.4.3. Jump relationships

As shown in section 3.1.2, the characteristic form of the equation is based on the implicit assumption that the solution is continuous. It cannot be applied across discontinuities. An alternative technique is needed for the treatment of discontinuous solutions. Jump relationships, also known as “Rankin-Hugoniot relationships”, are derived from a balance over a control volume that contains the discontinuity (Figure 3.11). Equation [3.23] is applied to the control volume in the limit of an infinitesimal volume width and time interval.

Consider first that the conservation law is scalar. Denoting by c_s the speed of the discontinuity, the variation between the times t_1 and t_2 in the total amount of U contained in the control volume is given by:

$$\int_{x_1}^{x_2} U(x, t_2) dx - \int_{x_1}^{x_2} U(x, t_1) dx = (t_2 - t_1)(U_1 - U_2)c_s \quad [3.24]$$

where U_1 and U_2 are respectively the values of U on the left- and right-hand side of the discontinuity.

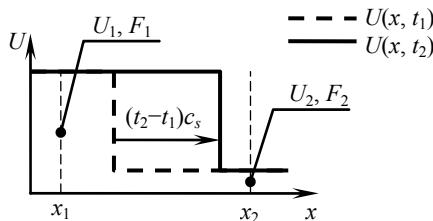


Figure 3.11. Definition sketch for the Rankin-Hugoniot relationships

The amount of U that crosses the boundaries of the control volume between t_1 and t_2 is given by:

$$\int_{t_1}^{t_2} F(x_2, t) dt - \int_{t_1}^{t_2} F(x_1, t) dt = (F_1 - F_2)(t_2 - t_1) \quad [3.25]$$

where F_1 and F_2 denote $F(U_1)$ and $F(U_2)$ respectively. The integral of the source term is:

$$\int_{t_1}^{t_2} \int_{x_1}^{x_2} S(x, t) dx dt = (t_2 - t_1)(x_2 - x_1)\bar{S} \quad [3.26]$$

where \bar{S} is the average of S over the space-time domain $[x_1, x_2] \times [t_1, t_2]$. Substituting equations [3.24–26] into equation [3.23] and dividing by $(t_2 - t_1)$ yields:

$$(U_1 - U_2)c_s + F_2 - F_1 + (x_2 - x_1)\bar{S} = 0 \quad [3.27]$$

When the width of the control volume tends to zero, the quantity $(x_2 - x_1)\bar{S}$ tends to zero and equation [3.27] becomes:

$$(U_1 - U_2)c_s = F_1 - F_2 \quad [3.28]$$

Equation [3.28] is generalized to hyperbolic systems of conservation laws by noticing that it is applicable to each of the components of the vectors U and F individually. The vector form of equation [3.28] is therefore:

$$(U_1 - U_2)c_s = F_1 - F_2 \quad [3.29]$$

where F_1 and F_2 denote $F(U_1)$ and $F(U_2)$ respectively. Equations [3.28–29] may be used to determine the speed of the discontinuity. Note that a stationary shock (i.e. a shock that does not move) satisfies the following conditions:

$$\left. \begin{array}{l} F_1 = F_2 \quad (\text{scalar law}) \\ F_1 = F_2 \quad (\text{hyperbolic system}) \end{array} \right\} \quad [3.30]$$

Also note that when the amplitude of the shock tends to zero, the shock speed c_s tends to the wave speed λ . Indeed, equation [3.28] leads to the following equivalence:

$$c_s = \frac{F_2 - F_1}{U_2 - U_1} \underset{U_2 \rightarrow U_1}{\approx} \frac{\partial F}{\partial U} = \lambda \quad [3.31]$$

3.4.4. Non-uniqueness of weak solutions

3.4.4.1. Example 1. The inviscid Burgers equation

The weak and strong forms of hyperbolic equations not being equivalent, a nonlinear PDE may have several weak solutions, each of which is mathematically permissible. Only physical considerations allow the “correct” solution to be identified from the many possible ones. This is illustrated by the inviscid Burgers equation.

The inviscid Burgers equation derived in section 1.4.2 can be written in non-conservation form as in equation [1.66], recalled here:

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = 0$$

A possible conservation form of this equation is equation [1.69], recalled here:

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} \left(\frac{u^2}{2} \right) = 0$$

In this form the conserved variable is $U = u$ and the flux function is $F = U^2/2$. Applying the jump relationship [3.28] leads to the following formula for c_s :

$$c_s = \frac{1}{2} \frac{U_1^2 - U_2^2}{U_1 - U_2} = \frac{U_1 + U_2}{2} = \frac{u_1 + u_2}{2} \quad [3.32]$$

However, equation [3.32] stems directly from the choice of the conserved variable. Other choices could be made, such as:

$$u = V^2 \quad [3.33]$$

Substituting definition [3.33] into the non-conservation form [1.66] leads to:

$$\frac{\partial}{\partial t} (V^2) + V^2 \frac{\partial}{\partial x} (V^2) = 0 \quad [3.34]$$

Simplifying yields the following non-conservation form in V :

$$\frac{\partial V}{\partial t} + V^2 \frac{\partial V}{\partial x} = 0 \quad [3.35]$$

Equation [3.35] is equivalent to equation [1.66] because the speed λ of the wave in V is equal to V^2 , that is to u , exactly as in the original equation. A continuous solution of equation [3.35] behaves exactly as a continuous solution of equation [1.66]. Differences arise when discontinuous solutions are considered. The conservation form of equation [3.35] is:

$$\frac{\partial V}{\partial t} + \frac{\partial}{\partial x} \left(\frac{V^3}{3} \right) = 0 \quad [3.36]$$

where the conserved variable is V and the flux function is $F = V^3/3$. Using the relationship [3.28] leads to the following expression for the shock speed:

$$c_s = \frac{F_1 - F_2}{V_1 - V_2} = \frac{V_1^2 + V_1 V_2 + V_2^2}{3} = \frac{u_1 + (u_1 u_2)^{1/2} + u_2}{3} \quad [3.37]$$

When u_1 and u_2 tend identically to a fixed value u , equations [3.32] and [3.37] tend to the same wave speed $\lambda = u$. When the solution is discontinuous however, equations [3.32] and [3.37] give different results.

There are two reasons for this:

- Equations [1.66] and [3.32] are transformed into equations [1.69] and [3.37] respectively under the assumption that the derivatives of u are defined everywhere. This is not true when the solution is discontinuous.
- The conserved variable is not the same in equation [1.69] as in equation [3.37]. Equation [1.69] is based on the implicit assumption that u is the conserved variable. Equation [3.37] is based on the implicit assumption that $V = u^{1/2}$ is the conserved variable.

The example of equations [3.32] and [3.37] shows that the solutions of equation [1.66] are not unique. Only a proper choice of the conserved variable allows the uniqueness of the solution to be ensured. It is the modeler's responsibility to define the conserved variable on the basis of physical considerations. This can be done only based on the analysis of the physical process involved, including in the case where the solutions become discontinuous.

3.4.4.2. Example 2. The hydraulic jump

Hydraulics specialists usually derive the steady-state, open channel flow equations using the concept of energy, also known as the hydraulic head. Such equations, however, may be obtained directly from the momentum equations, with

the advantage that they remain valid even when the flow becomes discontinuous, which is not the case with the equation of energy. The equations for steady-state, open channel flow in a rectangular prismatic channel can be written as:

$$\left. \begin{aligned} \frac{\partial Q}{\partial x} &= 0 \\ \frac{\partial}{\partial x} \left(\frac{Q^2}{A} + \frac{P}{\rho} \right) &= (S_0 - S_f)gA \end{aligned} \right\} \quad [3.38]$$

Substituting the continuity equation into the momentum equation leads to:

$$Q^2 \frac{\partial}{\partial x} \left(\frac{1}{A} \right) + \frac{\partial}{\partial x} \left(\frac{P}{\rho} \right) = (S_0 - S_f)gA \quad [3.39]$$

By definition, $\partial(P/\rho)/\partial x = c^2 \partial A/\partial x$ and $\partial A/\partial x = b \partial h/\partial x$. Equation [3.39] can be rewritten as:

$$(c^2 - u^2)b \frac{\partial h}{\partial x} = (S_0 - S_f)gA \quad [3.40]$$

Noting that $c^2 = gA/b$ and using the Froude number $\text{Fr} = u/c$, equation [3.40] becomes:

$$\frac{\partial h}{\partial x} = \frac{S_0 - S_f}{1 - \text{Fr}^2} \quad [3.41]$$

A more classical approach, used in most textbooks, consists of defining the hydraulic head H as the ratio of the total energy of the fluid to the product ρg :

$$H = \zeta + \frac{u^2}{2g} = h + \frac{u^2}{2g} + z_b \quad [3.42]$$

and stating that the head loss is due to the work carried out by the friction forces:

$$\frac{\partial H}{\partial x} = -S_f \quad [3.43]$$

Substituting equation [3.42] into equation [3.43] leads to the following expression:

$$\frac{\partial h}{\partial x} + \frac{1}{2g} \frac{\partial}{\partial x} (u^2) - S_0 = -S_f \quad [3.44]$$

Noting that $u = Q/A$ and that Q is constant, equation [3.44] can be rewritten as:

$$\frac{\partial h}{\partial x} + \frac{Q^2}{2g} \frac{\partial}{\partial x} \left(\frac{1}{A^2} \right) = S_0 - S_f \quad [3.45]$$

Using the differential $dA = b dh$ again, equation [3.45] leads to:

$$\frac{\partial h}{\partial x} - \frac{Q^2 b}{2g A^3} \frac{\partial h}{\partial x} = S_0 - S_f \quad [3.46]$$

Noting that $u = Q/A$ and introducing the Froude number, equation [3.46] is easily shown to be equivalent to equation [3.41]. Consequently, stating the conservation of momentum is equivalent to stating the conservation of energy in the continuous case.

Consider now a hydraulic jump in a rectangular channel of width b (Figure 3.12).

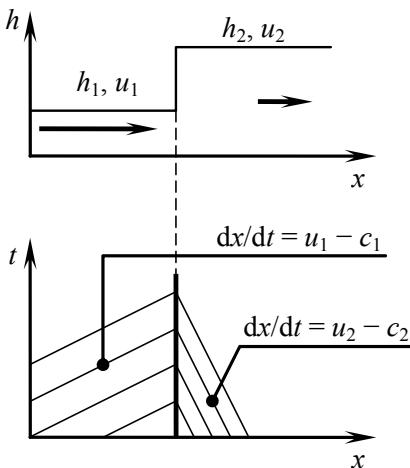


Figure 3.12. Stationary hydraulic jump. Definition sketch in the physical space (top) and in the phase space for the characteristic $u - c$ (bottom)

A hydraulic jump is a stationary shock, the flow upstream of which is supercritical and subcritical downstream. Note that the depth h_1 upstream of the jump is necessarily smaller than the depth h_2 downstream of the jump because the jump is a shock for the characteristic $dx/dt = u - c$. For a stationary jump, equation [3.30] leads to:

$$\left. \begin{aligned} Q_1 &= Q_2 = Q \\ \frac{Q^2}{bh_1} + bg \frac{h_1^2}{2} &= \frac{Q^2}{bh_2} + bg \frac{h_2^2}{2} \end{aligned} \right\} \quad [3.47]$$

These two equations state the conservation of mass and momentum across the shock. Note that this is not equivalent to stating the conservation of energy. If this was the case, the head would be identical on both sides of the jump and the following relationship would hold:

$$h_1 + \frac{Q^2}{2gb^2} \frac{1}{h_1^2} = h_2 + \frac{Q^2}{2gb^2} \frac{1}{h_2^2} \quad [3.48]$$

Equation [3.48] is not equivalent to the second relation [3.47]. Using equation [3.47], the head loss ΔH across the shock can be shown to be:

$$\Delta H = H_1 - H_2 = \frac{(h_2 - h_1)^3}{4h_1 h_2} \quad [3.49]$$

ΔH is always positive because $h_1 < h_2$ by definition. The head loss expresses the fact that part of the mechanical energy of the fluid is dissipated across the jump. The dissipation takes the form of a heat transfer to the fluid.

As in the example of the inviscid Burgers equation, the principle of conservation of momentum and energy are equivalent as long as the flow variables remain continuous. As soon as the solution becomes discontinuous, the conservation of momentum and the conservation of mechanical energy cease to be equivalent.

3.4.5. The entropy condition

The entropy condition allows mathematically permissible solutions that are not satisfactory from a physical point of view to be eliminated. It is also used to ensure the uniqueness of solutions of initial value problems such as the Riemann problem (see Chapter 4) [LIU 75, LIU 76]. The entropy condition is based on the following

consideration. The jump relationship [3.28] allows for the existence of “rarefaction shocks”, that is, solutions that satisfy the criterion (S1) in section 3.2.1 and do not satisfy the criterion (S2). Such “rarefaction shocks” would verify the following conditions:

$$\left. \begin{array}{l} U_L \neq U_R \\ \lambda_L^{(p)} < \lambda_R^{(p)} \quad \forall p = 1, \dots, m \end{array} \right\} \quad [3.50]$$

As illustrated by Figure 3.13, the characteristics converge to a shock in the phase space (Figure 3.13a), while they diverge from a “rarefaction shock” (Figure 3.13b).

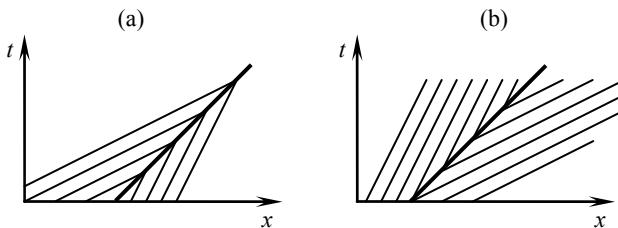


Figure 3.13. Two mathematically permissible, discontinuous solutions: shock (a), “rarefaction shock” (b)

The entropy principle states that “rarefaction shocks” are not physically permissible, because a discontinuous solution with wave speeds on the left-hand side of the discontinuity smaller than those on the right-hand side is not permissible. The term “entropy principle” was introduced by Courant and Friedrichs [COU 48] in their study of the Euler equations (see section 2.6). The entropy may be seen as an analog for aerodynamics of the mechanical energy, or hydraulic head, used in open channel hydraulics (the hydraulic head may be used as an entropy function for the Saint Venant or shallow water equations). As shown in section 3.4.4.2, the hydraulic head is not conserved across a shock. In a similar fashion, entropy always increases when a shock is passed in the direction of the flow.

The entropy principle may be justified as follows. A discontinuous solution may be viewed as the limit case of a continuous profile, where both sides of the discontinuity are connected to each other within a very short distance ε (see Figure 3.14). If the wave speed on the left-hand side of the discontinuity is larger than the wave speed on the right-hand side of the discontinuity, the profile becomes steeper and a discontinuity appears (Figure 3.14a). Conversely, if the wave speed on the left-hand side of the discontinuity is smaller than the wave speed on the right-hand side, a rarefaction wave appears and the profile becomes smoother (Figure 3.14b).

A shock may be seen as a self-stabilizing wave pattern in that any local smoothing of the profile (due for example to the presence of an additional source term, etc.) is automatically eliminated because the wave speed on the left-hand side of the shock is larger than on the right-hand side and the solution remains discontinuous. In contrast, a “rarefaction shock” is not a self-stabilizing wave pattern because if the profile becomes locally smooth for some reason, the difference between the wave speed on both sides of the discontinuity leads to a smoothing of the profile, thus destroying the discontinuous character of the solution.

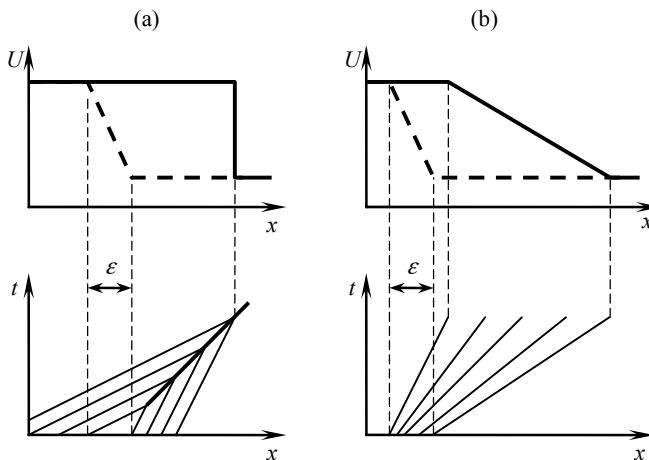


Figure 3.14. Discontinuous profile seen as a limit of a continuous profile.
Initial profile (dashed line), final profile (solid line) for a physically permissible shock (left)
and for a physically non-permissible shock (right)

3.4.6. Irreversibility

A salient feature of weak solutions is that their behavior is not reversible in time. In fact, two different initial conditions may lead to the same discontinuous solution, as shown in the example hereafter.

Consider the conservation form [1.69] of the inviscid Burgers equation. The initial condition is given by (see Figure 3.15a):

$$u(x,0) = \begin{cases} u_1 & \text{for } x \leq x_1 \\ u_1 \frac{x_2 - x}{x_2 - x_1} + u_2 \frac{x - x_1}{x_2 - x_1} & \text{for } x_1 \leq x \leq x_2 \\ u_2 & \text{for } x \geq x_2 \end{cases} \quad [3.51]$$

where $x_1 < x_2$ and $u_1 > u_2 > 0$. The initial condition takes the form of a ramp that connects the constant states u_1 and u_2 linearly between the abscissas x_1 and x_2 . Since $u_1 > u_2$ the left-hand part of the profile travels faster than the right-hand part.

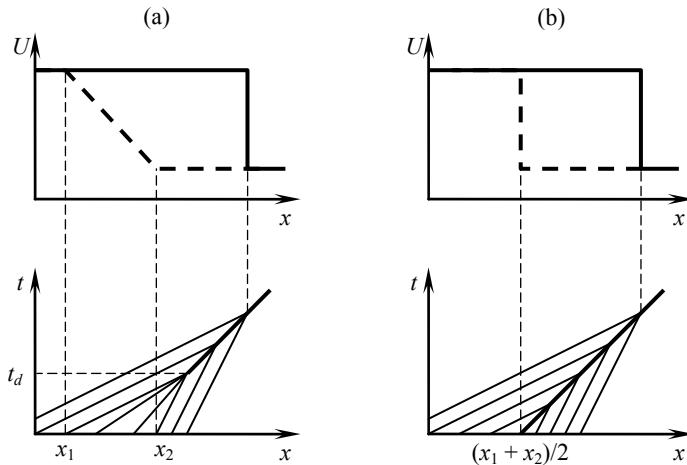


Figure 3.15. Two different initial conditions (dashed lines) leading to identical solutions at $t > t_d$ (solid lines)

The profile becomes steeper and becomes discontinuous at a time t_d :

$$t_d = \frac{x_2 - x_1}{u_1 - u_2} \quad [3.52]$$

At $t = t_d$ the discontinuity is located at $x = x_d$:

$$x_d = x_1 + u_1 t_d = x_2 + u_2 t_d = \frac{u_1 x_2 - u_2 x_1}{u_1 - u_2} \quad [3.53]$$

At further times the discontinuity propagates at a speed given by the average between u_1 and u_2 (see equation [3.31]). Note however that another initial condition may be defined (Figure 3.15b):

$$u(x,0) = \begin{cases} u_1 & \text{for } x < \frac{x_1 + x_2}{2} \\ u_2 & \text{for } x > \frac{x_1 + x_2}{2} \end{cases} \quad [3.54]$$

It is easy to check that the initial conditions [3.51] and [3.54] yield exactly the same solution for $t > t_d$.

In other words, different initial conditions may lead to the same final discontinuous solution. Consequently, a discontinuous solution may not be used as a starting point to “travel backwards in time” and calculate the solution at earlier times. The irreversible behavior of the solutions stems directly from the nonlinear character of the equations that makes discontinuous solutions possible.

3.4.7. Approximations for the jump relationships

This section gives two approximations for the jump relationships. Such approximations have been used by a number of authors in the development of numerical methods for the calculation of discontinuous solutions. Any reader interested in the details of the proof may find it useful to refer to [COU 48] and [LAX 57].

Theorem 1. The speed of a shock for the p th wave in a convex conservation law can be approximated with the arithmetic mean of the wave speeds $\lambda^{(p)}$ on both sides of the shock. The approximation is second-order with respect to the variation ΔU across the shock.

Theorem 2. The variation in the p th generalized Riemann invariant across a p -shock is of third order with respect to the variation ΔU across the shock.

Theorem 1 is best illustrated by the application to the inviscid Burgers equation (see equations [3.31] and [3.37]). Equation [3.31] is applicable if u is defined as the conserved variable. Equation [3.37] applies if $u^{1/2}$ is defined as the conserved variable. It is easy to check that equation [3.37] is a second-order approximation of equation [3.31].

Theorem 2 is useful when discontinuous solutions are to be calculated. In fact, its direct implication is that the Riemann invariants provide reasonably accurate approximations of the Rankin-Hugoniot relationships. Such a property has been used to derive approximate solvers for the Riemann problem covered in Chapter 4. The Riemann problem serves as a basis for a number of numerical techniques for the solution of hyperbolic systems of conservation laws with discontinuous solutions.

3.5. Summary

3.5.1. *What you should remember*

Three main types of wave may be distinguished: shock waves (see section 3.2.1), rarefaction waves (see section 3.2.2) and contact discontinuities (see section 3.2.3). Compound waves may appear when the flux function is non-convex. A compound wave is formed by the conjunction of a shock and a rarefaction wave.

A simple wave is a wave along the characteristics of which the conserved variable is a constant. In an $m \times m$ hyperbolic system, the generalized Riemann invariants provide $m - 1$ differential relationships across simple waves.

When the flux function is nonlinear, discontinuous solutions may arise from initially continuous profiles. This is because the dependence of the wave speed on the value of the conserved variable induces a deformation in the solution profile.

A discontinuous solution of a hyperbolic conservation law is called a weak solution because it is the solution of the weak form [3.32] of the original equation [3.1]. Both formulations are non-equivalent. The “strong form” is a particular case of the weak form under the assumption of continuous and differentiable solutions.

Weak solutions may be discontinuous. They are not unique. The “correct” weak solution of a conservation law must be chosen on the basis of physical considerations, in the light of the physical processes involved that allow the conserved variable to be identified.

The behavior of weak solutions is irreversible in time. Several initial conditions may lead to the same discontinuous solution. Consequently, inverse modeling (that is, retrieving the initial condition from the solution at a later time) cannot be carried out in a straightforward manner in the presence of weak solutions.

The characteristic form of the equations, that is based on the assumption of continuous and differentiable variables, is not applicable across discontinuities. The equal area rule allows weak solutions to be calculated using the method of characteristics in the scalar case. In the general case, the jump relationships [3.28–29], also called the Rankin-Hugoniot relationships, must be used.

The main two types of discontinuity in a solution are shocks and contact discontinuities. The wave speed on the left-hand side of the shock is always larger than on the right-hand side, while they are identical in the case of a contact discontinuity.

The entropy principle states that “rarefaction shocks”, the wave speed on the left-hand side of which would be smaller than on the right-hand side, are not physically permissible, even though they satisfy the jump relationship, thus being mathematically permissible.

The Riemann invariants may be viewed as an approximation of the jump relationships across shocks of small amplitude.

3.5.2. Application exercises

3.5.2.1. Exercise 3.1: the kinematic wave equation

Consider the rectangular channel used in Exercise 2.5 (see section 2.7.2.5). The flow is assumed to obey Strickler’s friction law [1.81]. Steady state is assumed.

1) The initial water depth is uniformly equal to $h_0 = 1$ m. Assuming that the wide channel approximation is applicable, compute the initial discharge into the channel under the assumption of a uniform, steady flow ($S_0 = S_f$). Provide the expression of the wave speed for the kinematic wave. Carry out the numerical application for the parameters in Table 2.1.

2) A perturbation $\Delta h = 0.5$ m appears instantaneously at the upstream end of the channel. Show that a shock wave appears. Provide the expression of the propagation speed of the shock wave. Carry out the numerical application for the parameters in Table 2.1.

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

3.5.2.2. Exercise 3.2: the kinematic wave equation

Consider the channel of Exercise 3.1, with the same geometry and initial conditions. The water depth at the upstream end of the channel is now assumed to increase linearly from 1 m to 1.25 m between $t = 0$ and $t = 100$ s, and to decrease linearly from 1.25 m to 1 m between $t = 100$ s and $t = 200$ s.

1) Assuming that the kinematic wave approximation is applicable, provide the expression of the time t_d at which the solution becomes discontinuous. Compute t_d and the location of the shock at $t = t_d$ from the parameters in Table 2.1.

2) Plot the water level profile at $t = 150$ s, 300 s, 450 s and 600 s. *N.B.:* it is advised to express both h and x as functions of the time t_L at which the characteristic leaves the left-hand end of the channel.

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

3.5.2.3. Exercise 3.3: the Buckley-Leverett equation

Consider an aquifer, the characteristics of which are given in Table 1.5 in Exercise 1.4 (see section 1.8.2.4). The aquifer is now assumed to be uniformly contaminated with an initial hydrocarbon saturation of 90% (i.e. the initial water saturation is assumed to be 10% everywhere). As in Exercise 1.4, the aquifer is decontaminated by injecting pure water with a Darcy velocity V at the left-hand end of the domain.

- 1) Show that the saturation profile at $t > 0$ is a compound wave.
- 2) Compute the propagation speed of the shock.
- 3) Compute the time at which the average contamination (i.e. the average hydrocarbon saturation) in the aquifer is 5%, 1% and 0.5%.

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

3.5.2.4. Exercise 3.4: the Saint Venant equations

Consider the channel of Exercise 3.1, where the Saint Venant equations are to be applied instead of the kinematic wave approximation.

- 1) The initial water depth is assumed to be uniformly equal to 1 m. Compute the speeds of the waves for the hydraulic parameters given in Table 2.1. Show that the flow regime depends on the slope. Provide the expression of the slope S_c for which the flow is critical.
- 2) A perturbation $\Delta h = 1$ m in the water level appears instantaneously at the upstream end of the channel. This triggers a moving bore that propagates to the right. Assuming that the flow regime is subcritical, provide the expression satisfied by the variation ΔQ in the discharge. Carry out the numerical approximation for $S_0 = 10^{-3}$.

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

3.5.2.5. Exercise 3.5: the Euler equations

An airplane moves at Mach 1 in immobile air. For the sake of simplicity, the coordinate system is attached to the airplane.

- 1) Write the continuity equation and the momentum equation under the assumption of steady state. The flow velocity is assumed to be zero on the hull of the airplane. Show that the assumption of a steady state flow necessarily induces a

multidimensional flow pattern and that the air must be “evacuated” in the lateral direction.

2) Determine the lateral flow, the pressure rise and the air density next to the hull. Carry out the numerical application for the parameters in Table 3.1.

3) Check that the entropy principle is verified across the shock.

Symbol	Meaning	Value
M_0	Far field Mach number upstream of the airplane	1
p_0	Far field pressure upstream of the airplane	10^5 Pa
γ	Polytropic constant for a perfect gas	1.4
ρ_0	Far field air density upstream of the airplane	1.2 kg/m ³

Table 3.1. Parameters for Exercise 3.5

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

Chapter 4

The Riemann Problem

4.1. Definitions – solution properties

4.1.1. The Riemann problem

The Riemann problem is a basic tool in a number of numerical methods for wave propagation problems. It is defined as the combination of a Partial Differential Equation (PDE) and a piecewise constant initial condition. The PDE must be solved for the piecewise constant initial condition. The Riemann problem is said to be an Initial Value Problem (IVP), that is, a problem to be solved over an infinite domain, for which boundary conditions are meaningless. The Riemann problem for a hyperbolic system of conservation laws is written as follows:

$$\left. \begin{aligned} \frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} &= 0 \\ U(x,0) &= \begin{cases} U_L & \text{for } x < x_0 \\ U_R & \text{for } x > x_0 \end{cases} \end{aligned} \right\} \quad [4.1]$$

where the constant values U_L and U_R are the so-called left and right states of the Riemann problem, and x_0 is the location of the initial discontinuity (Figure 4.1a). The flux function F is assumed to depend only on U .

The Riemann problem for a scalar law is a particular case of the Riemann problem for hyperbolic system of conservation laws where U and F have only one component:

$$\left. \begin{aligned} \frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} &= 0 \\ U(x,0) &= \begin{cases} U_L & \text{for } x < x_0 \\ U_R & \text{for } x > x_0 \end{cases} \end{aligned} \right\} [4.2]$$

Note that the PDE to be solved contains no source term. The assumption of a zero source term allows a very specific behavior to be identified for the solution. In particular, analytical or semi-analytical solutions can be derived for the Riemann problem. The properties of the solution of the Riemann problem are dealt with in sections 4.2 and 4.3.

4.1.2. The generalized Riemann problem

The Generalized Riemann Problem (GRP) is an Initial Value Problem (IVP), the initial condition of which is not necessarily piecewise constant:

$$\left. \begin{aligned} \frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} &= 0 \\ U(x,0) &= \begin{cases} U_L(x) & \text{for } x < x_0 \\ U_R(x) & \text{for } x > x_0 \end{cases} \end{aligned} \right\} [4.3]$$

where $U_L(x)$ and $U_R(x)$ are two functions defined independently of each other on the left- and right-hand side of the location x_0 of the initial discontinuity (Figure 4.1b).

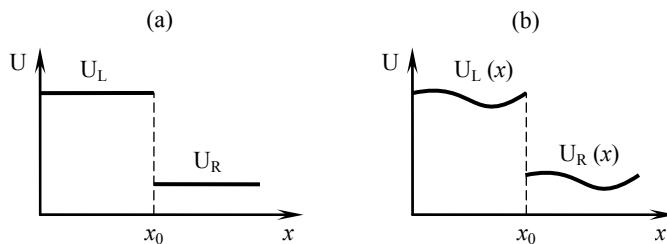


Figure 4.1. Definition sketch for the Riemann problem (a) and the generalized Riemann problem (b)

4.1.3. Solution properties

4.1.3.1. Properties

The structure of the solution is given in [LAX 57]. The following properties hold for the solution:

- Property (P4.1). The solution of the Riemann problem is self-similar: it is a function of the ratio $\xi = (x - x_0)/t$ only.
- Property (P4.2). The continuous and differentiable part of the solution of the Riemann problem verifies the following vector equation:

$$\left(A - \frac{x - x_0}{t} I \right) \frac{\partial U}{\partial x} = 0 \quad [4.4]$$

where I is the identity matrix.

- Property (P4.3). The solution of the Riemann problem for an $m \times m$ hyperbolic system of conservation laws is made of m waves separating $m + 1$ regions of constant state (the left and right states included).
- Property (P4.4). The solution of the Riemann problem at $x = x_0$ is independent of the time coordinate t .
- Property (P4.5). The solution is made of simple waves. This allows the generalized Riemann invariants to be applied across rarefaction waves and contact discontinuities. The generalized Riemann invariants [3.19] are recalled:

$$\frac{dU_1}{K_1^{(p)}} = \frac{dU_2}{K_2^{(p)}} = \dots = \frac{dU_m}{K_m^{(p)}} \quad \text{across } \frac{dx}{dt} = \lambda^{(p)}$$

4.1.3.2. Proofs

Property (P4.1) is proved as follows. Consider the non-conservation form [2.5] of the solution with a zero source term:

$$\frac{\partial U}{\partial t} + A \frac{\partial U}{\partial x} = 0 \quad [4.5]$$

An auxiliary variable V is defined as follows:

$$V(x, t) = U[x_0 + (x - x_0)k, kt] \quad [4.6]$$

where k is an arbitrary real number. The time and space derivatives of V can be expressed as:

$$\left. \begin{aligned} \frac{\partial V}{\partial t}(x, t) &= k \frac{\partial}{\partial t} U[x_0 + (x - x_0)k, kt] \\ \frac{\partial V}{\partial x}(x, t) &= k \frac{\partial}{\partial x} U[x_0 + (x - x_0)k, kt] \end{aligned} \right\} [4.7]$$

Consequently, we have:

$$\begin{aligned} \frac{\partial V}{\partial t}(x, t) + A \frac{\partial V}{\partial x}(x, t) \\ = k \frac{\partial}{\partial t} U[x_0 + (x - x_0)k, kt] + kA \frac{\partial}{\partial x} U[x_0 + (x - x_0)k, kt] = 0 \end{aligned} [4.8]$$

Consequently, if U is a solution of equation [4.5], V as defined by equation [4.6] is also a solution of equation [4.5]. The initial condition for V is obtained from equation [4.3]:

$$V(x, 0) = \begin{cases} U_L & \text{for } x < x_0 \\ U_R & \text{for } x > x_0 \end{cases} [4.9]$$

The auxiliary variable V verifies the same equation as U and has the same initial conditions. Consequently, $V = U$. From definition [4.6], we have:

$$U(x, t) = U[x_0 + (x - x_0), t] = U[x_0 + (x - x_0)k, kt] \quad \forall k \in \Re [4.10]$$

U is constant for a constant value of $\xi = (x - x_0)/t$. Consequently, U is a function of ξ alone.

Property (P4.1) can be rewritten as follows:

$$dU = 0 \quad \text{for } \frac{dx}{dt} = \frac{x - x_0}{t} [4.11]$$

Introducing the partial derivatives of U with respect to x and t into equation [4.11] yields:

$$\frac{\partial U}{\partial t} dt + \frac{\partial U}{\partial x} dx = 0 \quad \text{for } \frac{dx}{dt} = \frac{x - x_0}{t} [4.12]$$

Substituting equation [4.5] into equation [4.12] allows the time derivative to be eliminated from the equation:

$$-A \frac{\partial U}{\partial x} dt + \frac{\partial U}{\partial x} dx = 0 \quad \text{for } \frac{dx}{dt} = \frac{x - x_0}{t} \quad [4.13]$$

Equation [4.13] is rewritten as:

$$\left(A \frac{\partial U}{\partial x} - \frac{\partial U}{\partial x} \frac{dx}{dt} \right) dt = 0 \quad \text{for } \frac{dx}{dt} = \frac{x - x_0}{t} \quad [4.14]$$

Dividing by dt and factorizing $\partial U / \partial x$ leads to equation [4.4]:

$$\left(A - \frac{x - x_0}{t} I \right) \frac{\partial U}{\partial x} = 0$$

Property (P4.3) is a direct consequence of Properties (P4.1) and (P4.2). Equation [4.4] indicates that either $\xi = (x - x_0)/t$ is an eigenvalue of A (which means that the point $(x - x_0, t)$ is located within a wave centered at (x_0, t)), or $\partial U / \partial x$ is zero, which means that U is constant.

Property (P4.4) is a direct consequence of Property (P4.1). In fact, for $x = x_0$, $(x - x_0)k = x - x_0 = 0$ for any value of k . Consequently, $U(x_0, t) = U(x_0, kt)$ for all k , thus $U(x_0, t)$ does not depend on t .

Property (P4.5) is a direct consequence of Properties (P4.1) and (P4.2).

4.2. Solution for scalar conservation laws

4.2.1. The linear advection equation

The Riemann problem is written in the form [4.2] with $F = \lambda U$. As stated in the assumptions, λ is a constant in space and time. Then, as shown in section 1.1.4, equation [4.2] is equivalent to the characteristic form [1.30], recalled here:

$$U = \text{Const} \quad \text{for } \frac{dx}{dt} = \lambda$$

Since all the parts of the profile move at the speed λ , the discontinuity moves at a speed λ and the solution at $t > 0$ is given by:

$$U(x, t) = \begin{cases} U_L & \text{for } x < x_0 + \lambda t \\ U_R & \text{for } x > x_0 + \lambda t \end{cases} \quad [4.15]$$

The profile at $t > 0$ is illustrated in Figure 4.2. Note that solution [4.15] verifies the properties (P4.1) to (P4.4). Property (P4.5) is meaningless for a scalar law.

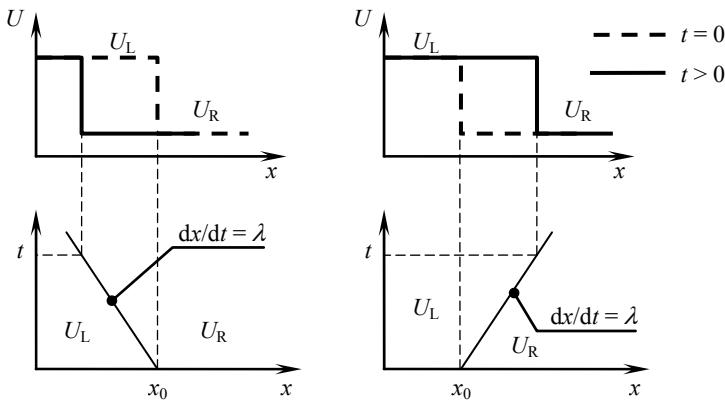


Figure 4.2. Solution of the Riemann problem in the physical space (top) and in the phase space (bottom) in the case of a negative wave speed (left) and a positive wave speed (right)

4.2.2. The inviscid Burgers equation

The conservation form [1.69] of the Burgers equation is used. The conserved variable is the flow velocity u . The flux F and the wave speed λ are given by:

$$\left. \begin{aligned} F &= \frac{u^2}{2} \\ \lambda &= u \end{aligned} \right\} \quad [4.16]$$

Two configurations may occur:

1) If $u_L < u_R$, the wave speed on the left-hand side of the initial discontinuity is smaller than the wave speed on the right-hand side and the profile cannot remain discontinuous, as stated by the entropy principle (see section 3.4.5). A rarefaction wave appears between the left and right states (Figure 4.3). The rarefaction wave extends from the abscissa $x = x_L$ to the abscissa $x = x_R$ defined as:

$$\left. \begin{aligned} x_L &= x_0 + \lambda_L t = x_0 + u_L t \\ x_R &= x_0 + \lambda_R t = x_0 + u_R t \end{aligned} \right\} \quad [4.17]$$

The rarefaction waves being issued from the point $(x_0, 0)$, it stems from Property (P4.2) that the wave speed λ is equal to the ratio $(x - x_0)/t$. Since $\lambda = u$, the equation of the profile becomes:

$$u(x, t) = \begin{cases} u_L & \text{for } x \leq x_L \\ \frac{x_R - x}{x_R - x_L} u_L + \frac{x - x_L}{x_R - x_L} u_R & \text{for } x_L \leq x \leq x_R \\ u_R & \text{for } x \geq x_R \end{cases} \quad [4.18]$$

2) If $u_L > u_R$, the wave speed on the left-hand side of the initial discontinuity is larger than the wave speed on the right-hand side and a shock appears (Figure 4.3). The shock propagates at a speed given by the jump relationship [3.28] and the solution is expressed as:

$$u(x, t) = \begin{cases} u_L & \text{for } x < x_s \\ u_R & \text{for } x > x_s \end{cases} \quad [4.19]$$

where x_s is the abscissa of the shock:

$$x_s = x_0 + c_s t = x_0 + \frac{u_L + u_R}{2} t \quad [4.20]$$

The various possible configurations are summarized in Table 4.1 and Figure 4.3.

Configuration	Wave pattern	Figure
$u_L < u_R < 0$	Rarefaction wave travelling to the left	4.3a
$u_L < 0, u_R > 0$	Rarefaction wave centered around x_0	4.3b
$0 < u_L < u_R$	Rarefaction wave travelling to the right	4.3c
$u_L > u_R, u_L + u_R < 0$	Shock travelling to the left	4.3d
$u_L > u_R, u_L + u_R = 0$	Stationary shock	4.3e
$u_L > u_R, u_L + u_R < 0$	Shock travelling to the right	4.3f

Table 4.1. The various possible configurations for the Riemann problem

Note that in the trivial case $u_L = u_R$, both the rarefaction wave conditions [4.17, 4.18] and the jump relationships [4.19, 4.20] are satisfied. Also note that when the rarefaction wave extends over $x = 0$, the value of u at $x = x_0$ is necessarily zero because the point $x = x_0$ corresponds to a zero value of the wave speed $\lambda = u$.

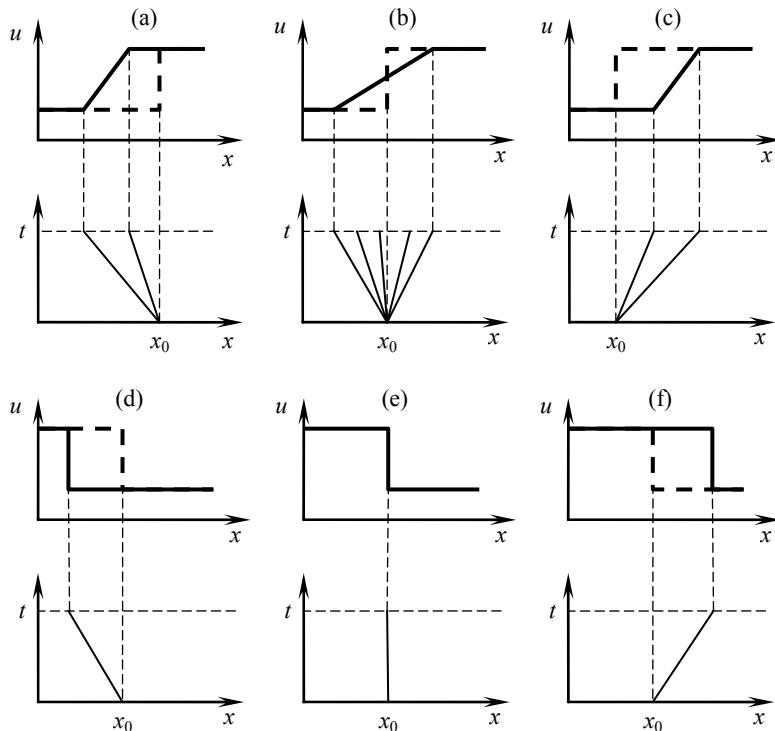


Figure 4.3. Solution of the Riemann problem in the physical space and in the phase space. Solution profiles for the configurations given in Table 4.1. Rarefaction waves (top), shock waves (bottom). Initial profile (dashed lines) and final profiles (solid lines)

4.2.3. The Buckley-Leverett equation

The water saturation s is the conserved variable in the Buckley-Leverett equation. It is also the Riemann invariant. The non-convex flux function and the wave speed are given by equations [1.111] and [1.117], recalled here:

$$\left. \begin{aligned} F &= \frac{s^2}{s^2 + (1-s)^2 b_{BL}} V_d \\ \lambda &= 2 \frac{(1-s)s}{[s^2 + (1-s)^2 b_{BL}]^2} b_{BL} V_d \end{aligned} \right\}$$

The Darcy velocity and the shape parameter b_{BL} are assumed constant in the solution of the Riemann problem. The flux function's being non-convex has the consequence that the wave speed is maximum for a saturation s_{max} between 0 and 1 (see Figure 1.19). For the sake of simplicity, V_d is assumed positive in what follows. The non-trivial case $s_L \neq s_R$ is examined hereafter. The following possible configurations are examined hereafter:

1) If $s_L < s_R \leq s_{max}$, then $\lambda_L < \lambda_R$ and the function $\lambda(s)$ varies monotonically between λ_L and λ_R . From the entropy principle, the solution cannot remain discontinuous and a rarefaction wave appears, that travels to the right. The rarefaction wave extends from the abscissa x_L to the abscissa x_R :

$$\left. \begin{array}{l} x_L = x_0 + 2\lambda_L t \\ x_R = x_0 + 2\lambda_R t \end{array} \right\} \quad [4.21]$$

The profile in the rarefaction wave obeys the relationship:

$$\frac{x - x_0}{t} = \lambda(s) \quad [4.22]$$

and the saturation profile is given by the reciprocal function $\lambda^{-1}[(x - x_0)/t]$ over the interval $[x_L, x_R]$. The solution is expressed as:

$$s(x, t) = \begin{cases} s_L & \text{for } x \leq x_L \\ \lambda^{-1}[(x - x_0)/t] & \text{for } x_L \leq x \leq x_R \\ s_R & \text{for } x \geq x_R \end{cases} \quad [4.23]$$

2) If $s_R < s_L \leq s_{max}$, then $\lambda_L < \lambda_R$ and the function $\lambda(s)$ varies monotonically between λ_L and λ_R . A shock appears, that travels at a speed c_s given by:

$$c_s = \frac{F(s_L) - F(s_R)}{s_L - s_R} = \frac{\frac{1}{1 + (1/s_L - 1)^2 b_{BL}} - \frac{1}{1 + (1/s_R - 1)^2 b_{BL}}}{s_L - s_R} V_d \quad [4.24]$$

The solution is then:

$$s(x, t) = \begin{cases} s_L & \text{for } x < x_0 + c_s t \\ s_R & \text{for } x > x_0 + c_s t \end{cases} \quad [4.25]$$

3) If $s_{\max} \leq s_L < s_R$, then $\lambda_L > \lambda_R$ and the function $\lambda(s)$ varies monotonically between λ_L and λ_R . A shock appears, that travels at a speed c_s given by equation [4.24]. The solution is given by equation [4.25].

4) If $s_{\max} \leq s_R < s_L$, then $\lambda_L < \lambda_R$ and the function $\lambda(s)$ varies monotonically between λ_L and λ_R . The profile cannot remain discontinuous and a rarefaction wave appears between the abscissa x_L and x_R as given by equation [4.21] The solution is given by equation [4.23].

5) If $s_L < s_{\max} < s_R$, then the function $\lambda(s)$ does not vary monotonically between λ_L and λ_R . A compound wave appears. The solution profile is determined using the method of characteristics followed by the equal area rule to enforce uniqueness (Figure 4.4). The solution method is best described by expressing x as a function of s . The initial condition may be written as:

$$x(s, 0) = x_0, \quad s_L < x < s_R \quad [4.26]$$

As seen in section 1.6.2 (equation [1.118]), the saturation s is invariant along the characteristic lines. Consequently, the wave speed associated with a given value of s depends only on s :

$$\frac{dx}{dt}(s, t) = \lambda(s) \quad \forall t \quad [4.27]$$

Equation [4.27] is integrated into:

$$x(s, t) = x_0 + \lambda(s)t, \quad s_L < s < s_R \quad [4.28]$$

The profile $x(s, t)$ reaches a local maximum for $s = s_{\max}$, which accounts for the multiple-valued character of the solution $s(x, t)$ (Figure 4.4). Uniqueness is enforced by correcting the profile according to the equal area rule. The corrected profile is made of a shock followed by a rarefaction wave. It satisfies mass conservation, that is:

$$\int_{s_L}^{s_R} \lambda(s) ds = \int_{s_L}^{s_c} \lambda(s) ds + \int_{s_c}^{s_R} c_s ds \quad [4.29]$$

where s_c denotes the saturation immediately behind the shock. The left- and right-hand sides of the equation represent respectively the integrals of the profile before and after the correction. The corrected profile is divided into two regions; between

s_L and s_c , the characteristic form [4.26] is applicable, while in the region $[s_c, s_R]$ the profile moves at the constant speed c_s . Noting that the integral of λ between two given values of s is by definition the difference between the fluxes, equation [4.29] becomes:

$$F(s_R) - F(s_L) = F(s_R) - F(s_c) + (s_R - s_c)c_s \quad [4.30]$$

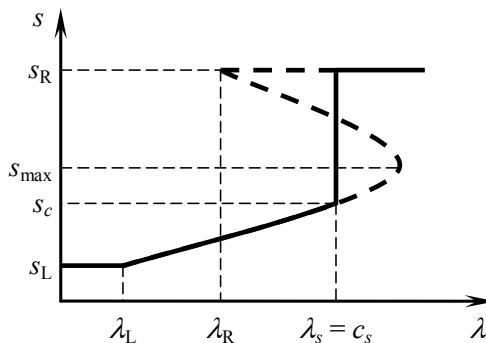


Figure 4.4. Applying the equal area rule to the solution of the Riemann problem for the Buckley-Leverett equation. Solution profile before the correction (dashed line) and after the correction (solid line)

Noting that $c_s = \lambda(s_c)$, equation [4.30] is simplified into:

$$F(s_c) - (s_R - s_c)\lambda(s_c) = F(s_L) \quad [4.31]$$

Equation [4.31] must be solved iteratively for s_c . The complete solution is:

$$x(s, t) = \begin{cases} x_0 + \lambda(s)t & \text{for } s_L < s \leq s_c \\ x_0 + c_s t & \text{for } s_c \leq s < s_L \end{cases} \quad [4.32]$$

6) If $s_R < s_{\max} < s_L$, the non-convex character of the flux function leads to the same configuration as in 5). A compound wave appears. The solution is calculated exactly in the same way as in 5).

The various possible wave configurations are summarized in Table 4.2. Figure 4.5 illustrates the different cases.

Configuration	Wave pattern	Figure
$s_L > s_R \geq s_{\max}$	Rarefaction wave travelling to the right	4.5a
$s_{\max} \leq s_L < s_R$	Shock travelling to the right	4.5b
$s_R < s_{\max} < s_L$	Compound wave travelling to the right	4.5c
$s_L < s_R \leq s_{\max}$	Rarefaction wave travelling to the right	4.5d
$s_{\max} \geq s_L > s_R$	Shock wave travelling to the right	4.5e
$s_L < s_{\max} < s_R$	Compound wave travelling to the right	4.5f

Table 4.2. The various possible configurations for the Riemann problem in the case of a positive Darcy velocity

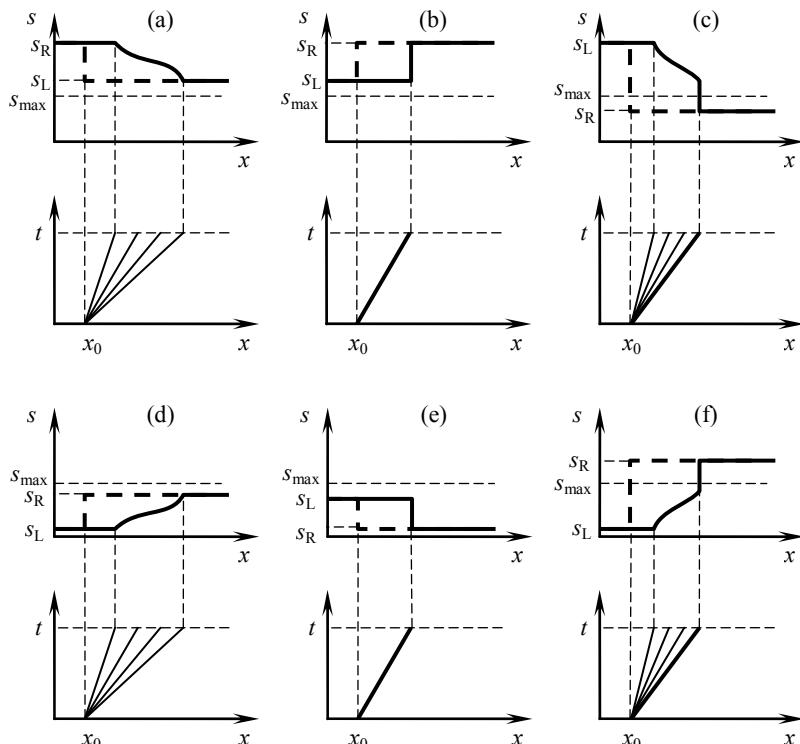


Figure 4.5. Solution of the Riemann problem for the Buckley-Leverett equation. Sketch in the phase space and in the physical space for a positive Darcy velocity. Initial profile (dashed lines), final profile (solid lines)

4.3. Solution for hyperbolic systems of conservation laws

4.3.1. General principle

As stated in Property (P4.3) (see section 4.1.3), the solution of the Riemann problem is made of $m + 1$ regions of constant state (among which the left and right state of the Riemann problem are to be counted) separated by m waves. Consequently, there are $m - 1$ regions of constant state over which the value of \mathbf{U} must be calculated (Figure 4.6).

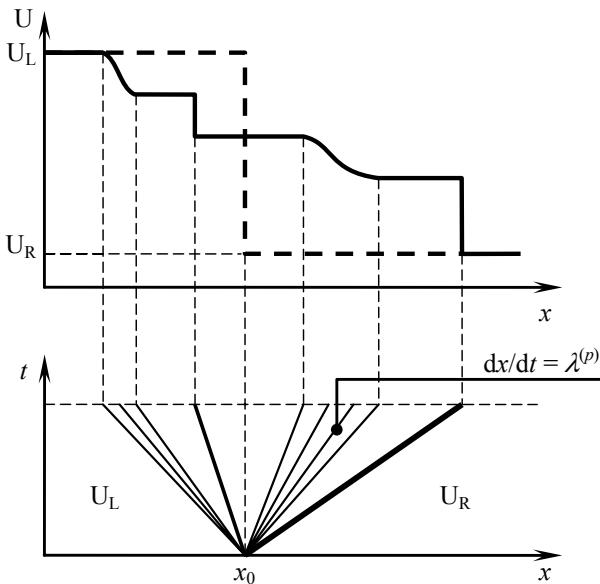


Figure 4.6. Solution of the Riemann problem for an $m \times m$ hyperbolic system of conservation laws. Structure of the solution in the physical space (top) and in the phase space (bottom)

The nature of the waves that separate the regions of constant state is not known *a priori*. Consequently, the nature of the relationship to be used across the waves (Riemann invariants/generalized Riemann invariants or jump relationships) is not known in advance. Counting the unknowns and the available relationships leads however to the conclusion that the solution can be determined uniquely. In fact:

- there are $m - 1$ regions where the constant state is unknown. The size of the vector \mathbf{U} being m , there are $(m - 1)m$ unknowns;
- if the wave p is a shock, the Rankine-Hugoniot relationships provide m relationships across the wave, while introducing an additional unknown (the shock

speed c_s). Eliminating c_s from the jump relationships leaves out $m - 1$ independent equations for the wave p ;

– if the p th wave is a rarefaction wave or a contact discontinuity, the generalized Riemann invariants [3.19] provide $m - 1$ independent equations for the wave p .

In other words, $m - 1$ independent equations are available across each of the m waves. Consequently, the number of equations matches the number of unknowns and the solution is unique.

The solution of the Riemann problem usually involves two nested iteration loops.

1) In a first step, an assumption is made regarding the wave pattern (that is, the nature of the various waves in the solution). This assumption yields the system of equations to be solved.

2) The system of equations is solved. This generally requires an iterative procedure because the jump relationships and the Riemann invariants are nonlinear functions of the components of the variable U .

3) The solution obtained at the end of the step 2) is used to check the validity of the wave pattern assumed in Step 1). If the assumptions made in 1) prove not to be compatible with the solution obtained in 2), another wave pattern must be assumed and the steps 1) – 2) must be completed again.

The following heuristic rule leads to correct guesses in many cases:

- 1) If $\lambda^{(p)}(U_L) < \lambda^{(p)}(U_R)$, the wave p is likely to be a rarefaction wave.
- 2) If $\lambda^{(p)}(U_L) = \lambda^{(p)}(U_R)$, the wave p is likely to be a contact discontinuity.
- 3) If $\lambda^{(p)}(U_L) > \lambda^{(p)}(U_R)$, the wave p is likely to be a shock wave.

Quite obviously, this rule is always verified in the scalar case. It is also applicable to hyperbolic systems in a number of cases, as shown in the next sections.

4.3.2. Application to the water hammer problem: sudden valve failure

Consider a pipe separated into two parts by a valve. The valve is initially closed, the water is at rest. The pressure is assumed to be higher on the left-hand side of the valve than on the right-hand side. The purpose is to study the behavior of the flow in the case of a sudden failure of the valve. The most severe transient occurs when the valve collapses instantaneously in a frictionless pipe. In what follows, the valve is assumed to disappear instantaneously at $t = 0$, thus leaving the two parts of the pipes in direct contact with each other (Figure 4.7, dashed line).

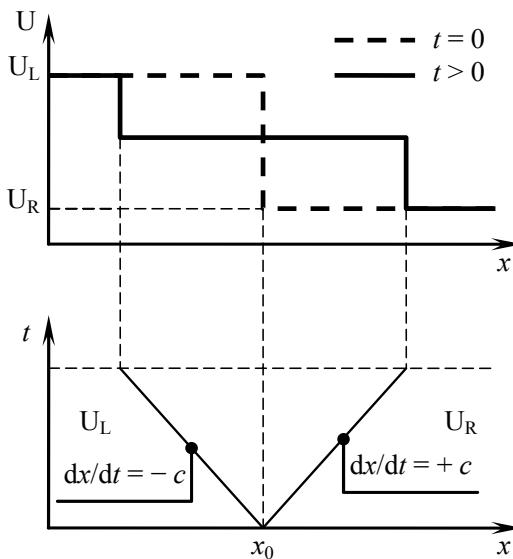


Figure 4.7. Definition sketch for the pipe transient induced by the sudden failure of a valve.
Sketch in the physical space (top) and in the phase space (bottom)

As seen in section 2.4.3, the wave speeds are constant, independent from U . Consequently, the waves are contact discontinuities that separate the left and right states of the Riemann problem from an intermediate region of constant state (Figure 4.7). The pressure and the velocity are given by:

$$p(x, t) = \begin{cases} p_L & \text{for } x < x_0 - ct \\ p^* & \text{for } x_0 - ct < x < x_0 + ct \\ p_R & \text{for } x > x_0 + ct \end{cases} \quad [4.33]$$

$$u(x, t) = \begin{cases} u_L & \text{for } x < x_0 - ct \\ u^* & \text{for } x_0 - ct < x < x_0 + ct \\ u_R & \text{for } x > x_0 + ct \end{cases}$$

where p^* and u^* denote the pressure and the flow velocity in the region of constant state respectively. The expressions for p^* and u^* may be obtained from two approaches.

1) In the first approach, the intermediate region of constant state is connected to the left and right states of the Riemann problem using characteristic lines and the Riemann invariants (Figure 4.8a). The left state is connected to the intermediate

region by the characteristic $dx/dt = +c$, while the right state is connected to the intermediate region by the characteristic $dx/dt = -c$. This leads to system [2.85] derived in section 2.4.4.1.

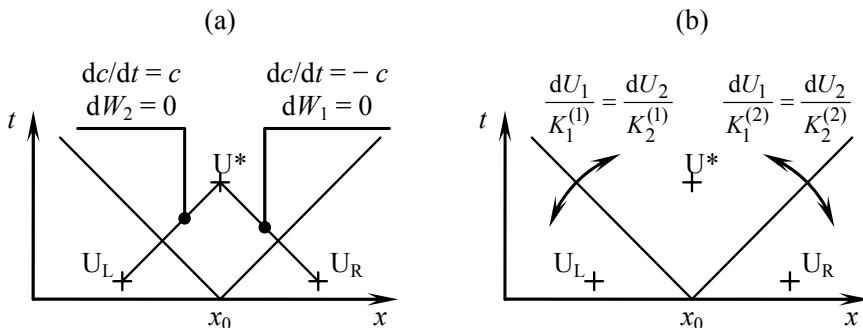


Figure 4.8. Two possible options for the solution of the Riemann problem.
Riemann invariants (a), generalized Riemann invariants (b)

$$\left. \begin{array}{l} p^* + \rho c u^* = p_L + \rho c u_L \\ p^* - \rho c u^* = p_R - \rho c u_R \end{array} \right\} \quad [4.34]$$

with $u_L = u_R = 0$ because the water is initially immobile. The solution of the system [4.34] is:

$$\left. \begin{array}{l} p^* = \frac{p_L + p_R}{2} \\ u^* = \frac{p_L - p_R}{2\rho c} \end{array} \right\} \quad [4.35]$$

2) In a second approach, the generalized Riemann invariants [3.19] are used across the waves (Figure 4.8b). Remember that the conserved variable and the eigenvectors are given by equations [2.68] and [2.70]:

$$\mathbf{U} = \begin{bmatrix} \rho A \\ \rho A u \end{bmatrix}, \quad \mathbf{K}^{(1)} = \begin{bmatrix} 1 \\ -c \end{bmatrix}, \quad \mathbf{K}^{(2)} = \begin{bmatrix} 1 \\ +c \end{bmatrix}$$

Applying equation [3.19] yields the following system:

$$\left. \begin{array}{l} \frac{d(\rho A)}{1} = \frac{d(\rho Au)}{-c} \\ \frac{d(\rho A)}{1} = \frac{d(\rho Au)}{c} \end{array} \right\} \begin{array}{l} \text{across } \frac{dx}{dt} = -c \\ \text{across } \frac{dx}{dt} = +c \end{array} \quad [4.36]$$

Remembering that $d(\rho A) = A \ dp/c^2$ by definition and that $d(\rho Au) \approx \rho A \ du$, equations [4.36] become:

$$\left. \begin{array}{l} dp = -\rho cdu \quad \text{across } \frac{dx}{dt} = -c \\ dp = \rho cdu \quad \text{across } \frac{dx}{dt} = +c \end{array} \right\} \quad [4.37]$$

The first relationship is valid across the wave $dx/dt = -c$. It can be used to connect the left state of the Riemann problem to the intermediate region of constant state. The second relationship is used to connect the right state of the Riemann problem to the intermediate region of constant state. Integrating relationships [4.37] leads to:

$$\left. \begin{array}{l} p^* - p_L = (u_L - u^*)\rho c \\ p^* - p_R = (u^* - u_R)\rho c \end{array} \right\} \quad [4.38]$$

This system, that is equivalent to system [4.34], has the same solution [4.35].

4.3.3. Free surface flow: the dambreak problem

4.3.3.1. Introduction

The dambreak problem is a Riemann problem for the shallow water equations [STO 57]. The following assumptions are made. The dam is idealized as an infinitely thin wall separating a left state (the upstream side of the dam, full of water) from a right state (the downstream valley). The bottom is horizontal, the effect of friction is neglected. The channel is assumed to be rectangular, the water is initially at rest on both sides of the dam. The downstream side of the dam may or may not be dry, the dry case being a limit case of the wet case. The breaking of the dam is idealized by the instantaneous removal of the wall at $t = 0$. The problem is illustrated by Figure 4.9.

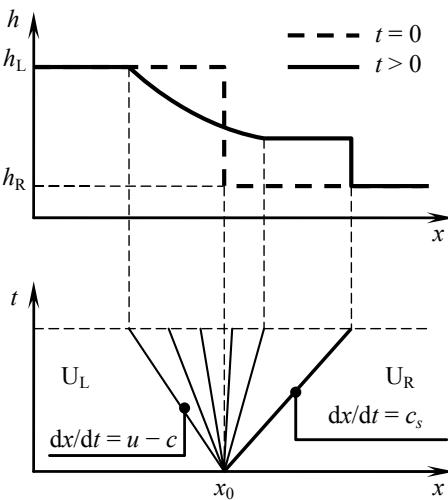


Figure 4.9. Definition sketch for the dam-break problem.
Sketch in the physical space (top) and in the phase space (bottom)

4.3.3.2. Wave pattern

The first step in the solution of the Riemann problem is the determination of the wave pattern. Several approaches may be used.

1) The heuristic approach of section 4.3.1 may be used. Using the assumption of a zero initial flow velocity on both sides of the dam, the initial wave speeds are:

$$\left. \begin{array}{l} \lambda_L^{(1)} = -c_L = -(gh_L) \\ \lambda_R^{(1)} = -c_R = -(gh_R) \\ \lambda_L^{(2)} = c_L = (gh_L) \\ \lambda_R^{(2)} = c_R = (gh_R) \end{array} \right\} [4.39]$$

Since $\lambda^{(1)}$ is smaller on the left-hand side of the dam than on the right-hand side, the first wave is likely to be a rarefaction wave. Conversely, $\lambda^{(2)}$ being larger on the left-hand side of the dam than on the right-hand side, the second wave is likely to be a shock.

2) Theorem 2 in section 3.4.7 may be used to provide an approximation of the flow variables in the intermediate region of constant state. Using the Riemann

invariant $u + 2c$ across the first wave and the Riemann invariant $u - 2c$ across the second wave yields the following system:

$$\left. \begin{aligned} u^* - 2c^* &= u_R - 2c_R = -2c_R \\ u^* + 2c^* &= u_L + 2c_L = 2c_L \end{aligned} \right\} \quad [4.40]$$

which leads to the following estimate for the wave speeds in the intermediate region of constant state:

$$\left. \begin{aligned} \lambda^{(1)*} &= u^* - c^* = \frac{1}{2}c_L - \frac{3}{2}c_R \\ \lambda^{(2)*} &= u^* + c^* = \frac{3}{2}c_L - \frac{1}{2}c_R \end{aligned} \right\} \quad [4.41]$$

Since $c_L > c_R$, $\lambda^{(1)}$ is larger in the intermediate region of constant state than on the left-hand side of the initial discontinuity. The first wave is a rarefaction wave. Conversely, $\lambda^{(2)}$ being larger in the intermediate region of constant state than in the right state of the Riemann problem, the second wave is a shock wave.

3) A physical reasoning may be used. The difference between the water levels on both sides of the dam induces a difference between the pressure on the left- and right-hand sides of the discontinuity. When the dam disappears, the pressure difference is exerted on both sides of a control volume that can be made infinitely thin. Therefore the water immediately to the left of the discontinuity is subjected to an infinite acceleration at $t = 0$. The water instantaneously reaches a finite speed. The wave speed $u + c$ in the intermediate region is larger than that on the right-hand side of the dam and the wave $u + c$ is indeed a shock. The movement of the water to the right necessarily implies that the water level in the intermediate region of constant state is smaller than in the left state. The wave speed $u - c$ in the intermediate region of constant state is larger than in the left state and the wave $u - c$ is a rarefaction wave.

4.3.3.3. Calculation of the solution

According to the wave pattern as identified in the previous section, the following relationships are to be used: the Riemann invariant $u - 2c$ should be used to connect the left state (i.e. the upstream side of the dam) to the intermediate region of constant state and the jump relationships should be used to connect the intermediate region of constant state to the right state. Noting that $u_L = u_R = 0$ and that $A = bh$, the following system is obtained:

$$\left. \begin{aligned} u^* - 2c^* &= 2c_L \\ (h^* - h_R)c_s &= (hu)^* \\ (hu)^* c_s &= (hu^2 + gh^2/2)^* - gh_R^2/2 \end{aligned} \right\} [4.42]$$

where c_s is the speed of the shock. The first equation [4.42] yields:

$$u^* = 2(c_L - c^*) [4.43]$$

Substituting equation [4.43] into the second equation [4.42] yields the following formula for c_s :

$$c_s = \frac{h^* u^*}{h^* - h_R} = 2h^* \frac{c_L - c^*}{h^* - h_R} [4.44]$$

Substituting equation [4.44] into the third relationship [4.42] leads to the following equation in h^* :

$$(c_L - c^*)^2 h_R h^* = \frac{g}{8} (h^* - h_R)^2 (h^* + h_R) [4.45]$$

with $c^* = (gh^*)^{1/2}$. Equation [4.45] can be solved using standard iterative techniques such as Newton's method. The value of h^* leads to that of c^* , thus yielding u^* and c_s from equations [4.43–44].

The following remarks can be made on the structure of the solution (see Figure 4.10):

- The flow velocity varies linearly between the left state and the intermediate region of constant state. Indeed, the self-similar character of the solution has the consequence that the wave speed $\lambda^{(1)}$ in the rarefaction wave is equal to $(x - x_0)/t$. Moreover, the invariance of $u + 2c$ may be stated across the rarefaction wave. These two conditions lead to:

$$\left. \begin{aligned} u - c &= \frac{x - x_0}{t} \\ u + 2c &= 2c_L \end{aligned} \right\} [4.46]$$

Solving the system [4.46] for u leads to:

$$u(x, t) = \frac{2}{3} \left(\frac{x - x_0}{t} + c_L \right) [4.47]$$

The flow velocity increases linearly from the left-hand boundary of the rarefaction wave $dx/dt = u - c$ to the left-hand boundary of the intermediate region of constant state. Note that if the intermediate region of constant state is such that $u^* > c^*$, the point $x = x_0$ is contained in the rarefaction wave and $u(x_0, t) = c(x_0, t) = 2c_L/3$.

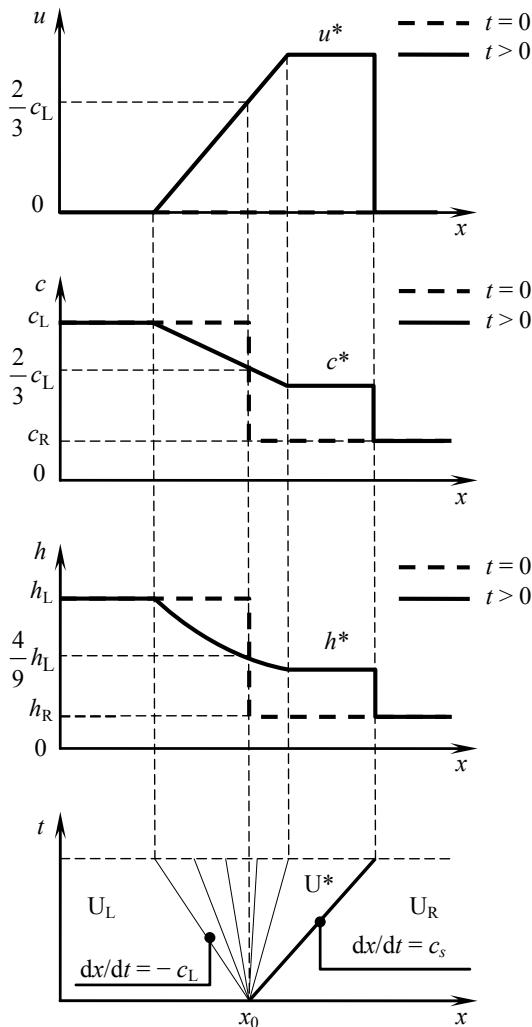


Figure 4.10. Dambreak problem. Structure of the solution and profiles of the flow variables

– The shape of the water depth profile in the rarefaction wave is parabolic. This can be deduced from equations [4.46] that leads to the following expression for c .

$$c(x, t) = \frac{1}{3} \left(2c_L - \frac{x - x_0}{t} \right) \quad [4.48]$$

Noting that $h = c^2/g$, $h(x, t)$ is given by:

$$h(x, t) = \frac{c^2(x, t)}{g} = \frac{1}{9g} \left(2c_L - \frac{x - x_0}{t} \right)^2 \quad [4.49]$$

Note that if the intermediate state is such that $u^* > c^*$, then $h(x_0, t) = 4h_L/9$.

4.3.3.4. A particular case: dambreak on a dry bed

In the particular case where the downstream valley is dry, $h_R = 0$ and system [4.42] simplifies into:

$$\left. \begin{aligned} u^* - 2c^* &= 2c_L \\ h^* c_s &= (hu)^* \\ (hu)^* c_s &= (hu^2 + gh^2/2)^* \end{aligned} \right\} \quad [4.50]$$

The second equation [4.50] yields the obvious solution $u^* = c_s$. Substituting this condition into the third equation [4.50] leads to:

$$h^* = 0 \quad [4.51]$$

The first equation [4.50] yields the following equality:

$$u^* = c_s = 2c_L \quad [4.52]$$

As a consequence of equation [4.52], the width of the intermediate region of constant state is zero. Indeed, since $c^* = 0$, the wave speed $u^* - c^*$ of the first wave on the left-hand side of the rarefaction wave is equal to u^* , and u^* is equal to the speed c_s of the shock. The amplitude of the shock is zero because $h^* = h_R = 0$. In other words, when the dam breaks on a dry bed, the intermediate region of constant state and the shock disappear. The structure of the solution is illustrated in Figure 4.11.

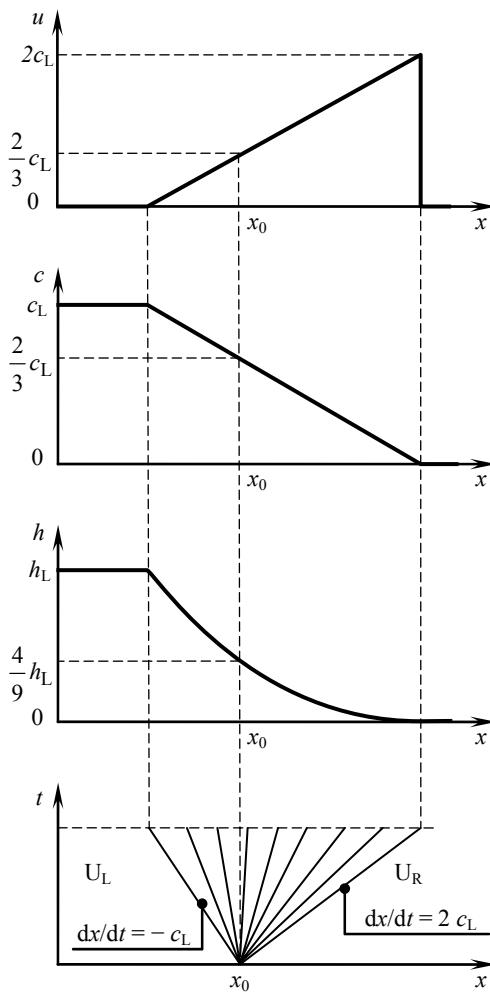


Figure 4.11. Dambreak on a dry bed. Structure of the solution and profiles of the flow variables

4.3.4. The Euler equations: the shock tube problem

4.3.4.1. Introduction

The shock tube is the analog for gas dynamics of the dam-break problem for the shallow water equations. The shock tube problem is an idealized representation of the sudden failure of a membrane separating a high pressure zone from a lower

pressure zone in a long and narrow tube (which allows the problem to be considered as one-dimensional). The particular case where the right-hand side of the tube is void is the exact transposition of the dambreak problem on a dry bed.

Note that the assumption of a perfect gas leads to equation [2.210]. Using equation [2.182], this condition leads to the property that p/ρ^γ is constant along the characteristics $dx/dt = u$. Therefore the following relationship is assumed:

$$p(\rho) = p_0 \left(\frac{\rho}{\rho_0} \right)^\gamma \quad [4.53]$$

The speed of sound is therefore (see equation [2.211]):

$$c = \left(\gamma \frac{p_0}{\rho_0^\gamma} p^{\gamma-1} \right)^{1/2} = \left(\gamma \frac{p}{\rho} \right)^{1/2} \quad [4.54]$$

Consequently the speed of sound is an increasing function of the pressure (or the density). This important property is used in the derivation of the wave pattern.

4.3.4.2. Wave pattern

The first step in the solution of the Riemann problem consists of guessing the wave pattern. The following approaches may be used.

1) The heuristic approach presented in section 4.3.1. The wave speeds for the left and right states are given by:

$$\left. \begin{array}{l} \lambda_L^{(1)} = -c_L \\ \lambda_R^{(1)} = -c_R \\ \lambda_L^{(2)} = 0 \\ \lambda_R^{(2)} = 0 \\ \lambda_L^{(3)} = c_L \\ \lambda_R^{(3)} = c_R \end{array} \right\} \quad [4.55]$$

The wave speed $\lambda^{(1)}$ is smaller on the left-hand side of the discontinuity than on the left-hand side because the pressure p_L is larger than the pressure p_R . Therefore, the first wave is likely to be a rarefaction wave. Conversely, $\lambda^{(3)}$ is larger on the left-

hand side of the discontinuity than on the right-hand side. The third wave is likely to be a shock. The second wave is known to be a contact discontinuity.

2) Using Theorem 2 in section 3.4.7 allows an approximation of the intermediate regions of constant state to be derived. Consider first the second wave. The generalized Riemann invariant for the second wave is derived from equations [3.19] and [2.201]:

$$\frac{d\rho}{0} = \frac{d(\rho u)}{0} = \frac{dE}{1} \quad \text{across } \frac{dx}{dt} = \lambda^{(2)} = u \quad [4.56]$$

Equation [4.45] is meaningful only if $du = d\rho = 0$ across the second wave. Since u is also the speed $\lambda^{(2)}$ of the second wave, the wave is a contact discontinuity. Applying the Riemann invariants across the first and third waves leads to:

$$\left. \begin{aligned} \beta_1 p^{*\beta_2} - u^* &= \beta_1 p_R^{\beta_2} - u_R \\ \beta_1 p^{*\beta_2} + u^* &= \beta_1 p_L^{\beta_2} + u_L \end{aligned} \right\} \quad [4.57]$$

remembering that $u_L = u_R = 0$, the following solution is obtained for equations [4.57]:

$$\left. \begin{aligned} u^* &= \frac{\beta_1}{2} \left(p_L^{\beta_2} - p_R^{\beta_2} \right) \\ p^* &= \left(\frac{p_L^{\beta_2} + p_R^{\beta_2}}{2} \right)^{1/\beta_2} \end{aligned} \right\} \quad [4.58]$$

u^* is positive because $p_L > p_R$. p^* lies between p_L and p_R . Consequently the wave speed $\lambda^{(3)} = u + c$ is larger in the intermediate region than in the right state of the Riemann problem and the third wave is a shock. Conversely, the wave speed $\lambda^{(1)} = u - c$ is larger in the intermediate region than in the left state and the first wave is a rarefaction wave.

The structure of the solution is the following (see Figure 4.12):

- The wave $dx/dt = \lambda^{(1)}$ is a rarefaction wave. The Riemann invariant W_3 is valid across the wave.
- The wave $dx/dt = \lambda^{(2)}$ is a contact discontinuity. The pressure p and the flow velocity u are constant across the wave. In contrast, s is discontinuous across the wave. Since p is identical on both sides of the second wave, the discontinuity in s implies a discontinuity in the density ρ .
- The wave $dx/dt = \lambda^{(3)}$ is a shock. The jump relationships hold across it.

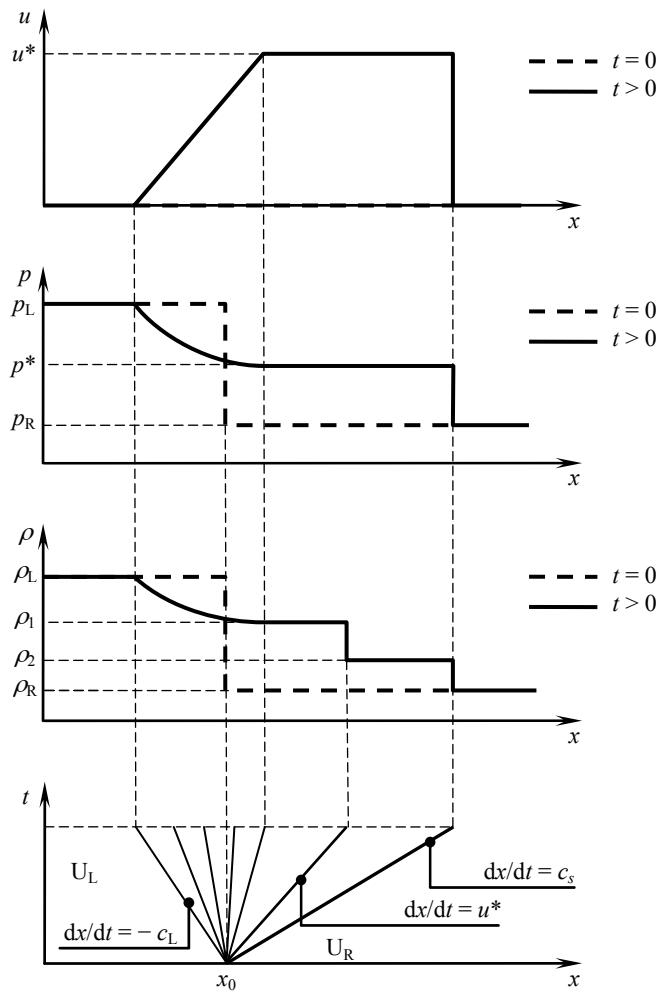


Figure 4.12. Structure of the solution for the shock tube problem.

4.3.4.3. Calculation of the solution

In a first step, the pressure p^* and the flow velocity u^* are calculated. This is done using the Riemann invariants across the first wave and the jump relationships across the third wave. Using the first Riemann invariant leads to:

$$\beta_1 p^{*\beta_2} + u^* = \beta_1 p_L^{\beta_2} + u_L \quad [4.59]$$

For the sake of clarity, the jump relationships are written in a coordinate system that moves with the shock:

$$\left. \begin{aligned} (u^* - c_s) \rho_2 &= (u_R - c_s) \rho_R \\ (u^* - c_s)^2 \rho_2 + p^* &= (u_R - c_s)^2 \rho_R + p_R \\ (u^* - c_s)(E_2 + p^*) &= (u_R - c_s)(E_R + p_R) \end{aligned} \right\} [4.60]$$

Substituting the first equation [4.60] into the second, factorizing with $q = (u - c_s) \rho_2$ leads to:

$$\left. \begin{aligned} u &= u_R + \frac{p_R - p^*}{q} \\ q^2 &= \frac{p_R - p^*}{\frac{1}{\rho_2} - \frac{1}{\rho_R}} \end{aligned} \right\} [4.61]$$

The density ρ_2 is eliminated from the second equation [4.61] by writing a relationship between the pressure and the density across the shock. To do so, the third equation [4.60] is used together with the definition of E in equation [2.177]:

$$e_2 - e_R = \frac{p + p_R}{2} \frac{\rho_2 - \rho_R}{\rho_2 \rho_R} [4.62]$$

Substituting equation [2.189] into equation [4.62] leads to:

$$\frac{\rho_2}{\rho_R} = \frac{\frac{p^*}{p_R} + \frac{\gamma - 1}{\gamma + 1}}{\frac{p^* \gamma - 1}{p_R \gamma + 1} + 1} [4.63]$$

which leads to the following expression for q :

$$q = \left[\frac{(\gamma + 1)p^* + (\gamma - 1)p_R}{2} \rho_R \right]^{1/2} [4.64]$$

Substituting equation [4.64] into equation [4.61] leads to:

$$u^* = u_R + (p^* - p_R) \left[\frac{2}{\rho_R} \frac{1}{(\gamma + 1)p^* + (\gamma - 1)p_R} \right]^{1/2} [4.65]$$

Substituting equation [4.65] into equation [4.59] allows u^* to be eliminated:

$$\beta_L p^{*\beta_2} + (p^* - p_R) \left[\frac{2}{\rho_R} \frac{1}{(\gamma + 1)p^* + (\gamma - 1)p_R} \right] = \beta_L p_L^{\beta_2} + u_L - u_R \quad [4.66]$$

Solving equation [4.66] for p^* allows u^* to be calculated from equation [4.65]. The density ρ_2 is then given by equation [4.63]. The shock speed c_s is obtained from the first equation [4.60]:

$$c_s = \frac{\rho_2 u^* - \rho_R u_R}{\rho_2 - \rho_R} = \frac{\rho_2 u^*}{\rho_2 - \rho_R} \quad [4.67]$$

The rarefaction wave $dx/dt = \lambda^{(1)}$ being issued from the abscissa x_0 of the initial discontinuity, the following relationship holds:

$$\lambda^{(1)}(x, t) = u - c = \frac{x - x_0}{t} \quad [4.68]$$

Combining equations [4.54], [4.59] and [4.68] leads to the following profiles:

$$\left. \begin{aligned} \rho(x, t) &= \left[\frac{2}{\gamma + 1} + \frac{\gamma - 1}{\gamma + 1} \frac{1}{c_L} \left(u_L - \frac{x - x_0}{t} \right) \right]^{\frac{2}{\gamma - 1}} \rho_L \\ u(x, t) &= \frac{2}{\gamma + 1} \left(c_L + \frac{x - x_0}{t} \right) \\ p(x, t) &= \left[\frac{2}{\gamma + 1} + \frac{\gamma - 1}{\gamma + 1} \frac{1}{c_L} \left(u_L - \frac{x - x_0}{t} \right) \right]^{\frac{2}{\gamma - 1}} p_L \end{aligned} \right\}, \lambda_L^{(1)} \leq \frac{x}{t} \leq \lambda^{(1)*} \quad [4.69]$$

The rarefaction wave extends to the point $x = x_1$ at which $p = p^*$ and $u = u^*$. At this abscissa, the density is equal to ρ_1 . Substituting the condition $u = u^*$ in equation [4.69] leads to:

$$\frac{2}{\gamma + 1} \left(c_L + \frac{x_1 - x_0}{t} \right) = u^* \quad [4.70]$$

which gives:

$$\frac{x_1 - x_0}{t} = \frac{\gamma + 1}{2} u^* - c_L \quad [4.71]$$

Substituting equation [4.71] into the first equation [4.69], noting that $u_L = 0$ leads to:

$$\rho_1 = \rho(x_1, t) = \left[\frac{2}{\gamma+1} - \frac{\gamma-1}{\gamma+1} \frac{1}{c_L} \left(\frac{\gamma+1}{2} u^* - c_L \right) \right]^{\frac{2}{\gamma-1}} \rho_L \quad [4.72]$$

Note that the flow velocity increases linearly with x within the rarefaction wave. As a consequence (see equation [4.69]), the speed of sound also varies linearly within the rarefaction wave. If the pressure p_R is low enough for the rarefaction wave to include the abscissa x_0 of the initial discontinuity, the flow variables at $x = x_0$ are given by

$$\left. \begin{aligned} \rho(x_0, t) &= \left(\frac{2}{\gamma+1} \right)^{\frac{2}{\gamma-1}} \rho_L \\ u(x_0, t) &= c(x_0, t) = \frac{2}{\gamma+1} c_L \\ p(x_0, t) &= \left(\frac{2}{\gamma+1} \right)^{\frac{2}{\gamma-1}} p_L \end{aligned} \right\} \quad [4.73]$$

4.3.4.4. A particular case: shock tube with a vacuum

A particular case occurs when the region on the right-hand side of the membrane in the tube contains no gas. Using $p_R = \rho_R = 0$ leads to the conclusion that the width of the region of constant state is zero because $u^* = c_s$ in equation [4.60] and $c^* = 0$. Then $p^* = 0$. The solution is made of a rarefaction wave that extends from $p = p_L$ to $p = p^* = 0$. The region to the right of the rarefaction wave is a void region. The rarefaction wave moves to the right at a speed $u^* + c^* = u^*$.

4.4. Summary

4.4.1. What you should remember

The Riemann problem is an Initial Value Problem (IVP) made of (i) a piecewise constant initial condition and (ii) a PDE to be solved for this initial condition.

The solution of the Riemann problem for a hyperbolic conservation law or a hyperbolic system of conservation laws is self-similar. The characteristics are straight lines in the phase space and the solution depends only on the ratio $(x - x_0)/t$, where x_0 is the abscissa of the discontinuity in the initial condition. The solution at $x = x_0$ is constant.

The solution of the Riemann problem for an $m \times m$ hyperbolic system of conservation laws is made of m waves separating regions of constant state.

The Riemann invariants and the generalized Riemann invariants [3.19] are applicable across a rarefaction wave or a contact discontinuity.

When the flux function is linear (e.g. the linear advection equation, see section 4.2.1, or the water hammer equations, see section 4.3.2), all the waves are contact discontinuities. When the flux function is convex or concave, the waves are rarefaction waves or shocks. When the flux function is non-convex, compound waves may appear for specific combinations of the left and right state (see the Buckley-Leverett equation in section 4.2.3).

The dambreak problem (section 4.3.3) is a Riemann problem for the Saint Venant equations. The solution is made of a rarefaction wave travelling upstream and a shock wave propagating downstream. When the dam breaks on a dry bed, the amplitude of the shock wave becomes zero and the solution is made of a rarefaction wave that connects directly the left state to the right state. The flow velocity u and the speed of the wave c in still water vary linearly with x within the rarefaction wave. The water depth follows a parabolic profile.

The shock tube (section 4.3.4) is a Riemann problem for the Euler equations. When the pressure in the left state is higher than in the right state, the following waves are found from left to right: a rarefaction wave, a contact discontinuity and a shock wave. The pressure and the velocity are constant across the contact discontinuity, while the density is not because the entropy that is transported at the speed of the flow is different on both sides of the contact discontinuity. When the right state is a void region, the intermediate region of constant state disappears and the solution is made of a rarefaction wave connecting directly the left state to the right state.

4.4.2. Application exercises

4.4.2.1. Exercise 4.1: the Saint Venant equations

Solve the Riemann problem for the Saint Venant equations in a rectangular channel for the following initial data (note the symmetry in the initial state):

$$\left. \begin{array}{l} h_L = h_R \\ u_L = -u_R \end{array} \right\} \quad [4.74]$$

Show that the nature of the waves in the solution depends on the sign of u_L . Provide a complete description of the solution in each case (nature of the waves, expressions for the water level and velocity in each of the regions). Show that a dry zone may appear if u_L is negative, smaller than a given threshold value to be determined. Explain how the symmetry of this Riemann problem may be used to solve boundary value problems, such as the sudden closure of a valve at the upstream and downstream ends of a canal.

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

4.4.2.2. Exercise 4.2: the Euler equations

Consider the Riemann problem formed by the Euler equations with the following initial conditions:

$$\left. \begin{array}{l} p_L = p_R \\ \rho_L = \rho_R \\ u_L = -u_R \end{array} \right\} \quad [4.75]$$

Show that the nature of the waves in the solution depends on the sign of u_L . Provide a complete description of the solution in each case (nature of the waves, expressions for the pressure, velocity and density in each of the regions). Show that a void zone may appear if u_L is negative, smaller than a given threshold value to be determined. Explain how the symmetry of this Riemann problem may be used to solve boundary value problems, such as the sudden closure of a valve in a pipe or the impact of a gas jet onto a wall.

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

Chapter 5

Multidimensional Hyperbolic Systems

5.1. Definitions

5.1.1. Scalar laws

A two-dimensional scalar hyperbolic conservation law is a PDE that can be written in conservation form as:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} + \frac{\partial G}{\partial y} = S \quad [5.1]$$

where F and G are respectively the fluxes in the x - and y -direction and S is the source term. As in the one-dimensional case, F and G are functions of U and, possibly, of x and t , but do not contain functions of any of the derivatives of U .

Equation [5.1] can be written in non-conservation form as:

$$\frac{\partial U}{\partial t} + \lambda_x \frac{\partial U}{\partial x} + \lambda_y \frac{\partial U}{\partial y} = S' \quad [5.2]$$

where λ_x and λ_y are respectively the wave speeds in the x - and y -direction. λ_x , λ_y and S' are given by:

$$\left. \begin{aligned} \lambda_x &= \frac{\partial F}{\partial U} \\ \lambda_y &= \frac{\partial G}{\partial U} \\ S' &= S - \left(\frac{\partial F}{\partial x} \right)_{U=\text{Const}} - \left(\frac{\partial G}{\partial y} \right)_{U=\text{Const}} \end{aligned} \right\} [5.3]$$

Equation [5.2] is equivalent to the following characteristic form:

$$\frac{dU}{dt} = S' \quad \text{for } \left\{ \begin{aligned} dx/dt &= \lambda_x \\ dy/dt &= \lambda_y \end{aligned} \right. [5.4]$$

In the particular case $S' = 0$, equation [5.4] can be integrated into:

$$U = \text{Const} \quad \text{for } \left\{ \begin{aligned} dx/dt &= \lambda_x \\ dy/dt &= \lambda_y \end{aligned} \right. [5.5]$$

Equation [5.1] is extended to three dimensions of space as follows:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} + \frac{\partial G}{\partial y} + \frac{\partial H}{\partial z} = S [5.6]$$

The non-conservation form of equation [5.6] is:

$$\frac{\partial U}{\partial t} + \lambda_x \frac{\partial U}{\partial x} + \lambda_y \frac{\partial U}{\partial y} + \lambda_z \frac{\partial U}{\partial z} = S' [5.7]$$

where λ_x , λ_y , λ_z and S' are given by:

$$\left. \begin{aligned} \lambda_x &= \frac{\partial F}{\partial U}, \lambda_y = \frac{\partial G}{\partial U}, \lambda_z = \frac{\partial H}{\partial U} \\ S' &= S - \left(\frac{\partial F}{\partial x} \right)_{U=\text{Const}} - \left(\frac{\partial G}{\partial y} \right)_{U=\text{Const}} - \left(\frac{\partial H}{\partial z} \right)_{U=\text{Const}} \end{aligned} \right\} [5.8]$$

Equation [5.6] can also be written in characteristic form as:

$$\frac{dU}{dt} = S' \quad \text{for } \begin{cases} dx/dt = \lambda_x \\ dy/dt = \lambda_y \\ dz/dt = \lambda_z \end{cases} \quad [5.9]$$

Note that in the particular case $S' = 0$, equation [5.9] is integrated into:

$$U = \text{Const} \quad \text{for } \begin{cases} dx/dt = \lambda_x \\ dy/dt = \lambda_y \\ dz/dt = \lambda_z \end{cases} \quad [5.10]$$

5.1.2. Two-dimensional hyperbolic systems

A two-dimensional, $m \times m$ system of conservation laws is a system of m PDEs in the form [5.1]:

$$\left. \begin{array}{l} \frac{\partial U_1}{\partial t} + \frac{\partial F_1}{\partial x} + \frac{\partial G_1}{\partial y} = S_1 \\ \vdots \\ \frac{\partial U_m}{\partial t} + \frac{\partial F_m}{\partial x} + \frac{\partial G_m}{\partial y} = S_m \end{array} \right\} \quad [5.11]$$

where U_p, F_p, G_p and S_p ($p = 1, \dots, m$) are respectively the conserved variable, the flux in the x - and y -direction and the source term for the p th equation. In the general case, F_p, G_p and S_p are functions not only of U_p but also of the other conserved variables U_1, \dots, U_m . System [5.11] can be written in vector conservation form as:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} + \frac{\partial \mathbf{G}}{\partial y} = \mathbf{S} \quad [5.12]$$

where $\mathbf{U}, \mathbf{F}, \mathbf{G}$ and \mathbf{S} are defined as:

$$\mathbf{U} = \begin{bmatrix} U_1 \\ \vdots \\ U_m \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} F_1 \\ \vdots \\ F_m \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} G_1 \\ \vdots \\ G_m \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} S_1 \\ \vdots \\ S_m \end{bmatrix} \quad [5.13]$$

Equation [5.12] can be written in non-conservation form as:

$$\frac{\partial U}{\partial t} + A_x \frac{\partial U}{\partial x} + A_y \frac{\partial U}{\partial y} = S' \quad [5.14]$$

where A_x , A_y and S' are given by:

$$\left. \begin{aligned} A_x &= \frac{\partial F}{\partial U} \\ A_y &= \frac{\partial G}{\partial U} \\ S' &= S - \left(\frac{\partial F}{\partial x} \right)_{U=\text{Const}} - \left(\frac{\partial G}{\partial y} \right)_{U=\text{Const}} \end{aligned} \right\} \quad [5.15]$$

System [5.11] is said to be hyperbolic if any linear combination of the matrices A_x and A_y has m real distinct eigenvalues. This definition is justified and interpreted in section 5.3.1.2.

5.1.3. Three-dimensional hyperbolic systems

A three-dimensional, $m \times m$ system of conservation laws is a system of m PDEs in the form [5.6]:

$$\left. \begin{aligned} \frac{\partial U_1}{\partial t} + \frac{\partial F_1}{\partial x} + \frac{\partial G_1}{\partial y} + \frac{\partial H_1}{\partial z} &= S_1 \\ &\vdots \\ \frac{\partial U_m}{\partial t} + \frac{\partial F_m}{\partial x} + \frac{\partial G_m}{\partial y} + \frac{\partial H_m}{\partial z} &= S_m \end{aligned} \right\} \quad [5.16]$$

where U_p , F_p , G_p , H_p and S_p ($p = 1, \dots, m$) are the conserved variable, the flux in the x -, y - and z -direction respectively, and the source term for the p th equation. In the general case, F_p , G_p , H_p and S_p are functions not only of U_p but also of the other conserved variables U_1, \dots, U_m . System [5.16] can be written in vector conservation form as:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} + \frac{\partial G}{\partial y} + \frac{\partial H}{\partial z} = S \quad [5.17]$$

where the vectors U, F, G and S are defined as:

$$U = \begin{bmatrix} U_1 \\ \vdots \\ U_m \end{bmatrix}, \quad F = \begin{bmatrix} F_1 \\ \vdots \\ F_m \end{bmatrix}, \quad G = \begin{bmatrix} G_1 \\ \vdots \\ G_m \end{bmatrix}, \quad H = \begin{bmatrix} H_1 \\ \vdots \\ H_m \end{bmatrix}, \quad S = \begin{bmatrix} S_1 \\ \vdots \\ S_m \end{bmatrix} \quad [5.18]$$

Equation [5.17] can be written in non-conservation form as:

$$\frac{\partial U}{\partial t} + A_x \frac{\partial U}{\partial x} + A_y \frac{\partial U}{\partial y} + A_z \frac{\partial U}{\partial z} = S' \quad [5.19]$$

where A_x , A_y , A_z and S' are defined as:

$$\left. \begin{aligned} A_x &= \frac{\partial F}{\partial U}, A_y = \frac{\partial G}{\partial U}, A_z = \frac{\partial H}{\partial U} \\ S' &= S - \left(\frac{\partial F}{\partial x} \right)_{U=\text{Const}} - \left(\frac{\partial G}{\partial y} \right)_{U=\text{Const}} - \left(\frac{\partial H}{\partial z} \right)_{U=\text{Const}} \end{aligned} \right\} \quad [5.20]$$

The three-dimensional system of conservation laws [5.17] is said to be hyperbolic if any linear combination of the matrices A_x , A_y and A_z has m real, distinct eigenvalues.

5.2. Derivation from conservation principles

This section focuses on the derivation of two-dimensional systems of conservation laws. The generalization to three-dimensional systems is straightforward and will not be detailed hereafter. A system of conservation laws being formed by a set of scalar laws, only the derivation of scalar laws is dealt with hereafter.

Two-dimensional conservation laws are written using a mass balance over a two-dimensional control volume of size $\delta x \times \delta y$ over a time interval δt (Figure 5.1). The balance can be written as:

$$\begin{aligned} \delta U(t_0 + \delta t) - \delta U(t_0) &= \delta F(x_0) - \delta F(x_0 + \delta x) \\ &\quad + \delta G(y_0) - \delta G(y_0 + \delta y) + \delta S \end{aligned} \quad [5.21]$$

where $\delta U(t)$ is the total amount of U contained in the control volume at the time t , $\delta F(x)$ is the total amount of U that crosses the interface of width δy located at the

abscissa x over the time interval δt , $\delta G(y)$ is the total amount of U that crosses the interface of width δx located at the ordinate y over the time interval δt , and δS is the total amount of U that appears within the control volume over the time interval δt owing to the source term. By definition:

$$\left. \begin{aligned} \delta U(t) &= \int_{x_0}^{x_0 + \delta x} \int_{y_0}^{y_0 + \delta y} U(x, y, t) dy dx \\ \delta F(x) &= \int_{t_0}^{t_0 + \delta t} \int_{y_0}^{y_0 + \delta y} F(x, y, t) dy dt \\ \delta G(y) &= \int_{t_0}^{t_0 + \delta t} \int_{x_0}^{x_0 + \delta x} G(x, y, t) dx dt \\ \delta S &= \int_{t_0}^{t_0 + \delta t} \int_{x_0}^{x_0 + \delta x} \int_{y_0}^{y_0 + \delta y} U(x, y, t) dy dx dt \end{aligned} \right\} [5.22]$$

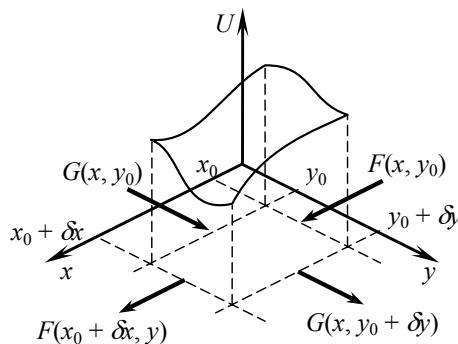


Figure 5.1. Definition sketch for the derivation of a two-dimensional, scalar conservation law

When δx , δy and δt tend to zero, the integral of a given function over a surface or a segment tends to the product of the point value of the function and the surface or the segment length. This leads to the following equivalence for δU :

$$\delta U(t) \underset{\delta x \rightarrow 0}{\underset{\delta y \rightarrow 0}{\approx}} U \delta x \delta y \quad [5.23]$$

and to the following equivalence for δF and δG :

$$\left. \begin{array}{l} \delta F(x) \approx F(x, y) \delta y \delta t \\ \quad \begin{array}{c} \delta x \rightarrow 0 \\ \delta t \rightarrow 0 \end{array} \\ \delta G(y) \approx G(x, y) \delta x \delta t \\ \quad \begin{array}{c} \delta x \rightarrow 0 \\ \delta t \rightarrow 0 \end{array} \end{array} \right\} [5.24]$$

The following equivalence holds for δS :

$$\delta S \approx S \delta x \delta y \delta t \quad \begin{array}{c} \delta x \rightarrow 0 \\ \delta y \rightarrow 0 \\ \delta t \rightarrow 0 \end{array} [5.25]$$

Substituting equations [5.23–25] into equation [5.21] leads to:

$$[U(t_0 + \delta t) - U(t_0)] \delta x \delta y = [F(x_0) - F(x_0 + \delta x)] \delta y \delta t + [G(y_0) - G(y_0 + \delta y)] \delta x \delta t + S \delta x \delta y \delta t \quad [5.26]$$

The differences in equation [5.26] can be expressed as functions of the derivatives of U , F and G with respect to time and space using the following equivalences:

$$\left. \begin{array}{l} U(x, y, t_0 + \delta t) - U(x, y, t_0) \approx \frac{\partial U}{\partial t} \delta t \\ F(x_0 + \delta x, y, t) - F(x_0, y, t) \approx \frac{\partial F}{\partial x} \delta x \\ G(x, y_0 + \delta y, t) - G(x, y_0, t) \approx \frac{\partial G}{\partial y} \delta y \end{array} \right\} [5.27]$$

Substituting equations [5.27] into equation [5.26] and simplifying leads to equation [5.1].

Note that the assumption of a continuous and differentiable solution is not necessary in the derivation of the weak form [5.21], while equation [5.1] is based on such an assumption. The assumption of a differentiable variable is used in equation [5.27] to introduce the derivatives of U , F and G with respect to time and space. As in the one-dimensional case, the “strong form” [5.1] is a particular case of the weak form [5.21]. As in the one-dimensional case, the weak solutions of equation [5.1] may be non-unique. These remarks also hold for three-dimensional scalar laws and for hyperbolic systems of conservation laws.

5.3. Solution properties

5.3.1. Two-dimensional hyperbolic systems

5.3.1.1. The bicharacteristic approach

This section deals with the properties of the solutions of two-dimensional hyperbolic systems of conservation laws, that is, systems in the form [5.12]. Several approaches may be used to characterize the behavior and properties of the solutions of such systems. The bicharacteristic approach, also called the characteristic surface approach, is one of them.

As in the one-dimensional case, the purpose is to find surfaces in the phase space (x, y, t) , over which certain quantities are invariant (Figure 5.2). The knowledge of the solution U at a given point (x_0, y_0) in the characteristic surface should allow the value of U to be computed at any other point that belongs to the surface. Conversely, a given characteristic surface cannot provide any information about the value of U at a point that does not belong to it. This property is used in the derivation of the characteristic form.

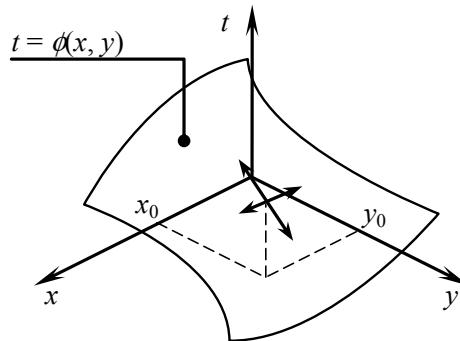


Figure 5.2. Definition sketch for a characteristic surface in the phase space for a two-dimensional hyperbolic system

Assume that U is known at a point (x_0, y_0, t_0) in the phase space. The value of U at this point is denoted by U_0 . U verifies the conservation form [5.12] and its non-conservation form [5.14]. The equation of the characteristic surface that passes at (x_0, y_0, t_0) is sought in the form:

$$t = \phi(x, y) \quad [5.28]$$

Assume that U is known over all the surface. The value of U at the point $(x, y, \phi(x, y))$ is denoted by $U_s(x, y)$. The differential dU_s is defined as:

$$dU_s = \frac{\partial U}{\partial t} d\phi + \frac{\partial U}{\partial x} dx + \frac{\partial U}{\partial y} dy \quad [5.29]$$

The differential $d\phi$ is given by:

$$d\phi = \frac{\partial \phi}{\partial x} dx + \frac{\partial \phi}{\partial y} dy \quad [5.30]$$

Substituting equation [5.29] into equation [5.30] leads to:

$$dU_s = \left(\frac{\partial \phi}{\partial x} dx + \frac{\partial \phi}{\partial y} dy \right) \frac{\partial U}{\partial t} + \frac{\partial U}{\partial x} dx + \frac{\partial U}{\partial y} dy \quad [5.31]$$

Hence the derivatives of U_s with respect to x and y :

$$\begin{cases} \frac{\partial U_s}{\partial x} = \frac{\partial \phi}{\partial x} \frac{\partial U}{\partial t} + \frac{\partial U}{\partial x} \\ \frac{\partial U_s}{\partial y} = \frac{\partial \phi}{\partial y} \frac{\partial U}{\partial t} + \frac{\partial U}{\partial y} \end{cases} \quad [5.32]$$

The system [5.32] can be rewritten as:

$$\begin{cases} \frac{\partial U}{\partial x} = \frac{\partial U_s}{\partial x} - \frac{\partial \phi}{\partial x} \frac{\partial U}{\partial t} \\ \frac{\partial U}{\partial y} = \frac{\partial U_s}{\partial y} - \frac{\partial \phi}{\partial y} \frac{\partial U}{\partial t} \end{cases} \quad [5.33]$$

Substituting equation [5.33] into the non-conservation form [5.14] leads to:

$$\left(I - \frac{\partial \phi}{\partial x} A_x - \frac{\partial \phi}{\partial y} A_y \right) \frac{\partial U}{\partial t} = S' - A_x \frac{\partial U_s}{\partial x} - A_y \frac{\partial U_s}{\partial y} \quad [5.34]$$

where I is the $m \times m$ identity matrix. U may be calculated at any point that does not belong to the surface $t = \phi(x, y)$ using a first-order expansion:

$$U(x, y, t) = U_s(x, y) + [\phi(x, y) - t] \frac{\partial U}{\partial t} \quad [5.35]$$

U cannot be calculated at a point that does not belong to the characteristic surface if $\partial U / \partial t$ cannot be computed. From equation [5.34], such a condition is equivalent to stating that the matrix $I - \partial\phi/\partial x A_x - \partial\phi/\partial y A_y$ has no inverse, that is, if its determinant is equal to zero:

$$\left| I - \frac{\partial\phi}{\partial x} A_x - \frac{\partial\phi}{\partial y} A_y \right| = 0 \quad [5.36]$$

Equation [5.36] provides a necessary condition on the x - and y -slopes of the surfaces $t = \phi(x, y)$. As shown in section 5.4, such surfaces may be approximated locally with cones or straight lines in the phase space.

The equation of the characteristic surfaces allows the generalization of the Riemann invariants to be derived along the surfaces. Equation [5.36] means that at least one of the rows in the matrix $I - \partial\phi/\partial x A_x - \partial\phi/\partial y A_y$ is a linear combination of the others. In other words, there exists a row vector v such that:

$$\left. \begin{aligned} v \left(I - \frac{\partial\phi}{\partial x} A_x - \frac{\partial\phi}{\partial y} A_y \right) \frac{\partial U}{\partial t} - v \left(S' - A_x \frac{\partial U_s}{\partial x} - A_y \frac{\partial U_s}{\partial y} \right) = 0 \\ v \left(I - \frac{\partial\phi}{\partial x} A_x - \frac{\partial\phi}{\partial y} A_y \right) = 0 \end{aligned} \right\} \quad [5.37]$$

Substituting equation [5.32] into the first equation [5.37] leads to:

$$v \left(\frac{\partial U}{\partial t} - A_x \frac{\partial U}{\partial x} - A_y \frac{\partial U}{\partial y} - S' \right) = 0 \quad [5.38]$$

which leads to a differential relationship between the various components of U .

5.3.1.2. The secant plane approach

The classical equations of fluid mechanics are invariant by rotation. The secant plane approach consists of finding the characteristic form of the restriction of the equations to a plane parallel to the time axis for any value of the angle θ between the plane and the x -axis (Figure 5.3). For a given value of θ , the intersection between the plane and the characteristic surface gives one or several characteristic curves. As in the one-dimensional case, Riemann invariants can be derived along these curves.

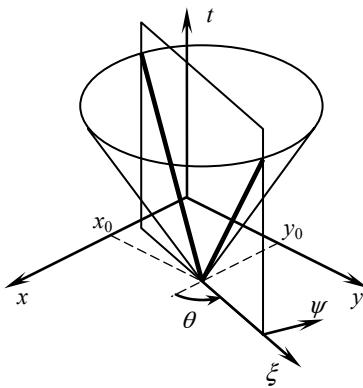


Figure 5.3. Definition sketch for the secant plane approach. Secant plane and characteristic surface (thin lines), characteristic curves in the secant planes (bold lines)

A local coordinate system (ξ, ψ) is attached to the secant plane (Figure 5.3). The following relationships hold between the local and global coordinate systems:

$$\left. \begin{aligned} d\xi &= \cos \theta \, dx + \sin \theta \, dy \\ d\psi &= -\sin \theta \, dx + \cos \theta \, dy \end{aligned} \right\} \quad [5.39]$$

The partial derivatives with respect to x and y are related to those with respect to ξ and ψ via:

$$\left. \begin{aligned} \frac{\partial U}{\partial x} &= \frac{\partial U}{\partial \xi} \frac{\partial \xi}{\partial x} + \frac{\partial U}{\partial \psi} \frac{\partial \psi}{\partial x} = \cos \theta \frac{\partial U}{\partial \xi} - \sin \theta \frac{\partial U}{\partial \psi} \\ \frac{\partial U}{\partial y} &= \frac{\partial U}{\partial \xi} \frac{\partial \xi}{\partial y} + \frac{\partial U}{\partial \psi} \frac{\partial \psi}{\partial y} = \sin \theta \frac{\partial U}{\partial \xi} + \cos \theta \frac{\partial U}{\partial \psi} \end{aligned} \right\} \quad [5.40]$$

Substituting equation [5.40] into equation [5.14] leads to the following equation:

$$\frac{\partial U}{\partial t} + (\cos \theta A_x + \sin \theta A_y) \frac{\partial U}{\partial \xi} + (-\sin \theta A_x + \cos \theta A_y) \frac{\partial U}{\partial \psi} = S' \quad [5.41]$$

Equation [5.41] can be rewritten as a one-dimensional equation in the direction of the secant plane:

$$\frac{\partial U}{\partial t} + A_\xi \frac{\partial U}{\partial \xi} = S_\xi \quad [5.42]$$

where A_ξ and A_ψ are defined as

$$\left. \begin{aligned} A_\xi &= \cos \theta A_x + \sin \theta A_y \\ S_\xi &= S' + (\sin \theta A_x - \cos \theta A_y) \frac{\partial U}{\partial \psi} \end{aligned} \right\} [5.43]$$

If A_ξ has m real, distinct eigenvalues for all θ , the system [5.42] is hyperbolic regardless of the orientation of the secant plane. The matrix A_ξ is a linear combination of the matrices A_x and A_y . It spans all the possible linear combinations of A_x and A_y as θ spans the interval $[0, 2\pi]$. The definition of a hyperbolic system as given in section 5.1.2 is justified as follows: system [5.14] is hyperbolic if its one-dimensional restriction [5.42] to all the possible secant planes is hyperbolic.

As in the one-dimensional case (see Chapter 2), equation [5.40] allows Riemann invariants to be defined in the secant plane. The vector W is defined in differential form as:

$$dW = K_\xi^{-1} dU [5.44]$$

where K_ξ is the matrix formed by the eigenvectors of A_ξ . Equation [5.42] is rewritten as:

$$\frac{\partial W}{\partial t} + \Lambda_\xi \frac{\partial W}{\partial \xi} = K_\xi^{-1} S_\xi [5.45]$$

where Λ_ξ is the diagonal matrix formed with the eigenvalues of A_ξ . Equation [5.45] is equivalent to:

$$\frac{dW_p}{dt} = K_\xi^{-1} S_\xi \quad \text{for } \frac{d\xi}{dt} = \lambda_\xi^{(p)} [5.46]$$

5.3.1.3. Domain of influence, domain of dependence

A hyperbolic system of conservation laws leads to several characteristic surfaces in the general case. The domain of influence of the solution is contained within the characteristic surface, the spatial extent of which is the largest. The domain of influence includes not only the characteristic surface but also all the points contained in the volume delineated by the surface. Such a surface is illustrated by Figure 5.4 in

the phase space. For the sake of clarity, only one family of characteristic surfaces is sketched in the figure.

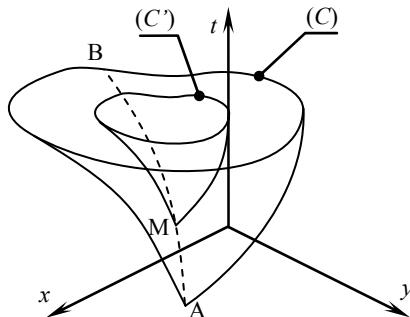


Figure 5.4. Definition sketch for the domain of influence in the phase space

When the equations are nonlinear the speeds of the various waves are variable in the phase space. This accounts for the curvature of the surfaces issued from A and M in Figure 5.4. Consider the characteristic surface issued from A. Point A influences all the points M of all the curves (AB) that belong to the surface. Several characteristic surfaces are issued from point B (for the sake of clarity, only the widest surface is drawn). The intersection between the characteristic surface issued from A and the plane ($t = t_B$) is a closed contour denoted by (C) in the figure. The intersection between the characteristic surface issued from M and the plane ($t = t_B$) is a closed contour denoted by (C') in the figure.

When $M = A$, $(C') = (C)$. When $M = B$, $(C') = B$. The size of (C') decreases gradually as M spans all the possible locations from A to B along the line (AB). Consequently, the domain of influence of A includes the curve (C) and all its inner points. Since this is true for any time t_B , the domain of influence of the solution includes all the points in the volume delineated by the widest of all the existing characteristic surfaces issued from A.

A similar reasoning leads to the conclusion that the domain of dependence of the solution at the point B is made of all the points contained in the volume that is delimited by the widest of all the characteristic surfaces passing at B (Figure 5.5). For the sake of clarity, only the widest characteristic surface (C) is represented in Figure 5.5. The characteristic surface is formed by an infinity of generating curves (AB), where the points A are points in the plane ($t = t_A$), $t_A < t_B$. The intersection between the characteristic surface and the plane ($t = t_A$) is a closed curve denoted by (C).

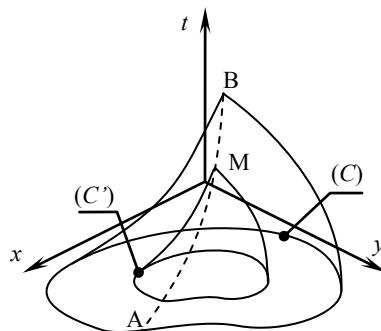


Figure 5.5. Definition sketch for the domain of dependence in the phase space

Consider a point M located along the curve (AB). Several characteristic surfaces pass at M. Only the widest one is sketched in the figure for the sake of clarity. The intersection between the surface and the plane ($t = t_A$) is a closed curve ((C')). $(C') = (C)$ for $M = B$, while $(C') = A$ for $M = A$. The curve (C') spans all the inner points of the curve (C) as the point M spans all the possible locations along the curve (AB). Consequently, all the points enclosed in (C) influence at least one point along the curve (AB), thus influencing indirectly the solution at the point B. Since this is true for all possible values of t_A , the domain of dependence of the solution at B includes the most extended characteristic surface that passes at B as well as all the points contained in the volume delineated by the surface.

5.3.2. Three-dimensional hyperbolic systems

In the case of a three-dimensional hyperbolic system, the phase space is four-dimensional. The characteristic surface approach leads us to seek a hyper-surface defined as:

$$t = \phi(x, y, z) \quad [5.47]$$

from which the solution cannot be calculated at points that do not belong to the surface. Applying the same reasoning as in section 5.3.1.1, the following condition is derived for the function ϕ .

$$\left| I - \frac{\partial \phi}{\partial x} A_x - \frac{\partial \phi}{\partial y} A_y - \frac{\partial \phi}{\partial z} A_z \right| = 0 \quad [5.48]$$

If the secant plane approach is to be used, a local coordinate system (ξ, ψ, ζ) is defined, that is obtained from the original coordinate system by applying a first rotation of angle θ in the (x, y) plane, followed by a second rotation in the thus obtained (ξ, z) plane (Figure 5.6). The two coordinate systems are related by the following differentials:

$$\left. \begin{aligned} d\xi &= \cos \varphi \cos \theta \, dx + \cos \varphi \sin \theta \, dy + \sin \varphi \, dz \\ d\psi &= -\sin \theta \, dx + \cos \theta \, dy \\ d\zeta &= -\sin \varphi \cos \theta \, dx - \sin \varphi \sin \theta \, dy + \cos \varphi \, dz \end{aligned} \right\} [5.49]$$

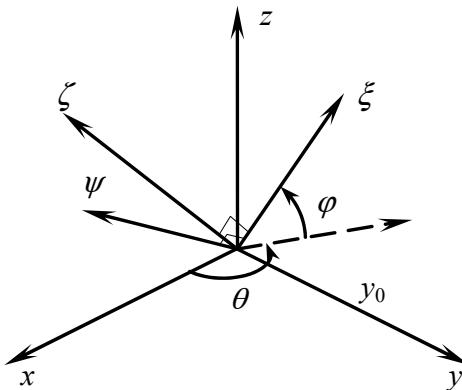


Figure 5.6. Definition sketch for the local coordinate system in the secant plane approach as applied to three-dimensional hyperbolic systems

Equation [5.19] can be rewritten in the form [5.42] by defining A_ξ and S_ξ as:

$$\left. \begin{aligned} A_\xi &= \cos \varphi \cos \theta \, A_x + \cos \varphi \sin \theta \, A_y + \sin \varphi \, A_z \\ S_\xi &= S' + (\sin \theta A_x - \cos \theta A_y) \frac{\partial U}{\partial \psi} \\ &\quad + [(\cos \theta A_x + \sin \theta A_y) \sin \varphi + \cos \varphi A_z] \frac{\partial U}{\partial \zeta} \end{aligned} \right\} [5.50]$$

The non-conservation form [5.45] and the characteristic form [5.46] are applicable in the local coordinate system.

5.4. Application: the two-dimensional shallow water equations

5.4.1. Governing equations

5.4.1.1. Physical context – assumptions

The two-dimensional free-surface flow equations, also called two-dimensional shallow water equations, can be viewed as a two-dimensional extension of the Saint Venant equations. They are often used in floodplain modeling studies or coastal modeling, where the classical, one-dimensional Saint Venant equations do not suffice to provide a correct description of the flow. This is the case in particular when sharp contrasts appear in the velocity field or in the water depths (Figure 5.7).

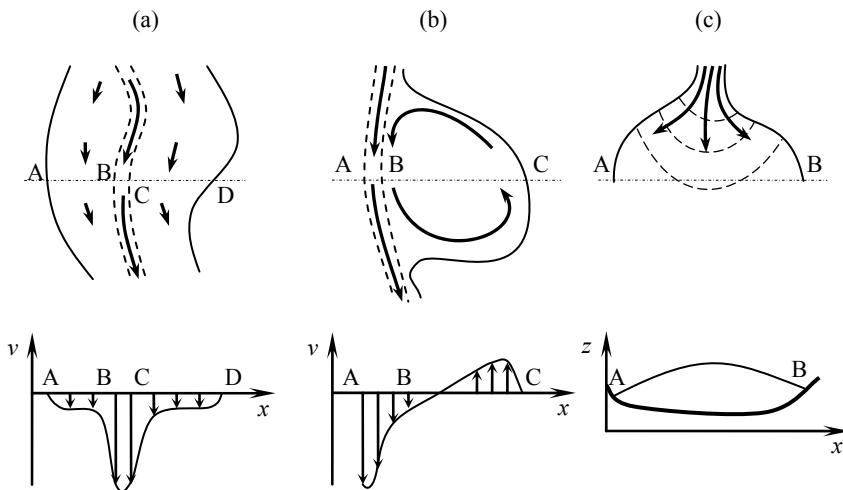


Figure 5.7. Typical situations where the one-dimensional approximation is invalid.
Plan view (top), cross-sectional view (bottom)

Figure 5.7 illustrates three typical situations where the one-dimensional approach is invalid:

- A wide floodplain, where the water depth is small compared to that in the main channel (Figure 5.7a), invalidates the one-dimensional assumption [CUN 80]. If a uniform water level is assumed over the cross-section, the difference between the depth in the main channel (indicated by a dashed line in Figure 5.7a) and the floodplain (solid lines) leads to different values for the flow velocity because the friction terms are larger in the floodplain than in the main channel. If a uniform velocity field is assumed across the entire section, the slope of the energy line cannot be the same in the floodplain and in the main channel because the depth is

not the same. Therefore the water depth cannot be the same in the main channel and in the floodplain.

b) Presence of a storage pocket in the floodplain (Figure 5.7b). The pocket may be filled as a result of water stage increase during a flood. As a result, a swirling zone, also called a dead zone, appears in the pocket. The swirl appears as a consequence of lateral momentum diffusion due to viscosity and turbulent diffusion. The direction of the flow changes along the segment [BC] that is drawn across the pocket. Under steady state conditions, or near the peak flow, the amount of water stored in the pocket is nearly constant, which means that the average discharge in the pocket is zero. In other words, the pocket does not participate in the dynamics of the river. Most of the discharge is transferred via the main channel [AB]. Neglecting this by assuming a uniform flow velocity over the whole cross-section [AC] may lead to tremendous error in the assessment of the friction term and in the subsequently calculated discharge and water depth.

c) Propagation of a flood wave over rapidly varying geometries (Figure 5.7c). A transient propagating into a region where the channel geometry varies strongly is subjected to two-dimensional effects. As mentioned in sections 5.4.2 and 5.4.3, multidimensional waves tend to expand in the radial direction. The elevation of the free surface and the velocity profile are therefore not constant along a cross-section when a wave enters a sudden widening as sketched in Figure 5.7c.

The governing equations for two-dimensional free surface flow are derived from the following assumptions:

- Assumption (A1). The water is assumed to be incompressible in the range of ordinary pressure and water levels. The density is constant.
- Assumption (A2). The vertical acceleration of the water molecules is negligible compared to the horizontal acceleration. The pressure field is considered to be hydrostatic.
- Assumption (A3). The flow is turbulent. The head loss due to bottom friction is proportional to the square of the velocity.
- Assumption (A4). The diffusion of momentum due to turbulence and viscosity, the Coriolis effect and the shear stress due to the wind are neglected.

Note that Assumption (A4) is introduced only for the sake of simplicity in the treatment of the equations, the main purpose being to focus on the hyperbolic part of the equations. It should be kept in mind however that momentum diffusion, Coriolis forces and wind-induced forces are taken into account in most simulation packages nowadays.

5.4.1.2. Continuity equation

The continuity equation is derived as explained in section 5.2. The conserved variable is the mass per unit length, the fluxes are the mass discharges per unit width in the x - and y -direction and the source term is zero:

$$\left. \begin{array}{l} U = \rho h \\ F = \rho h u \\ G = \rho h v \\ S = 0 \end{array} \right\} \quad [5.51]$$

where h is the water depth, u and v are the x - and y -velocity respectively and ρ is the (constant) water density. Substituting definitions [5.51] into equation [5.1] and using Assumption (A1) leads to the following PDE:

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial x}(hu) + \frac{\partial}{\partial y}(hv) = 0 \quad [5.52]$$

5.4.1.3. Equation for the momentum in the x -direction

The momentum equation is derived using the x -momentum per unit surface as the conserved variable:

$$U = \rho h u \quad [5.53]$$

The flux F in the x -direction results from two terms: (i) the inertial term, that accounts for the advection of the conserved variable U at a speed u in the x -direction, and (ii) the pressure force, expressed as the integral of the pressure between the bottom and the free surface (see sections 2.4.2.3 and 2.5.2.3 for the details of the proof):

$$F = \rho h u^2 + P \quad [5.54]$$

Note that Assumption (A2) leads to a hydrostatic pressure field, which leads to:

$$P = \int_0^h p(\eta) d\eta = \int_0^h \rho g \eta d\eta = \rho g \frac{h^2}{2} \quad [5.55]$$

where g is the gravitational acceleration and η is the distance to the free surface. Substituting equation [5.55] into equation [5.54] leads to:

$$F = \left(hu^2 + \frac{gh^2}{2} \right) \rho \quad [5.56]$$

The flux G in the y -direction accounts for the transport of the conserved variable $U = \rho hu$ at a speed v in the y -direction. G is expressed as:

$$G = \rho huv \quad [5.57]$$

The source term expresses the influence of two factors, namely the friction force $R_{f,x}$ against the bottom and the projection of the reaction of the bottom onto the x -axis. According to Assumption (A3), $R_{f,x}$ is proportional to the square of the intensity of the velocity vector; its orientation is opposite to that of the flow. As in the Saint Venant equations, the friction force is expressed so as to involve the friction slope $S_{f,x}$ in the x -direction:

$$R_{f,x} = \rho g h S_{f,x} \quad [5.58]$$

The friction slope is estimated using the two-dimensional generalization of the Chezy, Strickler or Manning formulations (see equations [2.107–108])

$$\begin{aligned} S_{f,x} &= \frac{(u^2 + v^2)^{1/2} u}{C^2 R_H} && \text{(Chézy)} \\ S_{f,x} &= \frac{(u^2 + v^2)^{1/2} u}{K_{\text{Str}}^2 R_H^{4/3}} && \text{(Strickler)} \\ S_{f,x} &= n_M^2 \frac{(u^2 + v^2)^{1/2} u}{R_H^{4/3}} && \text{(Manning)} \end{aligned} \quad [5.59]$$

The projection of the reaction of the bottom onto the x -axis is derived exactly as in the Saint Venant equations (see section 2.5.2.3, equations [2.112–113]):

$$R_{0,x} = -\rho g h \frac{\partial z_b}{\partial x} = \rho g h S_{0,x} \quad [5.60]$$

Substituting equations [5.53], [5.56–58] and [5.60] into equation [5.1], dividing by the density yields the x -momentum equation:

$$\frac{\partial}{\partial t}(hu) + \frac{\partial}{\partial x}\left(hu^2 + g \frac{h^2}{2}\right) + \frac{\partial}{\partial y}(huv) = (S_{0,x} - S_{f,x})gh \quad [5.61]$$

5.4.1.4. Equation for the y -momentum

The y -momentum equation is derived exactly in the same way as the x -momentum equation. Reproducing the reasoning of section 5.4.1.3 leads to the following equation:

$$\frac{\partial}{\partial t}(hv) + \frac{\partial}{\partial x}(huv) + \frac{\partial}{\partial y}\left(hv^2 + g\frac{h^2}{2}\right) = (S_{0,y} - S_{f,y})gh \quad [5.62]$$

where the friction slope in the y -direction is given by:

$$\begin{aligned} S_{f,y} &= \frac{(u^2 + v^2)^{1/2}v}{C^2 R_H} && \text{(Chezy)} \\ S_{f,y} &= \frac{(u^2 + v^2)^{1/2}v}{K_{su}^2 R_H^{4/3}} && \text{(Strickler)} \\ S_{f,y} &= n_M^2 \frac{(u^2 + v^2)^{1/2}v}{R_H^{4/3}} && \text{(Manning)} \end{aligned} \quad [5.63]$$

5.4.1.5. Vector form

System [5.52], [5.61], [5.62] can be written in vector conservation form as in equation [5.12], recalled here:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} + \frac{\partial \mathbf{G}}{\partial y} = \mathbf{S}$$

with the following definitions for \mathbf{U} , \mathbf{F} , \mathbf{G} and \mathbf{S} :

$$\begin{aligned} \mathbf{U} &= \begin{bmatrix} h \\ hu \\ hv \end{bmatrix}, & \mathbf{F} &= \begin{bmatrix} hu \\ hu^2 + gh^2/2 \\ huv \end{bmatrix}, \\ \mathbf{G} &= \begin{bmatrix} hv \\ huv \\ hv^2 + gh^2/2 \end{bmatrix}, & \mathbf{S} &= \begin{bmatrix} 0 \\ (S_{0,x} - S_{f,x})gh \\ (S_{0,y} - S_{f,y})gh \end{bmatrix} \end{aligned} \quad [5.64]$$

The non-conservation form [5.14], recalled here:

$$\frac{\partial \mathbf{U}}{\partial t} + \mathbf{A}_x \frac{\partial \mathbf{U}}{\partial x} + \mathbf{A}_y \frac{\partial \mathbf{U}}{\partial y} = \mathbf{S}'$$

is obtained with $\mathbf{S} = \mathbf{S}'$ and:

$$\left. \begin{aligned} \mathbf{A}_x &= \frac{\partial \mathbf{F}}{\partial \mathbf{U}} = \begin{bmatrix} 0 & 1 & 0 \\ c^2 - u^2 & 2u & 0 \\ -uv & v & u \end{bmatrix} \\ \mathbf{A}_y &= \frac{\partial \mathbf{G}}{\partial \mathbf{U}} = \begin{bmatrix} 0 & 0 & 1 \\ -uv & v & u \\ c^2 - v^2 & 0 & 2v \end{bmatrix} \end{aligned} \right\} [5.65]$$

where $c = (gh)^{1/2}$ is the speed of the waves in still water.

In what follows the characteristic surfaces and the Riemann invariants are derived using the secant plane approach. The bicharacteristic approach is described in detail in [DAU 67] and will not be detailed here.

5.4.2. The secant plane approach

5.4.2.1. Characteristic surfaces

As shown in section 5.3.1.2, the secant plane approach consists of using a projection of the equations onto a secant plane so as to obtain a one-dimensional equation in the form [5.42], recalled here:

$$\frac{\partial \mathbf{U}}{\partial t} + \mathbf{A}_\xi \frac{\partial \mathbf{U}}{\partial \xi} = \mathbf{S}_\xi$$

where \mathbf{A}_ξ and \mathbf{S}_ξ are defined as in equation [5.43], recalled here:

$$\left. \begin{aligned} \mathbf{A}_\xi &= \cos \theta \mathbf{A}_x + \sin \theta \mathbf{A}_y \\ \mathbf{S}_\xi &= \mathbf{S}' - (\sin \theta \mathbf{A}_x + \cos \theta \mathbf{A}_y) \frac{\partial \mathbf{U}}{\partial \psi} \end{aligned} \right\}$$

Substituting the definitions [5.65] into equation [5.43] yields:

$$A_\xi = \begin{bmatrix} 0 & \cos \theta & \sin \theta \\ (c^2 - u^2) \cos \theta - uv \sin \theta & 2u \cos \theta + v \sin \theta & u \sin \theta \\ (c^2 - v^2) \sin \theta - uv \cos \theta & v \cos \theta & u \cos \theta + 2v \sin \theta \end{bmatrix} \quad [5.66]$$

The components of the velocity in the ξ - and ψ -direction are introduced:

$$\left. \begin{aligned} u_\xi &= u \cos \theta + v \sin \theta \\ u_\psi &= -u \sin \theta + v \cos \theta \end{aligned} \right\} \quad [5.67]$$

Using definitions [5.67], the expression of A_ξ can be simplified into:

$$A_\xi = \begin{bmatrix} 0 & \cos \theta & \sin \theta \\ c^2 \cos \theta - uu_\xi & u \cos \theta + u_\xi & u \sin \theta \\ c^2 \sin \theta - vu_\xi & v \cos \theta & v \sin \theta + u_\xi \end{bmatrix} \quad [5.68]$$

The eigenvalues λ of A_ξ verify:

$$|A_\xi - \lambda I| = 0 \quad [5.69]$$

which gives the characteristic equation:

$$(\lambda - u_\xi)[(\lambda - u_\xi)^2 - c^2] = 0 \quad [5.70]$$

Equation [5.70] has the following solutions:

$$\left. \begin{aligned} \lambda^{(1)} &= u_\xi - c \\ \lambda^{(2)} &= u_\xi \\ \lambda^{(3)} &= u_\xi + c \end{aligned} \right\} \quad [5.71]$$

The surface associated with the first eigenvalue is defined as:

$$\frac{d\xi}{dt} = \lambda^{(1)} = u_\xi - c \quad [5.72]$$

that is:

$$\cos \theta \frac{dx}{dt} + \sin \theta \frac{dy}{dt} = \cos \theta u + \sin \theta v - c \quad [5.73]$$

The angle θ between the secant plane and the x -axis is eliminated from equation [5.73] by first differentiating equation [5.73] with respect to θ .

$$-\left(\frac{dx}{dt} - u\right) \sin \theta + \left(\frac{dy}{dt} - v\right) \cos \theta = 0 \quad [5.74]$$

and raising equations [5.73] and [5.74] to the square. The difference between the resulting equations leads to:

$$\left(\frac{dx}{dt} - u\right)^2 + \left(\frac{dy}{dt} - v\right)^2 = c^2 \quad [5.75]$$

Equation [5.75] is the equation of a circle in the plane (x, y) . The center of the circle moves at a speed (u, v) and the radius of the circle grows at a speed c (Figure 5.8). Note that this surface is identical to that associated with the eigenvalue $\lambda^{(3)}$ because changing $-c$ into $+c$ leads identically to equation [5.75]. Reasoning along the same line, the characteristic surface associated with the eigenvalue $\lambda^{(2)}$ is easily shown to be the straight line along which the center of the circle moves.

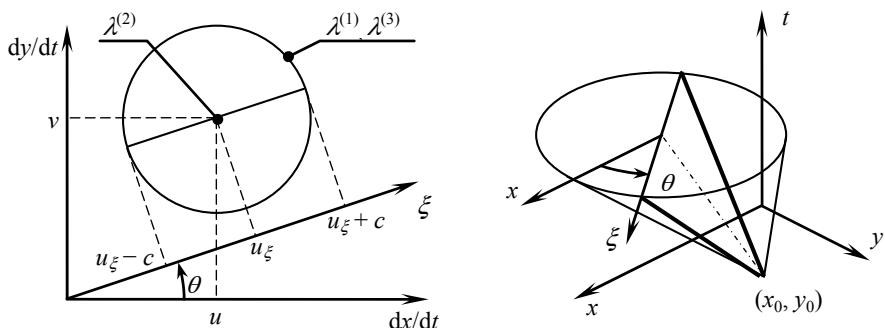


Figure 5.8. Bicharacteristic surfaces as derived using the secant plane approach.
Plan view (left), perspective view (right)

5.4.2.2. Derivation of the Riemann invariants

The Riemann invariants are derived using equation [5.42]. The eigenvectors of \mathbf{A}_ξ are:

$$\mathbf{K}_\xi^{(1)} = \begin{bmatrix} 1 \\ u - c \cos \theta \\ v - c \sin \theta \end{bmatrix}, \quad \mathbf{K}_\xi^{(2)} = \begin{bmatrix} 0 \\ -\sin \theta \\ \cos \theta \end{bmatrix}, \quad \mathbf{K}_\xi^{(3)} = \begin{bmatrix} 1 \\ u + c \cos \theta \\ v + c \sin \theta \end{bmatrix} \quad [5.76]$$

The inverse of the matrix \mathbf{K} formed by the eigenvectors is:

$$\mathbf{K}_\xi^{-1} = \begin{bmatrix} \frac{u_\xi + c}{2c} & -\frac{\cos \theta}{2c} & -\frac{\sin \theta}{2c} \\ -u_\psi & -\sin \theta & \cos \theta \\ \frac{-u_\xi + c}{2c} & \frac{\cos \theta}{2c} & \frac{\sin \theta}{2c} \end{bmatrix} \quad [5.77]$$

Consequently, the vector \mathbf{W} is defined by the differential:

$$d\mathbf{W} = \mathbf{K}_\xi^{-1} d\mathbf{U} = \frac{1}{2c} \begin{bmatrix} (u_\xi + c)dh - \cos \theta d(uh) - \sin \theta dv \\ [-u_\psi dh - \sin \theta d(uh) + \cos \theta dv]2c \\ (-u_\xi + c)dh + \cos \theta d(uh) + \sin \theta dv \end{bmatrix} \quad [5.78]$$

Equation [5.78] is simplified using the reciprocal of equations [5.67]:

$$\left. \begin{aligned} \cos \theta du + \sin \theta dv &= du_\xi \\ -\sin \theta du + \cos \theta dv &= du_\psi \end{aligned} \right\} \quad [5.79]$$

Substituting equations [5.79] into equation [5.78] and differentiating the equality $h = c^2/g$ into $dh = 2c \, dc/g$ leads to:

$$d\mathbf{W} = \mathbf{K}_\xi^{-1} d\mathbf{U} = \frac{1}{2c} \begin{bmatrix} (u_\xi + c)dh - d(hu_\xi) \\ -u_\psi dh + d(hu_\psi) \\ (-u_\xi + c)dh + d(hu_\xi) \end{bmatrix} = \frac{1}{2cg} \begin{bmatrix} -c^2 d(u_\xi - 2c) \\ 2c^2 du_\psi \\ c^2 d(u_\xi + 2c) \end{bmatrix} \quad [5.80]$$

The source term S_ξ is given by the second equation [5.43], recalled here:

$$S_\xi = S' + (\sin \theta A_x - \cos \theta A_y) \frac{\partial U}{\partial \psi}$$

Expanding the terms $d(uh)$, $d(vh)$, using $dh = c^2/g$, substituting equations [5.67] and rewriting equations [5.39] as:

$$\left. \begin{aligned} \cos \theta \frac{\partial}{\partial x} + \sin \theta \frac{\partial}{\partial y} &= \frac{\partial}{\partial \xi} \\ -\sin \theta \frac{\partial}{\partial x} + \cos \theta \frac{\partial}{\partial y} &= \frac{\partial}{\partial \psi} \end{aligned} \right\} \quad [5.81]$$

leads to:

$$S_\xi = \frac{1}{2c} \left[\begin{aligned} & -\frac{2c^2}{g} u_\psi \frac{\partial c}{\partial \psi} - \frac{c^3}{g} \frac{\partial u_\psi}{\partial \psi} \\ & + \frac{c^2}{g} u_\psi \frac{\partial u_\xi}{\partial \psi} + (S_{0,\xi} - S_{f,\xi})c^2 \\ & \frac{4c^3}{g} \frac{\partial c}{\partial \psi} + \frac{2c^2}{g} u_\psi \frac{\partial u_\psi}{\partial \psi} - 2(S_{0,\psi} - S_{f,\psi})c^2 \\ & \frac{2c^2}{g} u_\psi \frac{\partial c}{\partial \psi} + \frac{c^3}{g} \frac{\partial u_\psi}{\partial \psi} \\ & - \frac{c^2}{g} u_\psi \frac{\partial u_\xi}{\partial \psi} - (S_{0,\xi} - S_{f,\xi})c^2 \end{aligned} \right] \quad [5.82]$$

where $S_{0,\xi}$ and $S_{f,\xi}$ are respectively the bottom and friction slope in the ξ -direction, and $S_{0,\psi}$ and $S_{f,\psi}$ are respectively the bottom and friction slope in the ψ -direction. Substituting equations [5.80] and [5.82] into equation [5.42] yields the following differential relationships:

$$\left. \begin{array}{l} \frac{d}{dt}(u_\xi - 2c) = S_{\xi,1} \quad \text{for } \frac{d_\xi}{dt} = u_\xi - c \\ \frac{du_\psi}{dt} = S_{\xi,2} \quad \text{for } \frac{d_\xi}{dt} = u_\xi \\ \frac{d}{dt}(u_\xi + 2c) = S_{\xi,3} \quad \text{for } \frac{d_\xi}{dt} = u_\xi + c \end{array} \right\} [5.83]$$

where the components of the source term S_ξ are given by:

$$\left. \begin{array}{l} S_{\xi,1} = S_{\xi,3} = -u_\psi \frac{\partial u_\xi}{\partial \psi} + c \frac{\partial u_\psi}{\partial \psi} + 2u_\psi \frac{\partial c}{\partial \psi} - (S_{0,\xi} - S_{f,\xi})g \\ S_{\xi,2} = u_\psi \frac{\partial u_\psi}{\partial \psi} + 2c \frac{\partial c}{\partial \psi} - (S_{0,\psi} - S_{f,\psi})g \end{array} \right\} [5.84]$$

Note that the coordinate system (ξ, ψ) attached to the secant plane coincides with the global coordinate system (x, y) for $\theta = 0$. Then u_ξ and u_ψ coincide with u and v respectively and the first and third Riemann invariants coincide with the classical invariants $u - 2c$ and $u + 2c$ derived in section 2.5.3.3 for the one-dimensional flow equations in rectangular channels.

5.4.3. Interpretation – determination of the solution

5.4.3.1. Domain of influence, domain of dependence

The secant plane approach leads to the same result as the bicharacteristic approach developed in [DAU 67]. There are two characteristic surfaces in the phase space. The first surface is generated by a circle, the center of which moves at the speed of the flow and the radius of which expands at the propagation speed of the waves in still water. The second surface is the line generated by the successive locations of the center of the circle. The wave speeds $u - c$ and $u + c$ and the Riemann invariants $u - 2c$ and $u + 2c$ are recovered when u is redefined as the speed in the local coordinate ξ attached to the secant plane. This is because the shallow water equations are invariant by rotation. For the sake of clarity, u , v and c are considered constant hereafter. However, the reasoning remains valid for non-constant flow variables.

The domain of influence of the solution is located inside the first characteristic surface. It includes both the surface and the region of the phase space delimited by the surface. Indeed, a characteristic surface may be drawn from each point M located on a generating line (AB) of the surface issued from A (Figure 5.9). The segment

[MB] belongs to both characteristic surfaces. The intersection of the first surface with the plane $t = t_B$ is a circle (C) , the intersection of the second surface with the plane is a circle (C') . The point A influences the point M, thus influencing indirectly all the points on the circle (C') . The circle (C') spans the set of the internal points of the circle (C) as the point M spans all the possible locations along the segment [AB]. Consequently point A influences not only the points located on (C) but also all its internal points. The domain of influence of A includes all the points located inside the first characteristic surface.

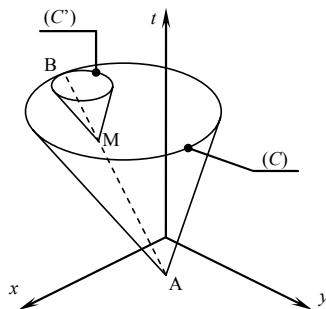


Figure 5.9. Definition sketch for the domain of influence in the phase space

The domain of dependence is determined by extending the reasoning to negative time intervals, that is, by travelling backward in time. Consider the characteristic surface passing at B (Figure 5.10). At any point M located on a generating curve, a characteristic surface can be drawn that passes at M. The intersection of the characteristic surface passing at M with the plane $t = t_A$ is a circle denoted by (C') . The intersection between the characteristic surface passing at B and the plane $t = t_A$ is denoted by (C) .

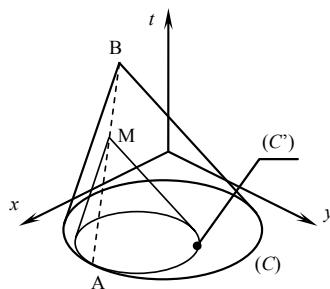


Figure 5.10. Definition sketch for the domain of dependence in the phase space

The circle (C') spans the complete set of interior points of the circle (C) as the point M spans all the possible locations along the line (AB). Consequently, the point B is influenced directly or indirectly by all the points located inside the characteristic surface passing at B.

5.4.3.2. Calculation of the solution

In spite of their apparent simplicity, the Riemann invariants derived in section 5.4.2 cannot be used to compute analytical solutions in a straightforward manner. Indeed, equation [5.84] includes derivatives in the direction normal to the secant plane. In the case of a genuinely two-dimensional flow field, such derivatives are non-zero and no analytical expression can be found for them, unless the flow configuration is very simple (e.g. radial symmetry). The characteristic form must then be approximated. Two possible approaches have been reported in the literature:

1) The first approach consists of selecting three bicharacteristic lines [A_1M], [A_2M] and [A_3M] passing at the point M of interest and writing the characteristic relationships [5.83–84] along them. The transverse derivatives are estimated from the known solution at points A_1 , A_2 and A_3 (see Figure 5.11). The three characteristic relationships allow the three flow variables to be determined uniquely at M. The approach was applied to gas dynamics and magnetohydrodynamics by Sauerwein [SAU 66, SAU 67] and to the two-dimensional shallow water equations by Daubert and Graffe [DAU 67], Katopodes [KAT 77], Katopodes and Strelkoff [KAT 79] and Gerritsen [GER 82]. The question remains however of the optimal choice of points A_1 , A_2 and A_3 . In fact, each particular choice for these points leads to a different solution.

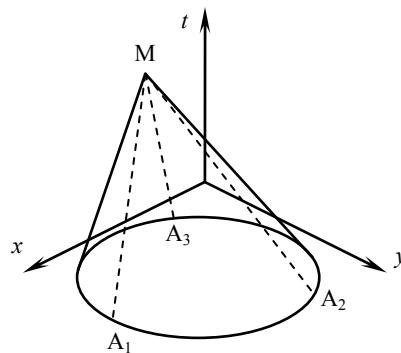


Figure 5.11. Calculating the solution at point M using three bicharacteristics

2) In another approach, proposed by the author [GUI 03b], the proper selection of the points A_1 , A_2 and A_3 is less crucial because the characteristic relationships are

integrated over the entire domain of dependence. Integrating equations [5.83–84] over the three surfaces $[A_1 A_2 M]$, $[A_2 A_3 M]$ and $[A_3 A_1 M]$ allows the three independent flow variables to be determined uniquely at M , while taking into account the details of the variations of the initial condition along the curve $(A_1 A_2 A_3)$.

5.5. Summary

5.5.1. What you should remember

An $m \times m$ multidimensional system of conservation laws is said to be hyperbolic if any linear combination of the Jacobian matrices of the fluxes with respect to the conserved variable has m real, distinct eigenvalues.

The notion of characteristic curve defined for one-dimensional problems may be extended to characteristic surfaces for two-dimensional problems and to characteristic volumes for three-dimensional problems.

The characteristic surfaces may be derived using the bicharacteristic approach presented in section 5.3.1.1 and the secant plane approach presented in section 5.3.1.2 indifferently.

In the case of a two-dimensional problem the characteristic surfaces are conical envelopes or curved lines in the phase space (x, y, t) .

5.5.2. Application exercises

5.5.2.1. Exercise 5.1: the Doppler effect

Consider a mobile sound source that moves at a speed u smaller than the speed of sound. The frequency N of the sound is constant. Using the secant plane approach, show that the frequency N' of the sound as heard by an immobile observer is given by:

$$N' = \frac{N}{(1 - M \cos \theta)} \quad [5.85]$$

where M is the Mach number and θ is the angle between the velocity vector of the source and the direction of the straight line drawn from the observer to the source (Figure 5.12).

This phenomenon is known as the Doppler effect.

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

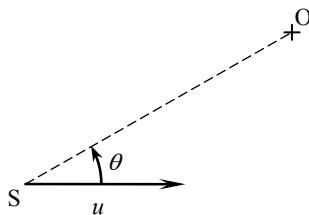


Figure 5.12. Mobile sound source

5.5.2.2. Exercise 5.2: visual assessment of the Mach number

Airplanes entering high moisture regions at supersonic speeds sometimes generate condensation patterns that develop next to the convex part of the wings and the hull. A condensation pattern indicates a sudden pressure drop, which is an indication that a shock wave is present (Figure 5.13).

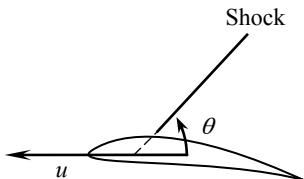


Figure 5.13. Condensation zone developing along the shock wave for a plane in supersonic flight

Show that the angle between the shock and the velocity vector of the airplane is given by:

$$M = \frac{1}{\sin \theta} \quad [5.86]$$

which allows the Mach number M to be determined visually.

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

Chapter 6

Finite Difference Methods for Hyperbolic Systems

6.1. Discretization of time and space

6.1.1. Discretization for one-dimensional problems

Hyperbolic systems of conservation laws can seldom be solved analytically when real-world problems are dealt with. This is because the initial and boundary conditions and the geometry of most real-world problems cannot be described analytically in a simple way. Most engineering applications therefore involve the solution of approximations of the governing equations. The solution of the approximate equations, that is easier to derive, can be expected to be “reasonably” close to that of the real equations provided that a number of criteria are satisfied (see Appendix B for basic notions in numerical analysis). The operation by which the original equations are approximated is called discretization. In the finite difference approach, presented in this chapter, time and space are treated identically in the discretization process. Other approaches exist, as shown in Chapters 7 and 8.

The discretization process consists of transforming the originally continuous time and space coordinates into discrete variables (see Figure 6.1). A set of points (called the computational points) is defined in time and space by the modeler. The solution is calculated at these points. The governing equations are approximated using the differences between the known and unknown values of the computational solution at the predefined points. Denoting by U_i^n the solution at the computational point x_i at the computational time t^n , the following difference may be seen as a

“good” approximation of the derivative $\partial U / \partial x$ over the interval $[x_i, x_{i+1}]$ at the time t^n .

$$D_x = \frac{U_{i+1}^n - U_i^n}{x_{i+1} - x_i} \quad [6.1]$$

Equation [6.1] is not the only possible approximation for the derivative $\partial U / \partial x$. Many alternative formulations may be proposed. The accuracy of the numerical solution depends on the accuracy with which the governing equations are approximated (see section B.1 in Appendix B). Approximation methods where the differences between the point values are used to estimate the derivatives are referred to as finite difference methods.

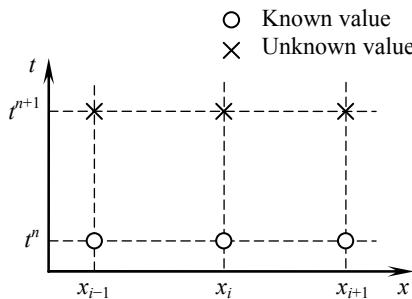


Figure 6.1. Discretization of time and space in the one-dimensional case

In what follows, the distance between the points i and $i + 1$ is denoted by $\Delta x_{i+1/2}$. It is often referred to as the grid spacing, or cell width. The difference between two successive computational times t^n and t^{n+1} , also called the computational time step, is usually denoted by Δt . The time t^n is usually referred to as “the time level n ”.

6.1.2. Multidimensional discretization

Several options are available for the discretization of multidimensional equations. Two main types of grid are distinguished (Figure 6.2):

1) Structured grids. The computational points of a structured grid are arranged along rows and columns. A computational point is located using two indices (i, j) in a two-dimensional space. The index i generally indicates the location of the point along the x -coordinate, while the index j indicates its location along the y -coordinate. A computational point is located using a triple index (i, j, k) in a three-

dimensional space. The value of the variable U at the point (i, j, k) at the time level n is usually denoted by $U_{i,j,k}^n$. The family of structured grid is divided into two subcategories:

1.1) Cartesian, or rectangular grids. The computational points form the intersection between two families of orthogonal, straight lines (Figure 6.2a).

1.2) Curvilinear grids. The computational points are located at the intersections between two families of curves that do not necessarily intersect at straight angles (Figure 6.2b). While curvilinear grids allow real-world geometries to be described more accurately, they require more work from the modeler in that it is the modeler's responsibility to define the curvature of the lines and the locations at which the computational points are to be distributed along the lines.

2) Unstructured grids. Such grids do not use the arrangement in lines and columns used by structured grids. The computational points are placed and linked to each other at the modeler's convenience, mostly based on the geometrical constraints imposed by the problem to be solved (Figure 6.2c).

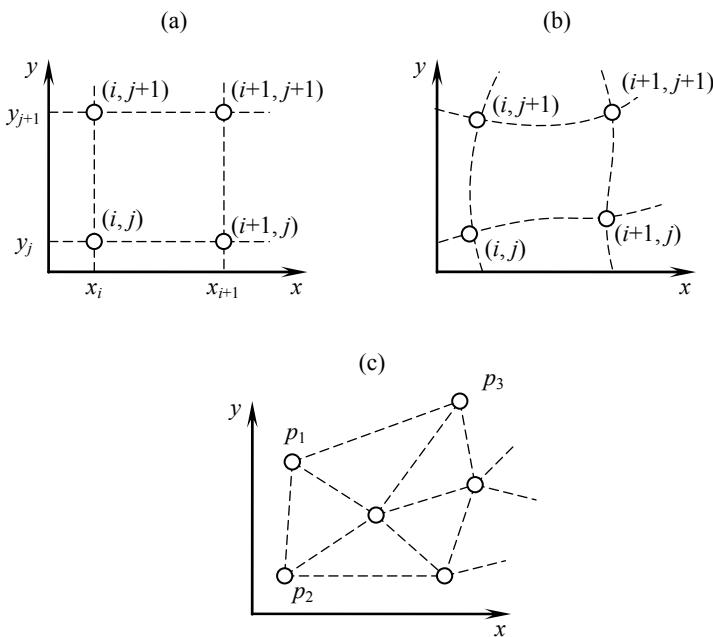


Figure 6.2. The various types of grid: structured, Cartesian (a), structured, curvilinear (b), unstructured (c). Sketch for a two-dimensional problem

6.1.3. *Explicit schemes, implicit schemes*

All scalar hyperbolic laws and hyperbolic systems of conservation laws contain derivatives with respect to time and space and a source term. The time derivative is that of the conserved variable, while the space derivative is that of the flux or the conserved variable, depending on whether the conservation form or the non-conservation form is being discretized.

When the space derivative and the source term are discretized using only the known values of U at the time level n , the discretization is said to be explicit because the unknown solution at the time level $n + 1$ can be computed directly from the known values at the time level n .

When the space derivative and/or the source term are discretized using the unknown values of U at the time level $n + 1$, the estimate of U^{n+1} is based on expressions that involve U^{n+1} itself. In such a case, the discretization is said to be implicit because U^{n+1} is defined as a function of itself.

Explicit schemes are easily programmed and maintained. However, they are subjected to a so-called stability constraint that yields a restriction in the range of permissible computational time steps (see section B.2 in Appendix B). The computational time step must remain smaller than a threshold value Δt_{\max} above which the numerical solution becomes unstable. Explicit methods usually lead to small computational time steps, thus increasing the number of calculations and leading to time-consuming simulations.

As shown in section B.2 of Appendix B, the so-called Courant number Cr (also called CFL number) is a key factor to the stability of explicit numerical methods. The Courant number expresses the ratio of the area covered by a wave of speed λ during the interval Δt to the total area of the grid cells. Most explicit methods are stable only when the absolute value of the Courant number is smaller than one.

Implicit schemes are not subjected to stability constraints. Most of them are indeed unconditionally stable, regardless of the Courant number. This makes them popular for industrial use because they allow larger computational time steps to be used, thus restricting the number of time steps and leading to faster applications. It should be remembered however that a fast method is not necessarily an accurate one.

6.2. The method of characteristics (MOC)

6.2.1. MOC for scalar hyperbolic laws

6.2.1.1. Principle of the method

The conservation form [1.1] of a scalar hyperbolic conservation law is recalled:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} = S$$

As shown in Chapter 1, equation [1.1] may be rewritten in the characteristic form [1.27], recalled here:

$$\frac{dU}{dt} = S' \quad \text{for } \frac{dx}{dt} = \lambda$$

where λ and S' are given as in equations [1.21], recalled here:

$$\left. \begin{aligned} \lambda &= \frac{\partial F}{\partial U} \\ S' &= S - \left(\frac{\partial F}{\partial x} \right)_{U=\text{Const}} \end{aligned} \right\}$$

Equation [1.27] is solved numerically by discretizing time and space as illustrated in Figure 6.3. The calculation of the unknown value U_i^{n+1} at the time level $n + 1$ is detailed hereafter.

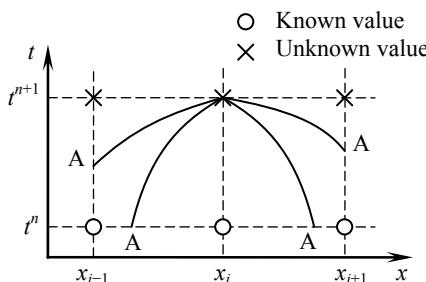


Figure 6.3. Definition sketch for the MOC

The principle of the MOC is the following. Assume first that the shape of the characteristic passing at (x_i, t^{n+1}) can be determined in the phase space. The intersection between the characteristic and the grid lines, also called the foot of the characteristic, is denoted by A. The various possible locations of the foot of the characteristic are illustrated in Figure 6.3. Integrating the characteristic form [1.27] between A and the point (x_i, t^{n+1}) leads to:

$$U_i^{n+1} = U_A + \int_A^{(x_i, t^{n+1})} S'(x, t) dt, \quad \frac{dx}{dt} = \lambda \quad [6.2]$$

The unknown value U_i^{n+1} can be calculated provided that the following two terms are estimated: (i) the value U_A of U at the foot of the characteristic and (ii) the integral of the source term between A and M. Classically, polynomial interpolation formulae are used. The most widely used are the first- and second-order and the Hermite polynomial-based interpolation introduced by Holly and Preissmann [HOL 77]. Interpolation issues are dealt with in the next three sections.

6.2.1.2. Interpolation at the foot of the characteristic: first-order formula

The value U_A of U at the foot of the characteristic is interpolated from the known values at the computational points. Consider first the case of the linear, or first-order interpolation, where U is estimated using the equation of a straight line. Two cases may be distinguished (Figure 6.4).

1) Point A is located between the points $(i-1, n)$ and (i, n) . The following first-order interpolation formula is used:

$$U_A = \frac{x_i - x_A}{\Delta x_{i-1/2}} U_{i-1}^n + \frac{x_A - x_{i-1/2}}{\Delta x_{i-1/2}} U_i^n \quad [6.3]$$

Defining the cell size and the average wave speed as:

$$\left. \begin{aligned} \Delta x_{i-1/2} &= x_i - x_{i-1} \\ \lambda_i^{n+1/2} &= \frac{x_i - x_A}{\Delta t} \end{aligned} \right\} \quad [6.4]$$

leads to:

$$U_A = \frac{\lambda_i^{n+1/2} \Delta t}{\Delta x_{i+1/2}} U_{i-1}^n + \frac{\Delta x_{i-1/2} - \lambda_i^{n+1/2} \Delta t}{\Delta x_{i-1/2}} U_i^n \quad [6.5]$$

where $\lambda_i^{n+1/2}$ represents the average wave speed of the characteristic that passes at the point $(i, n + 1)$ between the time levels n and $n + 1$. The expression [6.5] is simplified by introducing the Courant number, that represents the fraction of the cell covered by the characteristic over the time step Δt :

$$\text{Cr} = \frac{\lambda_i^{n+1/2} \Delta t}{\Delta x_{i-1/2}} \quad [6.6]$$

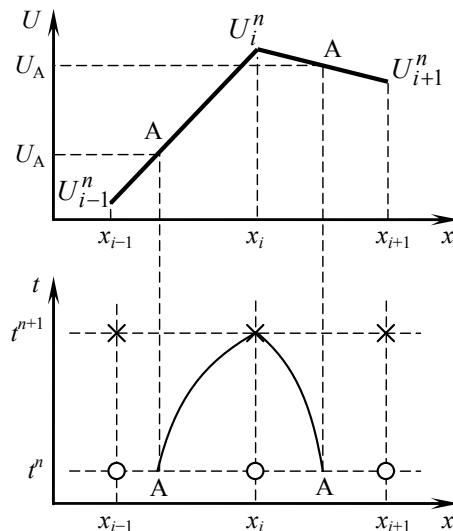


Figure 6.4. Interpolation at the foot of the characteristic. Sketch in the physical space (top) and in the phase space (bottom) for a first-order linear interpolation

Note that Cr is between 0 and 1 when A is located between the points $(i - 1, n)$ and (i, n) . Substituting definition [6.6] into equation [6.5] leads to the following expression:

$$U_A = \text{Cr} U_{i-1}^n + (1 - \text{Cr}) U_i^n, \quad 0 \leq \text{Cr} \leq 1 \quad [6.7]$$

2) The point A is located between the points (i, n) and $(i + 1, n)$. It is easy to check that the following formula applies:

$$U_A = -\text{Cr} U_{i+1}^n + (1 + \text{Cr}) U_i^n, \quad -1 \leq \text{Cr} \leq 0 \quad [6.8]$$

where the Courant number Cr is defined as:

$$\text{Cr} = \frac{\lambda_i^{n+1/2} \Delta t}{\Delta x_{i+1/2}} \quad [6.9]$$

3) Point A is located between $(i-1, n)$ and $(i-1, n+1)$. By definition, the Courant number is larger than unity. The following interpolation is used:

$$U_A = \frac{t^{n+1} - t_A}{\Delta t} U_{i-1}^n + \frac{t_A - t^n}{\Delta t} U_{i-1}^{n+1} \quad [6.10]$$

By definition of the average wave speed and the Courant number, we have:

$$\left. \begin{aligned} t^{n+1} - t_A &= \frac{\Delta x_{i-1/2}}{\lambda_i^{n+1/2}} = \frac{\Delta t}{\text{Cr}} \\ t_A - t^n &= \Delta t \frac{\Delta x_{i-1/2}}{\lambda_i^{n+1/2}} = \frac{\text{Cr} - 1}{\text{Cr}} \Delta t \end{aligned} \right\} \quad [6.11]$$

Substituting equations [6.11] into equation [6.10] leads to:

$$U_A = \frac{1}{\text{Cr}} U_{i-1}^n + \frac{\text{Cr} - 1}{\text{Cr}} U_{i-1}^{n+1}, \quad \text{Cr} \geq 1 \quad [6.12]$$

4) The point A is located between $(i+1, n)$ and $(i+1, n+1)$. By definition, the Courant number is smaller than -1 . The following interpolation formula is used:

$$U_A = -\frac{1}{\text{Cr}} U_{i+1}^n + \frac{\text{Cr} + 1}{\text{Cr}} U_{i+1}^{n+1}, \quad \text{Cr} \leq -1 \quad [6.13]$$

The formulae above can be summarized as follows:

$$U_A = \begin{cases} \frac{1}{\text{Cr}} U_{i-1}^n + \frac{\text{Cr} - 1}{\text{Cr}} U_{i-1}^{n+1} & \text{if } \text{Cr} \geq 1 \\ \text{Cr} U_{i-1}^n + (1 - \text{Cr}) U_i^n & \text{if } 0 \leq \text{Cr} \leq 1 \\ (1 + \text{Cr}) U_i^n - \text{Cr} U_{i+1}^n & \text{if } -1 \leq \text{Cr} \leq 0 \\ -\frac{1}{\text{Cr}} U_{i+1}^n + \frac{\text{Cr} + 1}{\text{Cr}} U_{i+1}^{n+1} & \text{if } \text{Cr} \leq -1 \end{cases} \quad [6.14]$$

where the Courant number is defined as:

$$\text{Cr} = \begin{cases} \frac{\lambda_i^{n+1/2} \Delta t}{\Delta x_{i-1/2}} & \text{if } \lambda_i^{n+1/2} \geq 0 \\ \frac{\lambda_i^{n+1/2} \Delta t}{\Delta x_{i+1/2}} & \text{if } \lambda_i^{n+1/2} \leq 0 \end{cases} \quad [6.15]$$

Note that the first and fourth formulae [6.14] are implicit because they involve the unknown values U_{i-1}^{n+1} and U_{i+1}^{n+1} . In contrast, the second and third formulae [6.14] are explicit because they allow U_i^{n+1} to be calculated directly from the known values at the time level n .

The calculation of the Courant number Cr requires that $\lambda_i^{n+1/2}$ be estimated. A number of possible formulae are proposed hereafter (the list is non-exhaustive):

$$\begin{aligned} \lambda_i^{n+1/2} &\approx \lambda_i^n && \text{(explicit without interpolation)} \\ \lambda_i^{n+1/2} &\approx \frac{\Delta x_{i+1/2} \lambda_{i-1}^n + \Delta x_{i-1/2} \lambda_{i+1}^n}{\Delta x_{i-1/2} + \Delta x_{i+1/2}} && \text{(explicit with interpolation)} \\ \lambda_i^{n+1/2} &\approx \lambda_i^{n+1} && \text{(implicit without interpolation)} \\ \lambda_i^{n+1/2} &\approx \frac{\Delta x_{i+1/2} \lambda_{i-1}^{n+1} + \Delta x_{i-1/2} \lambda_{i+1}^{n+1}}{\Delta x_{i-1/2} + \Delta x_{i+1/2}} && \text{(implicit with interpolation)} \end{aligned} \quad [6.16]$$

The last two formulae are said to be implicit because they involve the unknown values of λ at the time level $n + 1$. Since λ is a function of U in the general case, the determination of $\lambda_i^{n+1/2}$ in an implicit method is iterative.

Note that when the Courant number is an integer, the foot of the characteristic is located at a grid point. The interpolation formulae are exact and the numerical solution is the analytical solution. Also note that formulae [6.14] are defined over different intervals that coincide only at $\text{Cr} = -1$, $\text{Cr} = 0$ and $\text{Cr} = +1$. Both the first and second formula [6.14] give the analytical solution U_{i-1}^n when Cr tends to 1 by a lower or upper value; both the second and third formulae [6.14] yield the analytical solution U_i^n when the Courant number tends to zero; eventually, both the third and fourth formulae [6.14] give the solution U_{i+1}^n when the Courant number tends to -1 by a lower or upper value. This continuous behavior of the interpolation formula is referred to as a “continuous switch” in the literature.

The linear interpolation procedures introduce a numerical effect known as “numerical diffusion” (see section B.1.3). Numerical diffusion may be eliminated using higher-order interpolations that use more computational points. A second-order interpolation technique is presented in the next section.

6.2.1.3. Interpolation at the foot of the characteristic: second-order formula

This section deals with a second-order interpolation method for U_A (Figure 6.5):

$$U_A = U_i^n + (x_A - x_i)^2 a_i + (x_A - x_i) b_i \quad [6.17]$$

Equation [6.17] is valid when point A is located between the points $(i-1, n)$ and $(i+1, n)$. It is applicable for Courant numbers ranging from -1 to $+1$. The coefficients a_i and b_i are determined by stating that function [6.17] should coincide with the value U_{i-1}^n for $x_A = x_{i-1}$ and with U_{i+1}^n for $x_A = x_{i+1}$:

$$\left. \begin{aligned} U_{i-1}^n &= U_i^n + \Delta x_{i-1/2}^2 a_i - \Delta x_{i-1/2} b_i \\ U_{i+1}^n &= U_i^n + \Delta x_{i+1/2}^2 a_i + \Delta x_{i+1/2} b_i \end{aligned} \right\} \quad [6.18]$$

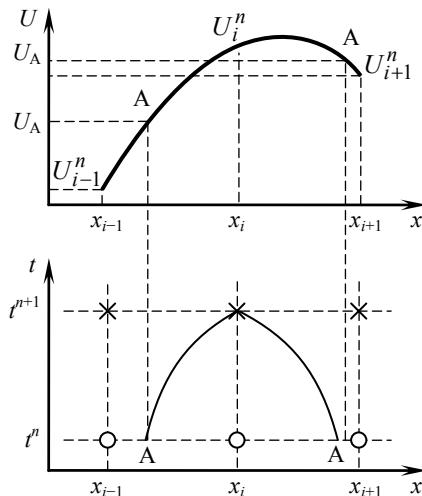


Figure 6.5. Second-order interpolation. Sketch in the physical space (top) and in the phase space (bottom)

Solving equations [6.18] for a_i and b_i leads to:

$$\left. \begin{aligned} a_i &= \frac{\Delta x_{i+1/2} U_{i-1}^n - (\Delta x_{i-1/2} + \Delta x_{i+1/2}) U_i^n + \Delta x_{i-1/2} U_{i+1}^n}{(\Delta x_{i-1/2} + \Delta x_{i+1/2}) \Delta x_{i-1/2} \Delta x_{i+1/2}} \\ b_i &= \frac{-\Delta x_{i+1/2}^2 U_{i-1}^n + (\Delta x_{i+1/2}^2 - \Delta x_{i-1/2}^2) U_i^n + \Delta x_{i+1/2}^2 U_{i+1}^n}{(\Delta x_{i-1/2} + \Delta x_{i+1/2}) \Delta x_{i-1/2} \Delta x_{i+1/2}} \end{aligned} \right\} \quad [6.19]$$

When the grid is regular, $\Delta x_{i-1/2} = \Delta x_{i+1/2} = \Delta x$ and equations [6.19] simplify into:

$$\left. \begin{aligned} a_i &= \frac{U_{i-1}^n - 2U_i^n + U_{i+1}^n}{2\Delta x^2} \\ b_i &= \frac{U_{i+1}^n - U_{i-1}^n}{2\Delta x} \end{aligned} \right\} \quad [6.20]$$

Substituting equation [6.20] into equation [6.17], using the relationship $x_A = x_i - Cr \Delta x$ leads to:

$$U_A = (Cr + 1) \frac{Cr}{2} U_{i-1}^n + (1 - Cr^2) U_i^n + (Cr - 1) \frac{Cr}{2} U_{i+1}^n \quad [6.21]$$

Note that for $Cr = -1$, $Cr = 0$ or $Cr = +1$, the foot of the characteristic coincides with a grid point and the interpolation is exact. The analytical solution is obtained.

6.2.1.4. Estimation of the source term

The integral of the source term must be estimated in equation [6.2]. Four options are proposed hereafter:

$$(x_i, t^{n+1}) \int_A S'(x, t) dt \approx \begin{cases} S'(U_i^n) \Delta t \\ S'(U_A) \Delta t \\ [S'(U_A) + S'(U_i^{n+1/2})] \Delta t / 2 \\ S'(U_i^{n+1}) \Delta t \end{cases} \quad [6.22]$$

The first two options are explicit because they use only known values of U , whether they are point values at the computational points or interpolated values. The

remaining two expressions are implicit because they require the knowledge of the unknown value U_i^{n+1} . The solution is determined iteratively in the general case.

6.2.1.5. Treatment of boundary conditions

Assume first that the wave speed $\lambda_1^{n+1/2}$ at the first computational point (i.e. at the left-hand boundary of the domain) is positive. Then the characteristic $dx/dt = \lambda$ enters the computational domain and the value of U at $i = 1$ cannot be computed from the internal points. It must be supplied in the form of a boundary condition (see Figure 6.6). Applying equations [6.14] with $i = 1$ in the case of a positive Courant number gives:

$$U_2^{n+1} = \begin{cases} \frac{1}{Cr} U_1^n + \frac{Cr-1}{Cr} U_1^{n+1} & \text{if } Cr \geq 1 \\ Cr U_1^n + (1-Cr) U_2^n & \text{if } 0 \leq Cr \leq 1 \end{cases} \quad [6.23]$$

Conversely, if the wave speed $\lambda_M^{n+1/2}$ at the right-hand boundary of the computational domain is negative, the value of U at $i = M$ must be supplied in the form of a boundary condition. Equations [6.14] lead to:

$$U_{M-1}^{n+1} = \begin{cases} (1+Cr) U_{M-1}^n - Cr U_M^n & \text{if } -1 \leq Cr \leq 0 \\ -\frac{1}{Cr} U_M^n + \frac{Cr+1}{Cr} U_M^{n+1} & \text{if } Cr \leq -1 \end{cases} \quad [6.24]$$

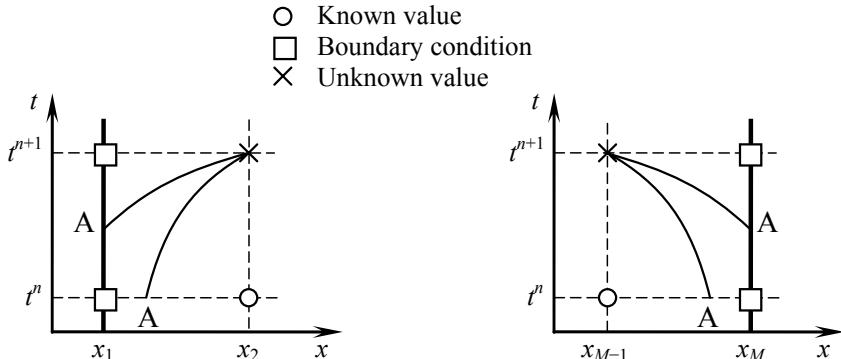


Figure 6.6. Treatment of boundary conditions. The boundary is indicated by a bold line

6.2.2. The MOC for hyperbolic systems of conservation laws

6.2.2.1. Principle of the method

The generalization of the MOC to hyperbolic systems of conservation laws is best known as the Courant-Isaacson-Rees (CIR) scheme [COU 52]. The CIR scheme can be viewed as a direct application of the scalar MOC to the Riemann invariants (Figure 6.7). The purpose is to solve the conservation form [2.2], recalled hereafter:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} = \mathbf{S}$$

As shown in Chapter 2, equation [2.2] can be written in characteristic form as in equation [2.25], recalled hereafter:

$$\frac{dW_p}{dt} = S''_p \quad \text{for } \frac{dx}{dt} = \lambda^{(p)} \quad \forall p = 1, 2, \dots, m$$

where $S'' = K^{-1}S'$ and the wave speeds $\lambda^{(p)}$ are the eigenvalues of the Jacobian matrix A of F with respect to U and K is the matrix formed by the eigenvectors of A (see section 2.1.3 for the details of the developments).

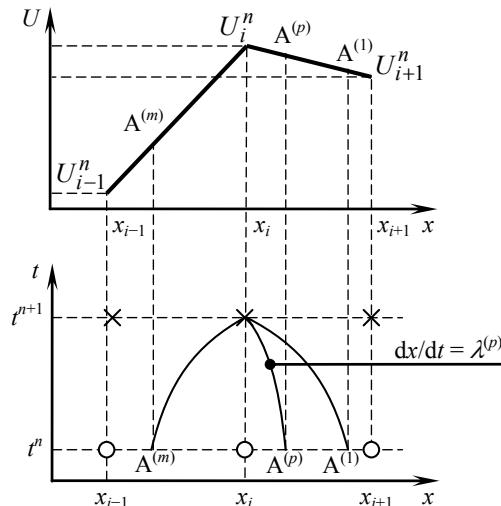


Figure 6.7. Interpolation at the feet of the characteristics. Definition sketch in the physical space (top) and in the phase space (bottom) for a linear interpolation

The foot of the p th characteristic passing at the point $(i, n+1)$ is denoted by $A^{(p)}$ (Figure 6.7). Integrating relationship [2.25] along the characteristic line $A^{(p)}M$ leads to:

$$(W_p)_i^{n+1} = (W_p)_{A^{(p)}} + \int_{A^{(p)}}^{(x_i, t^{n+1})} S_p''(x, t) dt, \frac{dx}{dt} = \lambda^{(p)}, p = 1, 2, \dots, m \quad [6.25]$$

Note that this formulation is identical to equation [6.2] for the scalar problem. Two options are available for the calculation of the p th Riemann invariant.

1) Interpolate W_p at the foot of the p th characteristic from the known values at the computational points:

$$(W_p)_{A^{(p)}} = \begin{cases} \frac{1}{Cr^{(p)}} (W_p)_{i-1}^n + \frac{Cr^{(p)} - 1}{Cr^{(p)}} (W_p)_{i-1}^{n+1} & \text{if } Cr^{(p)} \geq 1 \\ Cr^{(p)} (W_p)_{i-1}^n + (1 - Cr^{(p)}) (W_p)_i^{n+1} & \text{if } 0 \leq Cr^{(p)} \leq 1 \\ (1 + Cr^{(p)}) (W_p)_i^{n+1} - Cr^{(p)} (W_p)_{i+1}^n & \text{if } -1 \leq Cr^{(p)} \leq 0 \\ -\frac{1}{Cr^{(p)}} (W_p)_{i+1}^n + \frac{Cr^{(p)} + 1}{Cr^{(p)}} (W_p)_{i+1}^{n+1} & \text{if } Cr^{(p)} \leq -1 \end{cases} \quad [6.26]$$

where $(W_p)_i^n$ is the value of the p th Riemann invariant calculated from the value of U_i^n and $Cr^{(p)}$ is the Courant number for the p th wave:

$$Cr^{(p)} = \begin{cases} \frac{\lambda^{(p)}_i^{n+1/2} \Delta t}{\Delta x_{i-1/2}} & \text{if } \lambda^{(p)}_i^{n+1/2} \geq 0 \\ \frac{\lambda^{(p)}_i^{n+1/2} \Delta t}{\Delta x_{i+1/2}} & \text{if } \lambda^{(p)}_i^{n+1/2} \leq 0 \end{cases} \quad [6.27]$$

2) Calculate W_p from the interpolated value of U at the foot of the characteristic:

$$(W_p)_{A^{(p)}} = W_p(U_{A^{(p)}}) \quad [6.28]$$

where $U_{A^{(p)}}$ is the estimate of U at the foot $A^{(p)}$ of the p th characteristic:

$$U_{A^{(p)}} = \begin{cases} \frac{1}{Cr^{(p)}} U_{i-1}^n + \frac{Cr^{(p)} - 1}{Cr^{(p)}} U_{i-1}^{n+1} & \text{if } Cr^{(p)} \geq 1 \\ Cr^{(p)} U_{i-1}^n + (1 - Cr^{(p)}) U_i^{n+1} & \text{if } 0 \leq Cr^{(p)} \leq 1 \\ (1 + Cr^{(p)}) U_i^{n+1} - Cr^{(p)} U_{i+1}^n & \text{if } -1 \leq Cr^{(p)} \leq 0 \\ -\frac{1}{Cr^{(p)}} U_{i+1}^n + \frac{Cr^{(p)} + 1}{Cr^{(p)}} U_{i+1}^{n+1} & \text{if } Cr^{(p)} \leq -1 \end{cases} \quad [6.29]$$

When the governing equations are linear (as is the case with the water hammer equations), the Riemann invariants are linear combinations of the components of U and both options give the same result.

The wave speeds and the source term may be estimated using any of the formulae [6.16] and [6.22]. The number of boundary conditions to be provided at each boundary of the domain is equal to the number of characteristics that enter the domain.

The practical implementation of the CIR scheme for Courant numbers smaller than one is straightforward. Indeed, the case $|Cr| \leq 1$ uses only the second and third formulae [6.26] or [6.29], where only known values of U are required for the calculation of the solution. In contrast, when the absolute value of the Courant number is larger than one, an implicit formulation must be used, leading to a dependence between the unknown value of U at two neighboring points. The nonlinear dependence between the wave speed and the conserved variable U makes the procedure time-consuming in the general case.

6.2.2.2. Application example: the water hammer equations

This section deals with the application of the CIR scheme to the water hammer equations. Expression [2.79] is recalled:

$$\left. \begin{aligned} \frac{dp}{dt} - \frac{\rho c}{A} \frac{dQ}{dt} &= (k|u|u + \rho g A \sin \theta) \frac{c}{A} && \text{for } \frac{dx}{dt} = -c \\ \frac{dp}{dt} + \frac{\rho c}{A} \frac{dQ}{dt} &= (-k|u|u - \rho g A \sin \theta) \frac{c}{A} && \text{for } \frac{dx}{dt} = c \end{aligned} \right\}$$

For the sake of clarity, the source term is assumed to be zero hereafter. Equations [2.79] simplify into equations [2.83] where S'' is set to zero:

$$\left. \begin{array}{l} \frac{d}{dt} \left(p - \frac{\rho c}{A} Q \right) = 0 \quad \text{for } \frac{dx}{dt} = -c \\ \frac{d}{dt} \left(p - \frac{\rho c}{A} Q \right) = 0 \quad \text{for } \frac{dx}{dt} = c \end{array} \right\} \quad [6.30]$$

Integrating equations [6.30] along the characteristics $dx/dt = -c$ and $dx/dt = +c$ yields the following algebraic system:

$$\left. \begin{array}{l} p_i^{n+1} - \frac{\rho c}{A} Q_i^{n+1} = p_{A^{(1)}} - \frac{\rho c}{A} Q_{A^{(1)}} \\ p_i^{n+1} + \frac{\rho c}{A} Q_i^{n+1} = p_{A^{(2)}} + \frac{\rho c}{A} Q_{A^{(2)}} \end{array} \right\} \quad [6.31]$$

where $A^{(1)}$ and $A^{(2)}$ are the feet of the first and second characteristic passing at $(i, n+1)$ respectively (Figure 6.8).

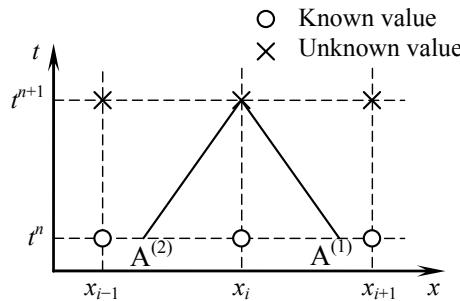


Figure 6.8. Application of the CIR scheme to the water hammer equations

Solving equations [6.31] for p and Q leads to:

$$\left. \begin{array}{l} p_i^{n+1} = \frac{p_{A^{(1)}} + p_{A^{(2)}}}{2} + \frac{\rho c}{2A} (Q_{A^{(2)}} - Q_{A^{(1)}}) \\ Q_i^{n+1} = \frac{Q_{A^{(1)}} + Q_{A^{(2)}}}{2} p_{A^{(2)}} + \frac{A}{2\rho c} (p_{A^{(1)}} - p_{A^{(2)}}) \end{array} \right\} \quad [6.32]$$

The value of p and Q at the feet of the characteristics may be interpolated indifferently from U and W. It is easy to check that the linear dependence between U and W leads to identical formulations for [6.26] and [6.29]. In the case of a regular cell size Δx , the following formula is obtained:

$$\left. \begin{aligned} p_{A^{(1)}} &= (1 + Cr^{(1)})p_i^n - Cr^{(1)}p_{i+1}^n \\ Q_{A^{(1)}} &= (1 + Cr^{(1)})Q_i^n - Cr^{(1)}Q_{i+1}^n \\ p_{A^{(2)}} &= Cr^{(2)}p_{i-1}^n + (1 - Cr^{(2)})p_i^n \\ Q_{A^{(2)}} &= Cr^{(2)}Q_{i-1}^n + (1 - Cr^{(2)})Q_i^n \end{aligned} \right\} [6.33]$$

Substituting [6.33] into [6.32], noting that $Cr^{(2)} = -Cr^{(1)} = Cr$, yields

$$\left. \begin{aligned} p_i^{n+1} &= \frac{Cr p_{i-1}^n + 2(1 - Cr)p_i^n + Cr p_{i+1}^n}{2} + Cr \frac{\rho c}{2A} (Q_{i-1}^n - Q_{i+1}^n) \\ Q_i^{n+1} &= \frac{Cr Q_{i-1}^n + 2(1 - Cr)Q_i^n + Cr Q_{i+1}^n}{2} + Cr \frac{A}{2\rho c} (p_{i-1}^n - p_{i+1}^n) \end{aligned} \right\} [6.34]$$

where Cr is given by

$$Cr = -Cr^{(1)} = Cr^{(2)} = \frac{c\Delta t}{\Delta x} [6.35]$$

Many software packages for water hammer simulation use the fact that the speed of sound c is constant in pipes with homogenous material and geometrical properties. The time step and/or cell size are adjusted in such a way that the Courant number is equal to one over the computational domain. The need for an interpolation procedure is then eliminated and the solution is exact. Equations [6.34] simplify into:

$$\left. \begin{aligned} p_i^{n+1} &= \frac{p_{i-1}^n + p_{i+1}^n}{2} + \frac{\rho c}{2A} (Q_{i-1}^n - Q_{i+1}^n) \\ Q_i^{n+1} &= \frac{Q_{i-1}^n + Q_{i+1}^n}{2} + \frac{A}{2\rho c} (p_{i-1}^n - p_{i+1}^n) \end{aligned} \right\} [6.36]$$

The reduced number of computations makes the particular application [6.36] approximately three times as fast as the general form [6.34].

6.2.3. Application examples

6.2.3.1. The linear advection equation

A major drawback of the MOC is that the equations are not solved in conservation form. Due to this, mass and/or momentum conservation may not always be guaranteed. The non-conservation form [1.48] of the linear advection equation, recalled hereafter, is solved:

$$\frac{\partial C}{\partial t} + u \frac{\partial C}{\partial x} = 0$$

The flow velocity u is assumed to be uniform over the computational domain. The parameters of the problem are given in Table 6.1. Such parameters are typical for the advection of a contaminant in a river. The initial condition is a top hat function, the width of which is 5 km. The calculation is carried out over an irregular computational grid. The cell size is 1 km over the entire domain, except between $x = 10$ km and $x = 14$ km, where $\Delta x = 4$ km. Since u is uniform, the Courant number is not. It is equal to one all over the domain, except between $x = 10$ km and $x = 14$ km, where it is equal to 1/4.

Symbol	Meaning	Value
A	River cross-sectional area	$1,000 \text{ m}^2$
C_i^0	Initial concentration	1 g/l for $1 \text{ km} \leq x \leq 5 \text{ km}$, 0 g/l otherwise
C_G	Concentration at the left-hand boundary	0 g/l
L	Length of the domain	30 km
u	Flow velocity	1 km/hr
Δt	Computational time step	1 h
Δx	Cell size	4 km for $10 \text{ km} \leq x \leq 14 \text{ km}$, 1 km otherwise

Table 6.1. Physical and numerical parameters for the solution of the advection equation over an irregular grid

The numerical solution computed by the first-order MOC is compared to the analytical solution in Figure 6.9. The concentration signal reaches the zone $\Delta x = 4$ km at $t = 4$ hr. Up to this time, the numerical solution is identical to the analytical solution because the concentration signal is transported with a Courant number equal to one. At $t = 5$ hr, the signal enters the zone $\Delta x = 4$ km. Owing to the interpolation between $x_{11} = 10$ km and $x_{12} = 14$ km, the solution is smoothed out and spreads artificially over the cell. Owing to the interpolation between the zero value

at x_{11} and the non-zero value at x_{12} , the concentration at $x = x_{12}$ does not return to zero, even after an infinite time.

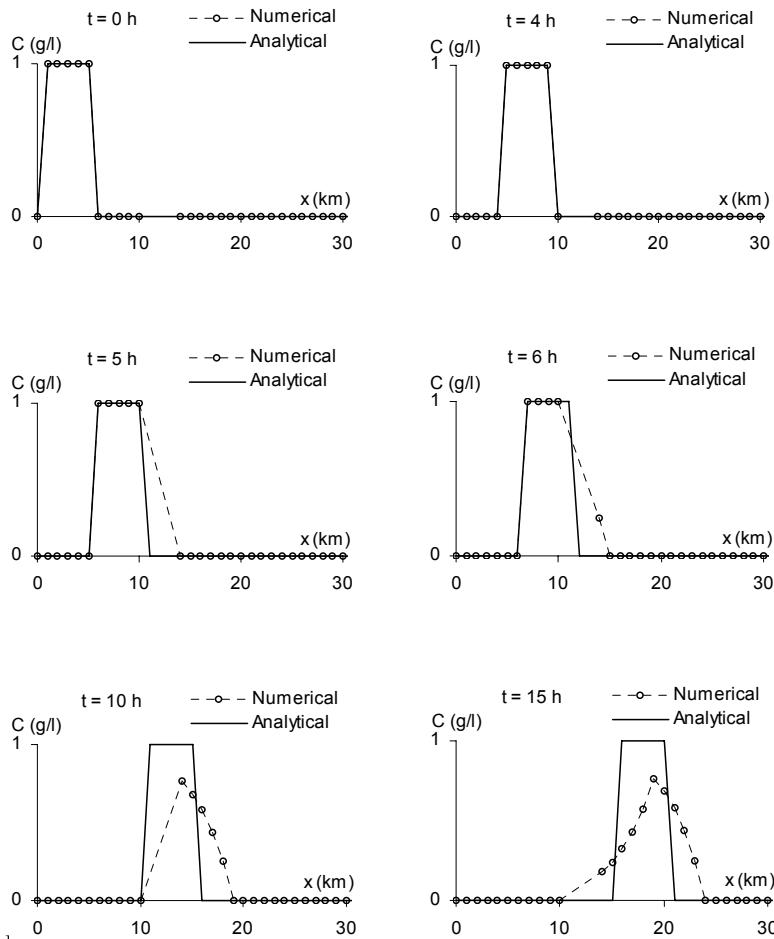


Figure 6.9. Advection of a square concentration signal. Comparison between the analytical solution and the numerical solution obtained using the first-order MOC on an irregular grid

The total mass of contaminant in the computational domain is plotted as a function of time in Figure 6.10. The mass is calculated as the integral of the concentration profile:

$$M_T(t^n) = \left[\sum_{i=1}^{M-1} (C_i^n + C_{i+1}^n) \Delta x_{i+1/2} + C_M^n \Delta x_{M-1/2} \right] \frac{A}{2} \quad [6.37]$$

The total mass of contaminant is not constant, which indicates that the conservation properties of the numerical solution are violated. The artificial spreading of the contaminant front at $t = 5$ hr results in a simultaneous increase in the total mass of contaminant. When the contaminant leaves the zone $\Delta x = 4$ km, the total mass suddenly decreases below its initial value. The initial value is recovered asymptotically as the contaminant travels downstream. However, it is never totally recovered, even for infinite times.

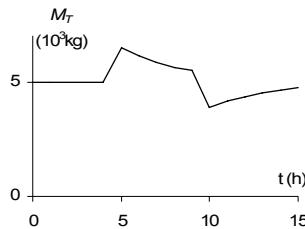


Figure 6.10. Advection of a square concentration signal. Total mass of contaminant in the domain as computed by the first-order MOC on an irregular grid

6.2.3.2. The inviscid Burgers equation

The characteristic form [1.68] of the inviscid Burgers equation is solved. A Riemann problem, the parameters of which are given in Table 6.2, is solved using the first-order MOC.

Symbol	Meaning	Value
L	Length of the domain	20 m
u_i^0	Initial speed	1 m/s for $x \leq 5$ m, 0 m/s otherwise
Δt	Computational time step	1 s
Δx	Cell size	1 m

Table 6.2. Physical and numerical parameters for the solution of the inviscid Burgers equation using the first-order MOC

The following options are used for the estimate of λ :

$$\lambda_i^{n+1/2} = \begin{cases} u_i^n & \text{(option 1)} \\ (u_{i-1}^n + u_i^n) / 2 & \text{(option 2)} \\ (u_{i-1}^n + u_{i+1}^n) / 2 & \text{(option 3)} \end{cases} \quad [6.38]$$

The numerical solution at $t = 20$ s is compared to the analytical solution for each of the three options in Figure 6.11. Note that the analytical solution is a shock that propagates at a speed given by the average value of the speeds on both sides of the discontinuity.

Option 1 gives a zero value for λ . The shock does not move.

Option 2 is based on the analytical formula for the shock speed. The numerical solution moves at the right speed, but the front is subjected to numerical diffusion. This is because the front moves at a speed $c_s = 0.5$ m/s, which corresponds to $Cr = 1/2$. As shown in section B.2.5, the phase portrait of the first-order MOC shows that numerical diffusion is maximum for $Cr = 1/2$, hence the smoothing in the neighborhood of the front.

Option 3 leads to an underestimated shock speed. The numerical profile is slower than the analytical profile.

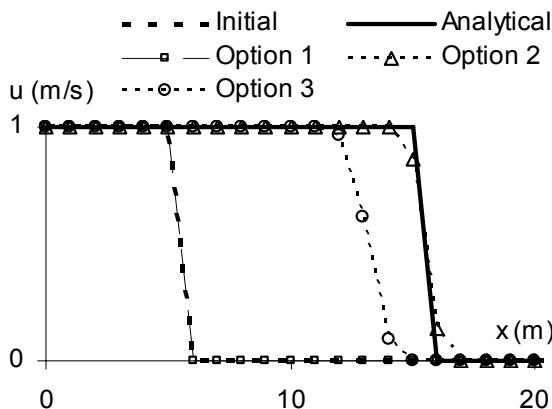


Figure 6.11. Analytical solution and numerical profiles computed by the first-order MOC for the three calculation options [6.38]

As shown by the two examples above, the MOC may lead to conservation problems when the Courant number is not uniform over the computational domain. This is because the MOC does not solve the governing equations in conservation form. This is one of the reasons why the method is seldom used in modern computational software packages (with the exception of the water hammer equations that are based on a constant wave speed).

6.3. Upwind schemes for scalar laws

6.3.1. The explicit upwind scheme (non-conservative version)

Upwind schemes aim to solve the conservation form [1.1] recalled here:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} = S$$

The derivative of U with respect to time is estimated as:

$$\frac{\partial U}{\partial t} \approx \frac{U_i^{n+1} - U_i^n}{\Delta t} \quad [6.39]$$

The scheme is said to be “upwind” because the derivative of F with respect to space is estimated using the computational point located upstream of the point i (Figure 6.12).

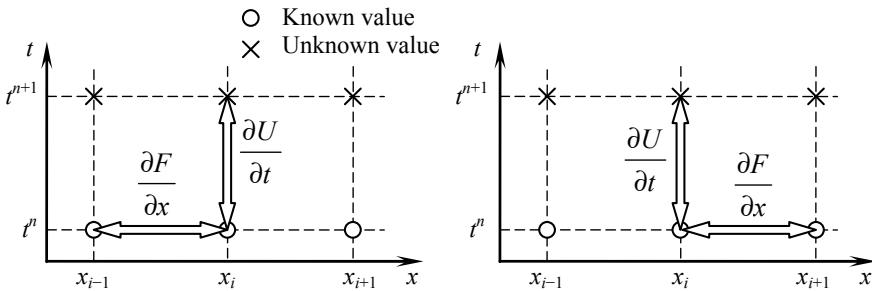


Figure 6.12. Definition sketch for the explicit upwind scheme. Sketch for a positive wave speed (left), for a negative wave speed (right)

In the explicit approach, the flux F and the source term S are estimated using the known values of U :

$$\begin{aligned} \frac{\partial F}{\partial x} &\approx \begin{cases} \frac{F_i^n - F_{i-1}^n}{\Delta x_{i-1/2}} & \text{if } \lambda_i^{n+1/2} \geq 0 \\ \frac{F_{i+1}^n - F_i^n}{\Delta x_{i+1/2}} & \text{if } \lambda_i^{n+1/2} \leq 0 \end{cases} \\ S &\approx S_i^n \end{aligned} \quad [6.40]$$

where $F_i^n = F(U_i^n)$ and $S_i^n = S(U_i^n)$. The numerical solution is stable if the absolute value of the Courant number is smaller than or equal to one. The Courant number is defined as in equation [6.15], recalled hereafter:

$$\text{Cr} = \begin{cases} \frac{\lambda_i^{n+1/2} \Delta t}{\Delta x_{i-1/2}} & \text{if } \lambda_i^{n+1/2} \geq 0 \\ \frac{\lambda_i^{n+1/2} \Delta t}{\Delta x_{i+1/2}} & \text{if } \lambda_i^{n+1/2} \leq 0 \end{cases}$$

Substituting approximations [6.39–40] into equation [1.1] leads to:

$$U_i^{n+1} = \begin{cases} U_i^n + \frac{\Delta t}{\Delta x_{i-1/2}} (F_{i-1}^n - F_i^n) + \Delta t S_i^n & \text{if } \lambda_i^{n+1/2} \geq 0 \\ U_i^n + \frac{\Delta t}{\Delta x_{i+1/2}} (F_i^n - F_{i+1}^n) + \Delta t S_i^n & \text{if } \lambda_i^{n+1/2} \leq 0 \end{cases} \quad [6.41]$$

In the particular case of the linear advection equation, $F = uC$. It is easy to check that equations [6.41] simplify into the same expression as the second and third equations [6.14]:

$$U_i^{n+1} = \begin{cases} \text{Cr} U_{i-1}^n + (1 - \text{Cr}) U_{i-1}^n & \text{if } \lambda_i^{n+1/2} \geq 0 \\ -\text{Cr} U_{i+1}^n + (1 + \text{Cr}) U_i^n & \text{if } \lambda_i^{n+1/2} \leq 0 \end{cases} \quad [6.42]$$

In other words, the explicit upwind scheme is equivalent to the MOC when the absolute value of the Courant number is smaller than one. It is stable only when the Courant number lies within the range $[-1, +1]$.

6.3.2. The implicit upwind scheme (non-conservative version)

In this scheme the time derivative is estimated as in equation [6.39], recalled here:

$$\frac{\partial U}{\partial t} \approx \frac{U_i^{n+1} - U_i^n}{\Delta t}$$

while the space derivative of the flux is estimated using the unknown values at the time level $n + 1$:

$$\frac{\partial F}{\partial x} \approx \begin{cases} \frac{F_i^{n+1} - F_{i-1}^{n+1}}{\Delta x_{i-1/2}} & \text{if } \lambda_i^{n+1/2} \geq 0 \\ \frac{F_{i+1}^{n+1} - F_i^{n+1}}{\Delta x_{i+1/2}} & \text{if } \lambda_i^{n+1/2} \leq 0 \end{cases} \quad [6.43]$$

$$S \approx S_i^{n+1}$$

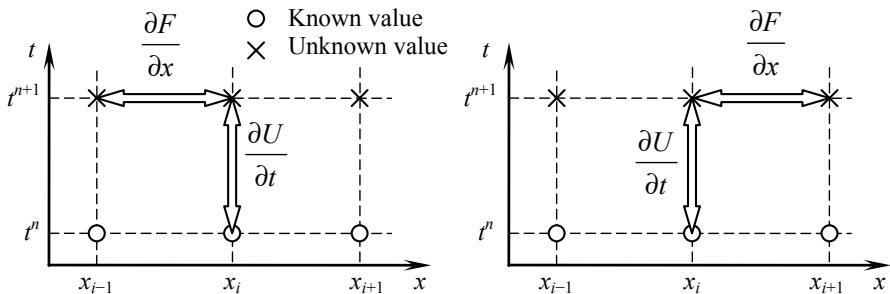


Figure 6.13. Definition sketch for the implicit upwind scheme. Sketch for a positive wave speed (left), for a negative wave speed (right)

Substituting approximations [6.39] and [6.43] into equation [1.1] leads to:

$$U_i^{n+1} = \begin{cases} U_i^n + \frac{\Delta t}{\Delta x_{i-1/2}} (F_{i-1}^{n+1} - F_i^{n+1}) + S_i^n & \text{if } \lambda_i^{n+1/2} \geq 0 \\ U_i^n + \frac{\Delta t}{\Delta x_{i+1/2}} (F_i^{n+1} - F_{i+1}^{n+1}) + S_i^n & \text{if } \lambda_i^{n+1/2} \leq 0 \end{cases} \quad [6.44]$$

Applying the implicit scheme to the particular case of the advection equation leads to the following formula that is not equivalent to the MOC:

$$U_i^{n+1} = \begin{cases} \frac{Cr}{1+Cr} U_{i-1}^{n+1} + \frac{1}{1+Cr} U_i^n & \text{if } \lambda_i^{n+1/2} \geq 0 \\ \frac{-Cr}{1-Cr} U_{i+1}^{n+1} + \frac{1}{1-Cr} U_i^n & \text{if } \lambda_i^{n+1/2} \leq 0 \end{cases} \quad [6.45]$$

A consistency analysis (see section B.1) shows that the implicit upwind scheme is more diffusive than the first-order MOC for Courant numbers larger than one. The numerical solution is stable for all values of the Courant number.

6.3.3. Conservative versions of the implicit upwind scheme

The explicit upwind scheme presented in section 6.3.1 leads to the same formulation as the first-order MOC when applied to the advection equation. Consequently, conservation is not guaranteed when the scheme is applied with irregular grids. Conservation can be restored via a minor modification in the estimate of the time derivative. This is done by attaching a control volume to each computational point. The control volume is delineated by interfaces located at mid-distance between two adjacent computational points (Figure 6.14).

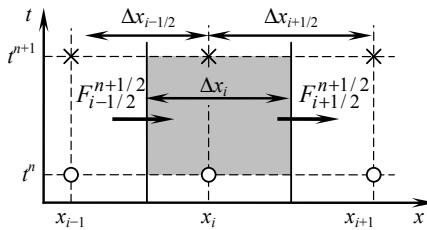


Figure 6.14. Defining control volumes for the upwind scheme

A mass balance over the control volume attached to the point i gives:

$$\Delta x_i U_i^{n+1} = \Delta x_i U_i^n + (F_{i-1/2}^{n+1/2} - F_{i+1/2}^{n+1/2}) + \Delta t S_i^{n+1/2} \quad [6.46]$$

where $F_{i-1/2}^{n+1/2}$ and $F_{i+1/2}^{n+1/2}$ are the average value of the fluxes across the left- and right-hand interface of the control volume respectively and $S_i^{n+1/2}$ is the average value of the source term over the control volume. The size of the control volume is given by $\Delta x_i = (\Delta x_{i-1/2} + \Delta x_{i+1/2})/2$. Note that in equation [6.46] the quantity U_i^n is not a point value but the average value of U over the control volume at the time level n . This approach is to be put in parallel with the finite volume approach described in Chapter 7.

In the conservative upwind approach, the flux is estimated at a given interface using the variable in the cell located immediately upstream of the interface. For instance, the flux at the interface $i - 1/2$ is computed using the cell $i - 1$ if the wave

speed is positive. It is computed using the cell i if the wave speed is negative. In the same way, the flux at the interface $i + 1/2$ is computed using the cell i or $i + 1$ for a positive and negative wave speed respectively. These formulations can be summarized as follows:

$$\begin{aligned} F_{i-1/2}^{n+1/2} &\approx \begin{cases} F_{i-1}^n & \text{if } \lambda_i^{n+1/2} \geq 0 \\ F_i^n & \text{if } \lambda_i^{n+1/2} \leq 0 \end{cases} \\ F_{i+1/2}^{n+1/2} &\approx \begin{cases} F_i^n & \text{if } \lambda_i^{n+1/2} \geq 0 \\ F_{i+1}^n & \text{if } \lambda_i^{n+1/2} \leq 0 \end{cases} \end{aligned} \quad [6.47]$$

Substituting equations [6.47] into equation [6.46] leads to:

$$U_i^{n+1} = \begin{cases} U_i^n + 2 \frac{F_{i-1}^n - F_i^n}{\Delta x_{i-1/2} + \Delta x_{i+1/2}} \Delta t + \Delta t S_i^n & \text{if } \lambda_i^{n+1/2} \geq 0 \\ U_i^n + 2 \frac{F_i^n - F_{i+1}^n}{\Delta x_{i-1/2} + \Delta x_{i+1/2}} \Delta t + \Delta t S_i^n & \text{if } \lambda_i^{n+1/2} \leq 0 \end{cases} \quad [6.48]$$

In the implicit version of the scheme, the flux and the source term are calculated using the unknown values at the time level $n + 1$:

$$U_i^{n+1} = \begin{cases} U_i^n + 2 \frac{F_{i-1}^{n+1} - F_i^{n+1}}{\Delta x_{i-1/2} + \Delta x_{i+1/2}} \Delta t + \Delta t S_i^{n+1} & \text{if } \lambda_i^{n+1/2} \geq 0 \\ U_i^n + 2 \frac{F_i^{n+1} - F_{i+1}^{n+1}}{\Delta x_{i-1/2} + \Delta x_{i+1/2}} \Delta t + \Delta t S_i^{n+1} & \text{if } \lambda_i^{n+1/2} \leq 0 \end{cases} \quad [6.49]$$

The total amount of U contained within the control volume is conserved. Indeed, the mass is defined as:

$$M_T^n = \sum_{i=1}^M \Delta x_i U_i^n \quad [6.50]$$

The flux $F_{i+1/2}^{n+1/2}$ leaving the control volume i across the interface $i + 1/2$ enters the control volume $i + 1/2$ through the same interface. Therefore, no mass is gained or lost during the time step.

This conservative version of the finite difference approach is sometimes referred to as “finite differences with control volume”.

6.3.4. Application examples

The conservative scheme is applied to the same test cases as in section 6.2.3, with the same test parameters (Tables 6.1 and 6.2).

The results of the first test, which deals with the linear advection of a concentration profile, are illustrated by Figures 6.15–16.

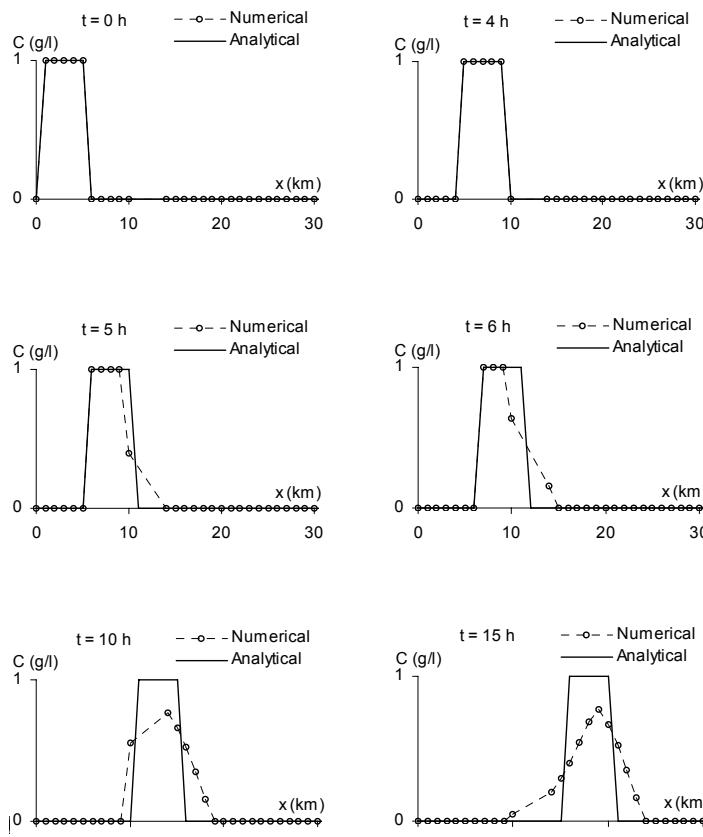


Figure 6.15. Pure advection of a square concentration signal. Analytical solution and numerical solution calculated by the explicit upwind scheme on an irregular grid

In contrast with the first-order MOC, the integral of the numerical profile obtained using the conservative upwind scheme is the same at all times. This is confirmed by Figure 6.16, which shows the variation in the total mass of contaminant in the domain. The mass is constant in the limit of the precision of the

computer. It must be stressed however that the conservative upwind scheme remains diffusive, which leads to an underestimation of the peak concentration compared to the analytical solution.

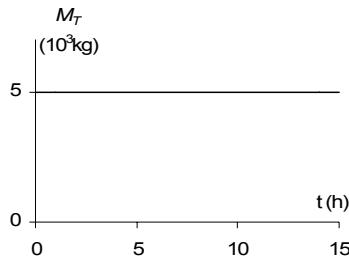


Figure 6.16. Pure advection of a square concentration signal calculated by the conservative, explicit upwind scheme. Total mass of contaminant in the domain as a function of time

The second test, that deals with a Riemann problem for the inviscid Burgers equation, is applied to the conservation form [1.69]. In contrast with the application to the first-order MOC in section 6.2.3, the three options [6.38] give the same result. Conservation being ensured intrinsically by the scheme, the front propagates at the correct speed (Figure 6.17). Nevertheless, the scheme remains diffusive and the shape of the front is altered by numerical smoothing.

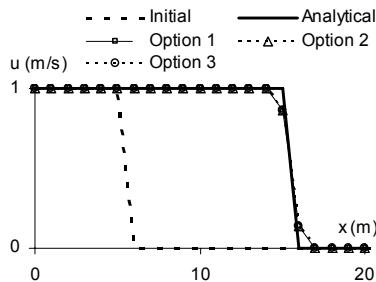


Figure 6.17. Analytical solution and numerical profiles computed by the conservative, explicit upwind scheme for the three calculation options [6.38]

6.4. The Preissmann scheme

6.4.1. Formulation

The Preissmann scheme [PRE 61a-c] is a conservative scheme. It is used in a number of commercially available packages for the simulation of open channel

hydraulics such as flows in rivers, urban drainage and sewer systems. The discretization follows the same rule for a hyperbolic system as for a scalar, conservation law. The scheme uses the four computational points that define the corners of a “box” in the phase space (Figure 6.18), hence the name of “box scheme”.

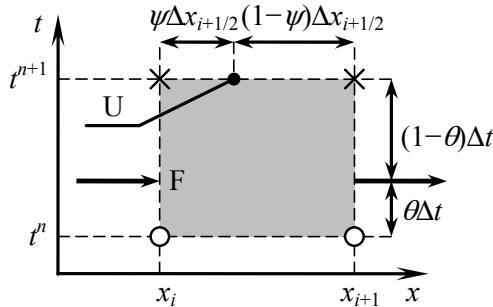


Figure 6.18. Definition sketch for the control volume in the Preissmann scheme

A balance over the control volume materialized by the gray-shaded area in Figure 6.18 gives:

$$\frac{U_{i+1/2}^{n+1} - U_{i+1/2}^n}{\Delta t} + \frac{F_{i+1}^{n+1/2} - F_i^{n+1/2}}{\Delta t} = S_{i+1/2}^{n+1/2} \quad [6.51]$$

where $U_{i+1/2}^n$ is the average value of U over the segment $[x_i, x_{i+1}]$ at the time level n , $F_i^{n+1/2}$ is the average of the flux F at the point i over the time step and $S_{i+1/2}^{n+1/2}$ is the average value of the source term S over the segment $[x_i, x_{i+1}]$ between the time levels n and $n + 1$. The following estimates are used for $U_{i+1/2}^n$ and $F_i^{n+1/2}$:

$$\left. \begin{aligned} U_{i+1/2}^n &\approx (1 - \psi)U_i^n + \psi U_{i+1}^n \\ F_i^{n+1/2} &\approx (1 - \theta)F_i^n + \theta F_{i+1}^n \end{aligned} \right\} \quad [6.52]$$

where θ and ψ are parameters ranging from 0 to 1. θ is usually called the time-centering, or implicitation, parameter. Most practical implementations of the scheme use the value $\psi = 1/2$. The same weight is then given to the waves with positive and negative wave speeds.

In this case the numerical solution is unconditionally stable for $\theta \geq 1/2$, while it is unconditionally unstable for $\theta < 1/2$. Substituting equations [6.52] into equation [6.51] gives:

$$\frac{(1-\psi)(U_i^{n+1} - U_i^n) + \psi(U_{i+1}^{n+1} - U_{i+1}^n)}{\Delta t} + \frac{(1-\theta)(F_{i+1}^n - F_i^n) + \theta(F_{i+1}^{n+1} - F_i^{n+1})}{\Delta x_{i+1/2}} = S_{i+1/2}^{n+1/2} \quad [6.53]$$

Multiplying by Δt and rearranging leads to:

$$(1-\psi)U_i^{n+1} + \psi U_{i+1}^{n+1} + \frac{\theta \Delta t}{\Delta x_{i+1/2}} (F_{i+1}^{n+1} - F_i^{n+1}) = (1-\psi)U_i^n + \psi U_{i+1}^n + \frac{(1-\theta)\Delta t}{\Delta x_{i+1/2}} (F_i^n - F_{i+1}^n) + \Delta t S_{i+1/2}^{n+1/2} \quad [6.54]$$

Equation [6.54] is a vector equation, that is, a system of m scalar equations. In the general case, the nonlinearity of F with respect to U makes the solution of the system [6.54] computationally intensive. In a number of approaches [CUN 80], solving a linearized version of the system makes the solution process easier:

$$\left. \begin{aligned} U_i^{n+1} &= U_i^n + \Delta U_i \\ F_i^{n+1} &= F_i^n + \frac{\partial F}{\partial U} \Delta U_i = F_i^n + A_i^{n+1/2} \Delta U_i \end{aligned} \right\} \quad [6.55]$$

where the matrix $A_i^{n+1/2}$ is an approximation of the average value of the Jacobian matrix $A = \partial F / \partial U$ between the time levels n and $n + 1$. How this matrix should be approximated is the subject of section 6.4.2. Equation [6.53] can be rewritten as:

$$\frac{(1-\psi)\Delta U_i + \psi \Delta U_{i+1}}{\Delta t} + \frac{\theta(A_{i+1}^{n+1/2} \Delta U_{i+1} - A_i^{n+1/2} \Delta U_i)}{\Delta x_{i+1/2}} = S_{i+1/2}^{n+1/2} - \frac{F_{i+1}^n - F_i^n}{\Delta x_{i+1/2}} \quad [6.56]$$

Simplifying by Δt leads to:

$$\left[(1 - \psi)I - \frac{\theta \Delta t A_i^{n+1/2}}{\Delta x_{i+1/2}} \right] \Delta U_i + \left[\psi I + \frac{\theta \Delta t A_{i+1}^{n+1/2}}{\Delta x_{i+1/2}} \right] \Delta U_{i+1} = \\ \Delta t S_{i+1/2}^{n+1/2} - \Delta t \frac{F_{i+1}^n - F_i^n}{\Delta x_{i+1/2}} \quad [6.57]$$

where I is the identity matrix. System [6.57] is linear and can be solved using standard matrix inversion techniques. When a scalar equation is to be solved, the matrix A becomes the wave speed λ and equation [6.57] simplifies into:

$$(1 - \psi - \theta C r_i^{n+1/2}) \Delta U_i + (\psi + \theta C r_{i+1}^{n+1/2}) \Delta U_{i+1} = \\ \Delta t S_{i+1/2}^{n+1/2} - \Delta t \frac{F_{i+1}^n - F_i^n}{\Delta x_{i+1/2}} \quad [6.58]$$

where the average Courant numbers at the points i and $i + 1$ are given by:

$$\left. \begin{aligned} C r_i^{n+1/2} &= \frac{\lambda_i^n + \lambda_i^{n+1}}{2} \frac{\Delta t}{\Delta x_{i+1/2}} \\ C r_{i+1}^{n+1/2} &= \frac{\lambda_{i+1}^n + \lambda_{i+1}^{n+1}}{2} \frac{\Delta t}{\Delta x_{i+1/2}} \end{aligned} \right\} \quad [6.59]$$

6.4.2. Estimation of nonlinear terms – algorithmic aspects

The Jacobian matrix A and the source term S being functions of U , the average values $A_i^{n+1/2}$ and $S_{i+1/2}^{n+1/2}$ over the time steps necessarily depend on the (unknown) value of U at the time level $n + 1$. The following expressions are used in practice:

$$\left. \begin{aligned} S_{i+1/2}^{n+1/2} &= (1 - \theta)[(1 - \psi)S_i^n + \psi S_{i+1}^n] + \theta[(1 - \psi)S_i^{n+1} + \psi S_{i+1}^{n+1}] \\ A_i^{n+1/2} &= (1 - \theta)A_i^n + \theta A_i^{n+1} \end{aligned} \right\} \quad [6.60]$$

The calculation procedure is iterative. It consists of the following steps:

- 1) Initialize $A_i^{n+1/2}$ and $S_{i+1/2}^{n+1/2}$ using the values at the beginning of the time step:

$$\left. \begin{aligned} S_{i+1/2}^{n+1/2} &\approx (1-\psi)S_i^n + \psi S_{i+1}^n \\ A_i^{n+1/2} &= A_i^n \end{aligned} \right\} [6.61]$$

2) Solve system [6.57] using estimates [6.61].

3) Update $A_i^{n+1/2}$ and $S_{i+1/2}^{n+1/2}$ using equation [6.60].

Steps 2–3 must be repeated until convergence is achieved. In practical applications such as the Saint Venant equations, only a few iterations are needed.

The Preissmann scheme has the drawback that it cannot be used in a straightforward manner in applications where the speed of the waves change sign over the computational domain [MES 97]. In such a case, instability may occur. A new version of the scheme, based on an approach similar to flux splitting (see section 6.7), has been proposed in [JOH 02].

6.4.3. Numerical applications

This chapter details the application of the Preissmann scheme to the test cases presented in sections 6.2.3 and 6.3.4, with the difference that the linear advection equation is solved on a regular grid (see Table 6.3). Using different values for the time step leads to different values of the Courant number. In this test, the influence of θ on the accuracy of the numerical solution is investigated.

The Preissmann scheme is applied to the linear advection equation by defining U and F as $U = AC$, $F = AuC$ in equation [6.58] and dividing by A .

Symbol	Meaning	Value
A	River cross-sectional area	1,000 m ²
C_i^0	Initial concentration	1 g/l for 1 km $\leq x \leq$ 5 km, 0 g/l otherwise
C_G	Concentration at the left-hand boundary	0 g/l
L	Length of the domain	30 km
u	Flow velocity	1 km/hr
Δt	Computational time step	0.5 hr, 1 hr, 2 hr
Δx	Cell size	1 km

Table 6.3. Physical and numerical parameters for the numerical solution of the linear advection equation on a regular grid

$$(\psi + \theta \text{Cr})C_{i+1}^{n+1} = (\psi - 1 + \theta \text{Cr})C_i^{n+1} + [1 - \psi + (1 - \theta)\text{Cr}]C_i^n + [\psi - (1 - \theta)\text{Cr}]C_{i+1}^n \quad [6.62]$$

Equation [6.62] is a recurrence relationship between the unknown values of C at the points i and $i+1$. In the case of a positive flow velocity, the computational domain is swept from left to right by rewriting equation [6.62] as:

$$\begin{aligned} C_{i+1}^{n+1} &= \frac{\psi - 1 + \theta \text{Cr}}{\psi + \theta \text{Cr}} C_i^{n+1} + \frac{1 - \psi + (1 - \theta)\text{Cr}}{\psi + \theta \text{Cr}} C_i^n \\ &\quad + \frac{\psi - (1 - \theta)\text{Cr}}{\psi + \theta \text{Cr}} C_{i+1}^n \end{aligned} \quad [6.63]$$

Figure 6.19 illustrates the behavior of the numerical solution after 10 hours for various values of the numerical parameter θ . Note that:

- for $\theta = \psi = 1/2$ the scheme is dispersive if the Courant number is different from one. The numerical dispersion is reflected by the oscillations in the computed profile. When the Courant number is smaller than one the oscillations propagate faster than the analytical solution. For Courant numbers larger than one the oscillations propagate more slowly than the analytical solution;
- increasing the value of θ induces numerical diffusion, which contributes to dampening the oscillations. Numerical dispersion still occurs but its effects are “hidden” by those of numerical diffusion. Using $\theta = 0.7$ with a Courant number $\text{Cr} = 1/2$ allows the oscillations to be almost completely eliminated (Figure 6.19).

In the second test the Preissmann scheme is applied to the Riemann problem for the inviscid Burgers equation. The parameters of the test case are given in Table 6.4.

Symbol	Meaning	Value
L	Length of the domain	30 m
u_i^0	Initial flow velocity	2 m/s for $x \leq 5$ m, 1 m/s otherwise
Δt	Computational time step	0.5 s, 1 s
Δx	Cell width	1 m

Table 6.4. Physical and numerical parameters for the solution of the inviscid Burgers equation

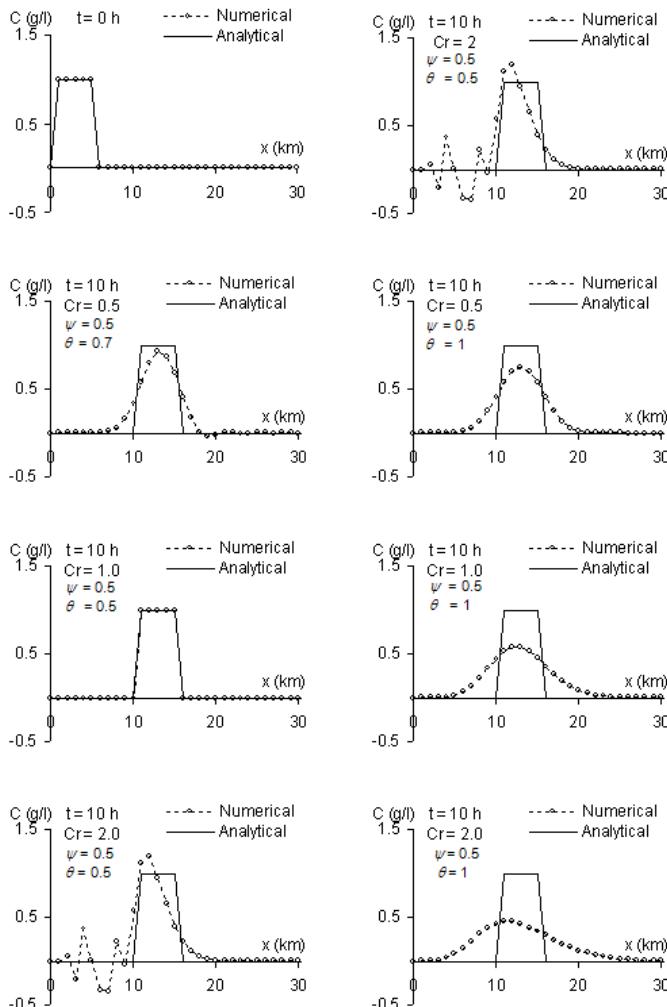


Figure 6.19. Pure advection of a square concentration signal.
Comparison between the numerical and analytical solutions for various values of the parameter θ and the Courant number

The linearized version [6.58] of the scheme is applied with $U = u$ and $F = u^2/2$:

$$\frac{1 - \psi - \theta Cr_i^{n+1/2}}{\Delta t} \Delta u_i + \frac{\psi + \theta Cr_{i+1}^{n+1/2}}{\Delta t} \Delta u_{i+1} = \frac{(u_i^n)^2 - (u_{i+1}^n)^2}{2\Delta x_{i+1/2}} \quad [6.64]$$

where the following expressions are used for the Courant number:

$$\left. \begin{aligned} \text{Cr}_i^{n+1/2} &= \frac{u_i^n + u_i^{n+1}}{2} \frac{\Delta t}{\Delta x_{i+1/2}} \\ \text{Cr}_{i+1}^{n+1/2} &= \frac{u_{i+1}^n + u_{i+1}^{n+1}}{2} \frac{\Delta t}{\Delta x_{i+1/2}} \end{aligned} \right\} [6.65]$$

For a positive flow velocity the wave speed is positive and the domain is swept directly from the left-hand boundary to the right-hand boundary:

$$\Delta u_{i+1} = \frac{\Delta t \frac{(u_i^n)^2 - (u_{i+1}^n)^2}{2\Delta x_{i+1/2}} - (1 - \psi - \theta \text{Cr}_i^{n+1/2}) \Delta u_i}{(\psi + \theta \text{Cr}_{i+1}^{n+1/2})} [6.66]$$

which allows u_{i+1}^{n+1} to be calculated as:

$$u_{i+1}^{n+1} = u_{i+1}^n + \Delta u_{i+1} [6.67]$$

The analytical solution is a shock, the speed of which is the average of the speeds on both sides, i.e. 1.5 m/s. The numerical profile obtained at $t = 10$ s for various values of the time step and the parameter θ are compared to the analytical profile in Figure 6.20.

Four iterations are used for the determination of the Courant number. However, 2 iterations would have been sufficient in that the values obtained using 2 iterations and those obtained using 4 iterations differ by less than 1%.

As in the linear case, using $\theta = 1/2$ eliminates numerical diffusion and the effect of numerical dispersion becomes clearly visible. When the Courant number is larger than one (as is the case for $\Delta t = 1$ s) the oscillations appear behind the shock. When the Courant number is smaller than one (as is the case for $\Delta t = 0.5$ s) the oscillations appear ahead of the shock. Increasing θ even by a slight amount leads to a dramatic damping of the oscillations. For $\theta = 0.7$ the oscillations are almost absent from the profile. They disappear completely for $\theta = 1$.

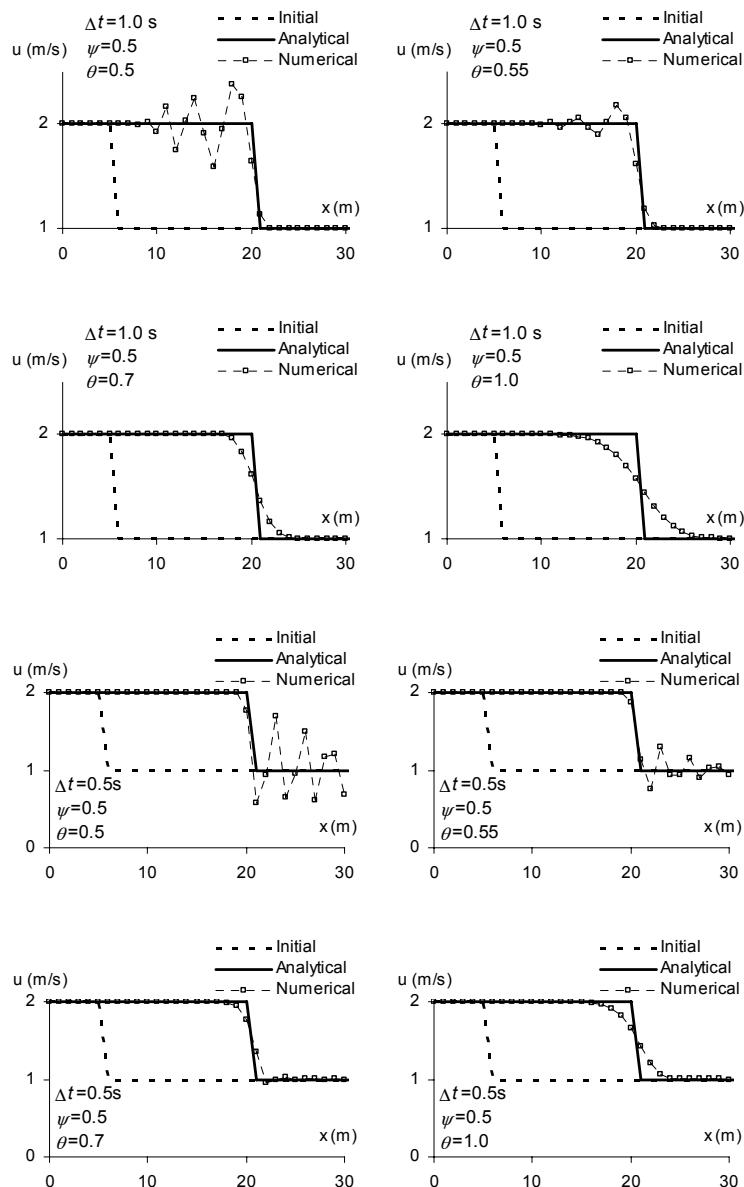


Figure 6.20. Riemann problem for the inviscid Burgers equation with the data in Table 6.4. Analytical and numerical solution at $t = 10$ s

The Preissmann scheme may lead to unstable solutions when applied to problems where the direction of the wave changes locally or becomes locally zero. This is the case with the Riemann problem specified in Table 6.2. The wave speed in the right state of the initial profile is zero. Adding numerical diffusion via the implicitation parameter θ does not always lead to a total damping of the oscillations. This is one of the reasons why the Preissmann scheme, although frequently applied to free surface flow simulations, remains restricted to the simulation of subcritical regimes.

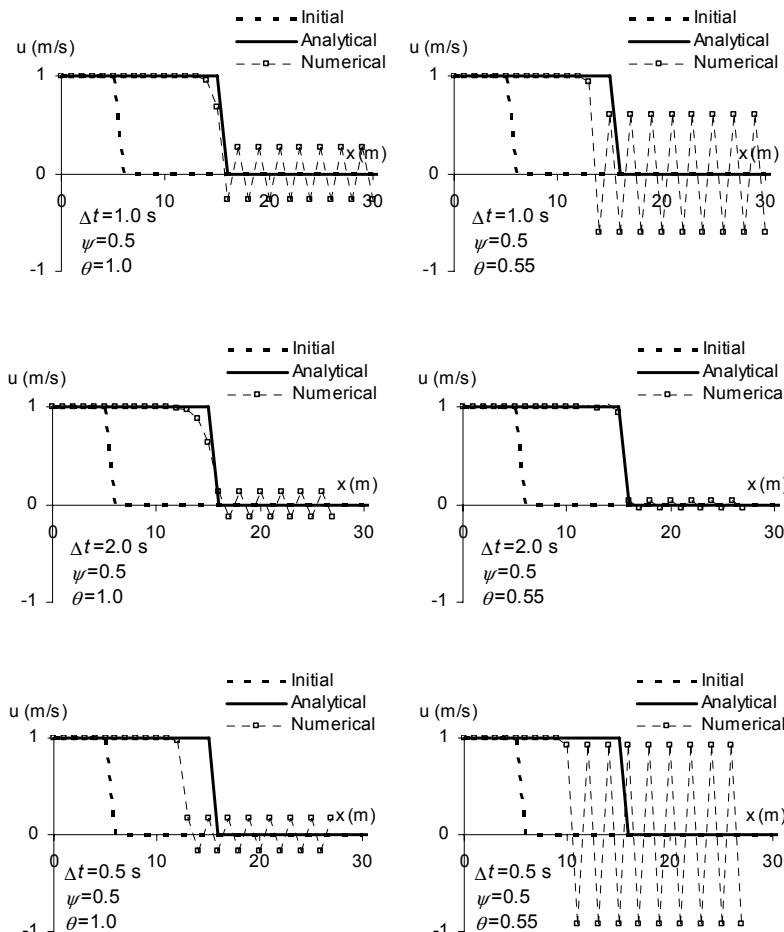


Figure 6.21. Riemann problem for the inviscid Burgers equation with the parameters given in Table 6.2. Analytical and numerical solution at $t = 10$ s

6.5. Centered schemes

6.5.1. The Crank-Nicholson scheme

The MOC and the upwind schemes presented in sections 6.2 and 6.3 are sensitive to the direction of the waves. As a consequence, a test should be carried out at each computational point in order to determine the direction from which the information comes and which computational points must be used. Moreover, characteristics-based and upwind schemes are considered to be too diffusive (see section B.1 for detailed considerations on numerical diffusion) in a number of computational applications, among which are turbulent flow simulations. Centered schemes aim to eliminate this drawback.

The derivative of the conserved variable with respect to time is estimated as in equation [6.39]:

$$\frac{\partial U}{\partial t} \approx \frac{U_i^{n+1} - U_i^n}{\Delta t}$$

The derivative of the flux with respect to space is estimated as:

$$\frac{\partial F}{\partial x} \approx \frac{1}{2} \frac{F_{i+1}^n - F_{i-1}^n}{2\Delta x} + \frac{1}{2} \frac{F_{i+1}^{n+1} - F_{i-1}^{n+1}}{2\Delta x} \quad [6.68]$$

Note that the estimate of the space derivative is symmetric with respect to the point i , hence the term “centered scheme”. Substituting equations [6.39] and [6.68] into equation [2.1] leads to:

$$U_i^{n+1} + \frac{\Delta t}{4\Delta x} (F_{i+1}^{n+1} - F_{i-1}^{n+1}) = U_i^n + \frac{\Delta t}{4\Delta x} (F_{i+1}^n - F_{i-1}^n) \quad [6.69]$$

A domain with M computational points (including the boundary points) allows $M-2$ equations [6.69] to be written. The values of U at the remaining points $i=1$ and $i=M$ cannot be computed using equation [6.69] because the points 0 and $M+1$ do not exist. Since the vector U has m components, $2m$ unknown values are to be determined. Except in the particular case where all the characteristics enter the computational domain across both boundaries, the boundary conditions do not allow the solution to be determined uniquely. Additional conditions, such as zero gradient conditions or fixed values, must be prescribed at the points $i=1$ and $i=M$. Another

possibility is to discretize the governing equations at the first and last points in the domain using schemes that involve only two adjacent points in space, such as the CIR scheme or the Preissmann scheme.

If the flux function is nonlinear, system [6.69] is a nonlinear system. Its solution may be computationally demanding. The computational effort can be reduced to some extent by linearizing the system as proposed in equations [6.55], recalled hereafter:

$$\left. \begin{aligned} U_i^{n+1} &= U_i^n + \Delta U_i \\ F_i^{n+1} &= F_i^n + \frac{\partial F}{\partial U} \Delta U_i = F_i^n + A_i^{n+1/2} \Delta U_i \end{aligned} \right\}$$

Substituting equations [6.55] into equation [6.69] leads to:

$$\Delta U_i + \frac{\Delta t}{4\Delta x} A_i^{n+1/2} (A_{i+1}^{n+1/2} \Delta U_{i+1} - A_{i-1}^{n+1/2} \Delta U_{i-1}) = 0 \quad [6.70]$$

where the Jacobian matrices $A_{i\pm 1}^{n+1/2}$ are estimated as explained in section 6.4.2.

6.5.2. Centered schemes with Runge-Kutta time stepping

The Crank-Nicholson scheme presented in the previous section is an implicit scheme. This implies an iterative linearization of the equations, followed by the solution of a system of algebraic equations in the form [6.70]. The question thus arises of the possibility to develop explicit, centered schemes so as to preserve the non-dissipative character of the centered formulation, while making the computational procedure simpler. The following discretization may be seen as a good candidate for the central discretization of the conservation form [2.2] or the non-conservation form [2.5]:

$$\left. \begin{aligned} \frac{\partial U}{\partial t} &\approx \frac{U_i^{n+1} - U_i^n}{\Delta t} \\ \frac{\partial F}{\partial x} &\approx \frac{F_{i+1}^n - F_{i-1}^n}{2\Delta x} \\ \frac{\partial U}{\partial x} &\approx \frac{U_{i+1}^n - U_{i-1}^n}{2\Delta x} \end{aligned} \right\} \quad [6.71]$$

Applying the discretization above to the linear advection equation [1.48] leads to the following numerical scheme:

$$C_i^{n+1} = C_i^n + \frac{u\Delta t}{2\Delta x} (C_{i-1}^n - C_{i+1}^n) \quad [6.72]$$

However, a simple stability analysis [VIC 82] reveals that the numerical solution obtained using equation [6.72] is unconditionally unstable. Stable solutions can be obtained only if the derivative with respect to space is integrated using multiple step integration procedures, such as the Runge-Kutta technique. The most widely used options are the second- and fourth-order Runge-Kutta time stepping techniques. The principle of such techniques is outlined hereafter.

Assume that the scheme can be expressed in the form:

$$U_i^{n+1} = U_i^n + MU_i^n \quad [6.73]$$

where M is a matrix operator. If equation [6.72] is to be used, M is given by:

$$M = \frac{u\Delta t}{2\Delta x} (\delta_{-1} + \delta_{+1}) \quad [6.74]$$

where δ is the shift operator, defined as:

$$\left. \begin{aligned} \delta_{-1} U_i^n &= U_{i-1}^n \\ \delta_{+1} U_i^n &= U_{i+1}^n \end{aligned} \right\} \quad [6.75]$$

The formula for the p th-order Runge-Kutta technique is:

$$U_i^{n+1} = U_i^n + \sum_{k=1}^p \frac{\Delta t^k}{k!} M^k U_i^n \quad [6.76]$$

where the notation M^k indicates that the operator M is to be applied k successive times. For instance, assuming that M is given by equation [6.72], the operator M^2 is defined as:

$$\begin{aligned} M^2 C_i^n &= M(MC_i^n) \\ &= \frac{u\Delta t}{2\Delta x} (MC_{i-1}^n - MC_{i+1}^n) \\ &= \frac{u\Delta t}{2\Delta x} \left[\frac{u\Delta t}{2\Delta x} (C_{i-2}^n - C_i^n) - \frac{u\Delta t}{2\Delta x} (C_i^n - C_{i+2}^n) \right] \\ &= \left(\frac{u\Delta t}{2\Delta x} \right)^2 (C_{i-2}^n - 2C_i^n + C_{i+2}^n) \end{aligned} \quad [6.77]$$

As shown by a linear stability analysis, second-order Runge-Kutta methods do not yield stable solutions. The minimal order for which the numerical solution can be made stable is $p = 3$. The stability domains of the third- and fourth-order Runge-Kutta time stepping methods are given by (see [VIC 82]):

$$\left. \begin{array}{ll} |\text{Cr}| \leq 1.8 & (\text{order } 3) \\ |\text{Cr}| \leq 2.85 & (\text{order } 4) \end{array} \right\} \quad [6.78]$$

The range of stability of the centered scheme with third- and fourth-order Runge-Kutta time stepping is wider than that of classical explicit schemes. In contrast, the computational effort required by Runge-Kutta time stepping schemes is larger than that required by classical schemes in that the operator M must be applied several times over a given time step and the results of the successive applications must be combined linearly as in equation [6.76].

6.6. TVD schemes

6.6.1. Definitions

Consider the numerical solution U_i^n of a scalar conservation law over a one-dimensional domain, $i = 1, \dots, M$. The total variation TV^n of the solution over the domain at the time level n is defined as:

$$\text{TV}(U)^n = \sum_{i=1}^{M-1} |U_{i+1}^n - U_i^n| \quad [6.79]$$

The total variation serves as a quantitative indicator for the oscillatory character of the solution. A numerical scheme is said to be Total Variation Diminishing (TVD) if it satisfies the following property:

$$\text{TV}(U)^{n+1} \leq \text{TV}(U)^n \quad [6.80]$$

Applying a TVD scheme to an initially monotone numerical solution necessarily yields a monotone numerical solution.

A scheme is said to be monotony-preserving if the following conditions hold:

$$\left. \begin{array}{ll} U_i^n \leq U_{i-1}^n & \forall i \\ U_i^n \geq U_{i-1}^n & \forall i \end{array} \Rightarrow \quad \begin{array}{ll} U_i^{n+1} \leq U_{i-1}^{n+1} & \forall i \\ U_i^{n+1} \geq U_{i-1}^{n+1} & \forall i \end{array} \right\} \quad [6.81]$$

A TVD scheme is always monotony-preserving.

The monotony property has the advantage that spurious oscillations do not arise in the numerical solution. Monotone schemes are developed with the purpose of minimizing numerical diffusion, while preserving the monotony of the solution.

6.6.2. General formulation of TVD schemes

This section focuses on three-point TVD schemes as applied to the linear advection equation. The generalization of such schemes to hyperbolic systems of conservation laws is dealt with in section 6.7. Applying the second-order MOC seen in section 6.2 to the linear advection equation on a regular grid leads to equation [6.21]. In the particular case of a zero source term, equation [6.21] simplifies into:

$$U_i^{n+1} = (\text{Cr} + 1) \frac{\text{Cr}}{2} U_{i-1}^n + (1 - \text{Cr}^2) U_i^n + (\text{Cr} - 1) \frac{\text{Cr}}{2} U_{i+1}^n \quad [6.82]$$

Equation [6.82] is also known as the Lax-Wendroff scheme [LAX 60]. Equation [6.82] can be obtained from a second-order consistency analysis by specifying a zero numerical diffusion condition. The scheme may also be written as the combination of an upwind scheme and a function of the variations in the conserved variable. Assuming that the advection velocity is positive, equation [6.82] can be rewritten as:

$$\begin{aligned} U_i^{n+1} &= \text{Cr} U_{i-1}^n + (1 - \text{Cr}) U_i^n \\ &\quad + \frac{(\text{Cr} - 1)\text{Cr}}{2} [(U_{i+1}^n - U_i^n) - (U_i^n - U_{i-1}^n)] \\ &= U_i^n - (U_i^n - U_{i-1}^n) \text{Cr} \\ &\quad + \frac{(\text{Cr} - 1)\text{Cr}}{2} [(U_{i+1}^n - U_i^n) - (U_i^n - U_{i-1}^n)] \end{aligned} \quad [6.83]$$

Reasoning by symmetry leads to the following expression for a negative advection velocity:

$$\begin{aligned} U_i^{n+1} &= U_i^n + (U_i^n - U_{i+1}^n) \text{Cr} \\ &\quad + \frac{(\text{Cr} + 1)\text{Cr}}{2} [(U_{i+1}^n - U_i^n) - (U_i^n - U_{i-1}^n)] \end{aligned} \quad [6.84]$$

Equations [6.83] and [6.84] can be rewritten in the following, condensed form:

$$\begin{aligned} U_i^{n+1} = & U_i^n - \frac{|\text{Cr}| + \text{Cr}}{2} (U_i^n - U_{i-1}^n) + \frac{|\text{Cr}| - \text{Cr}}{2} (U_{i+1}^n - U_i^n) \\ & + \frac{(|\text{Cr}| - 1)|\text{Cr}|}{2} [(U_{i+1}^n - U_i^n) - (U_i^n - U_{i-1}^n)] \end{aligned} \quad [6.85]$$

Scheme [6.85] yields oscillatory solutions in the neighborhood of steep fronts. TVD schemes aim to limit the contribution of the variations in U locally when the gradient of the solution becomes too large. To do so, scheme [6.85] is modified into:

$$\begin{aligned} U_i^{n+1} = & U_i^n - \frac{|\text{Cr}| + \text{Cr}}{2} (U_i^n - U_{i-1}^n) + \frac{|\text{Cr}| - \text{Cr}}{2} (U_{i+1}^n - U_i^n) \\ & + \frac{(|\text{Cr}| - 1)|\text{Cr}|}{2} [(U_{i+1}^n - U_i^n)\phi_{i+1/2}^n - (U_i^n - U_{i-1}^n)\phi_{i-1/2}^n] \end{aligned} \quad [6.86]$$

where the so-called limiting function ϕ , also referred to as a limiter, is a real number between 0 and 1. The value of ϕ depends on the regularity (or smoothness) of the profile. If the profile is regular (or smooth) enough the limiter is set to unity and scheme [6.86] is equivalent to the original Lax-Wendroff scheme. When the solution becomes irregular, with sudden changes in the local slope of the solution profile, ϕ is decreased, which is equivalent to adding numerical diffusion. In the particular case $\phi = 0$, equation [6.86] is equivalent to the upwind scheme. The smoothness of the solution is characterized by a monotony indicator θ defined as:

$$\theta_{i+1/2}^n = \begin{cases} \frac{U_i^n - U_{i-1}^n}{U_{i+1}^n - U_i^n} & \text{if } \text{Cr} \geq 0 \\ \frac{U_{i+2}^n - U_{i+1}^n}{U_{i+1}^n - U_i^n} & \text{if } \text{Cr} \leq 0 \end{cases} \quad [6.87]$$

The monotony indicator is the ratio of the slope of the profile upstream of the point i to the slope of the profile downstream of it (Figure 6.22). The indicator is positive when the solution is monotone and negative otherwise. It is zero if the slope of the profile is zero upstream of the point i . It is infinite if the profile is horizontal downstream of the point i . The various TVD schemes proposed in the literature use different formulations for the limiter $\phi(\theta)$. The conditions that must be fulfilled by the function $\phi(\theta)$ are detailed in the next section.

Note that a number of schemes proposed in the literature before the appearance of TVD schemes were found later to be particular cases of the general formulation

[6.86], provided that $\phi(\theta)$ is defined appropriately. This is the case with the schemes listed in Table 6.5. Note that not all of these schemes are TVD.

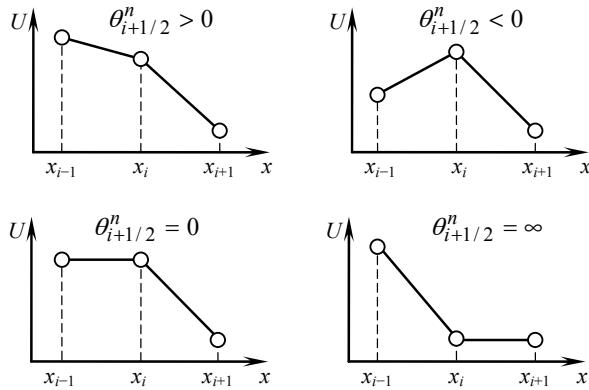


Figure 6.22. Monotony indicator for various possible configurations (sketched here only for a positive wave speed)

Scheme	Equivalent limiter
Upwind scheme	$\phi(\theta) = 0$
Lax-Wendroff scheme	$\phi(\theta) = 1$
Beam and Warming scheme	$\phi(\theta) = \theta$
Fromm scheme	$\phi(\theta) = (1 + \theta) / 2$

Table 6.5. Expression of the limiter for classical schemes

6.6.3. Harten's and Sweby's criteria

The necessary conditions for a scheme to be TVD were derived by Harten [HAR 83a, HAR 84]. Harten's analysis focuses on schemes in the form:

$$U_i^{n+1} = U_i^n - (U_i^n - U_{i-1}^n)a_{i-1/2}^n + (U_{i+1}^n - U_i^n)b_{i+1/2}^n \quad [6.88]$$

Such schemes are TVD provided that a and b satisfy the following conditions:

$$\left. \begin{aligned} a_{i-1/2}^n &\geq 0 \\ b_{i+1/2}^n &\geq 0 \\ a_{i+1/2}^n + b_{i+1/2}^n &\leq 1 \end{aligned} \right\} \quad [6.89]$$

Note that equation [6.86] can be written in the form [6.88] by defining a and b as:

$$\left. \begin{aligned} a_{i-1/2}^n &= Cr + \frac{(Cr-1)Cr}{2} \left[\phi(\theta_{i-1/2}^n) - \frac{\phi(\theta_{i+1/2}^n)}{\theta_{i+1/2}^n} \right] \\ b_{i+1/2}^n &= 0 \end{aligned} \right\} \quad [6.90]$$

Also note that definitions [6.90] are given only for a positive wave speed. Substituting equations [6.90] into equations [6.89] leads to:

$$0 \leq Cr + \frac{(Cr-1)Cr}{2} \left[\phi(\theta_{i-1/2}^n) - \frac{\phi(\theta_{i+1/2}^n)}{\theta_{i+1/2}^n} \right] \leq 1 \quad [6.91]$$

Since the Courant number is assumed to be smaller than one for the sake of stability, equation [6.91] becomes:

$$\left. \begin{aligned} 0 \leq \phi(\theta) \leq 2 \\ 0 \leq \phi(\theta) \leq 2\theta \end{aligned} \right\} \quad \text{for } \theta \geq 0 \quad [6.92]$$

If θ is negative the limiter ϕ is set to zero and the scheme becomes locally equivalent to the TVD, explicit upwind scheme. This leads to the final set of conditions for θ :

$$\left. \begin{aligned} \phi(0) &= 0 && \text{for } \theta \leq 0 \\ 0 \leq \phi(\theta) \leq \min \max(2, 2\theta) & && \text{for } \theta \geq 0 \end{aligned} \right\} \quad [6.93]$$

Figure 6.23 provides a representation of the TVD region in the (θ, ϕ) space. The TVD region is the gray-shaded area in the figure. Note that the Lax-Wendroff and the explicit upwind scheme are represented by the straight lines $\phi = 1$ and $\phi = 0$ respectively.

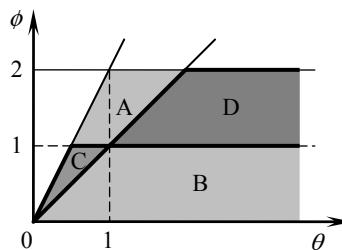


Figure 6.23. Representation of the TVD region in the (θ, ϕ) space. General TVD region (light gray-shaded area), Sweby's TVD region (dark gray-shaded area)

Limiters that belong to region A exhibit a compressive behavior that tends to make the fronts artificially steeper, while limiter functions located in region B induce an artificial smoothing of the solution via numerical diffusion. Sweby [SWE 84] suggested that optimal limiting is achieved in regions B and C in Figure 6.23, hence the following criteria for ϕ :

$$\left. \begin{array}{ll} \phi(\theta) = 0 & \text{for } \theta \leq 0 \\ \theta \leq \phi(\theta) \leq \min(2\theta, 1) & \text{for } \theta \in [0, 1] \\ 1 \leq \phi(\theta) \leq \min(2\theta, 2) & \text{for } \theta \geq 1 \end{array} \right\} \quad [6.94]$$

A number of classical limiters are presented in section 6.6.4.

6.6.4. Classical limiters

A number of limiters have been proposed in the literature. Well-known examples are the minmod, MC, Superbee and Van Leer limiters. The corresponding formulae are given in Table 6.6. A graphical representation of the limiting functions in the (θ, ϕ) space is provided in Figure 6.24. A limiter, that is very close to the MC limiter, is also proposed for the Lax-Wendroff scheme. The minmod limiter, that is the most diffusive limiter among the limiters presented here, follows the lower bound of Sweby's TVD region. The Superbee limiter, that follows the upper bound of Sweby's TVD region, may prove to be overcompressive in some cases. The MC limiter follows the minimum between of Beam and Warming's scheme [WAR 76], Fromm's scheme [FRO 68] and the maximum permissible value $\phi = 2$. Van Leer's limiter is a rational function of θ .

Limiter / scheme	Formula	Figure
Upwind scheme	$\phi(\theta) = 0$	6.24a
Lax-Wendroff scheme	$\phi(\theta) = 1$	6.24a
Monotone Lax-Wendroff scheme	$\phi(\theta) = \max[0, \min(1, 2\theta)]$	6.24b
Minmod limiter	$\phi(\theta) = \max[0, \min(1, \theta)]$	6.24c
Superbee limiter	$\phi(\theta) = \max[0, \min(1, 2\theta), \min(2, \theta)]$	6.24d
MC limiter	$\phi(\theta) = \max\{0, \min[(1 + \theta) / 2, 2, 2\theta]\}$	6.24e
Van Leer's limiter	$\phi(\theta) = (\theta + \theta) / (1 + \theta)$	6.24f

Table 6.6. Expression of the limiter ϕ for various classical schemes and limiters available from the literature. All limiters take a zero value for negative values of θ

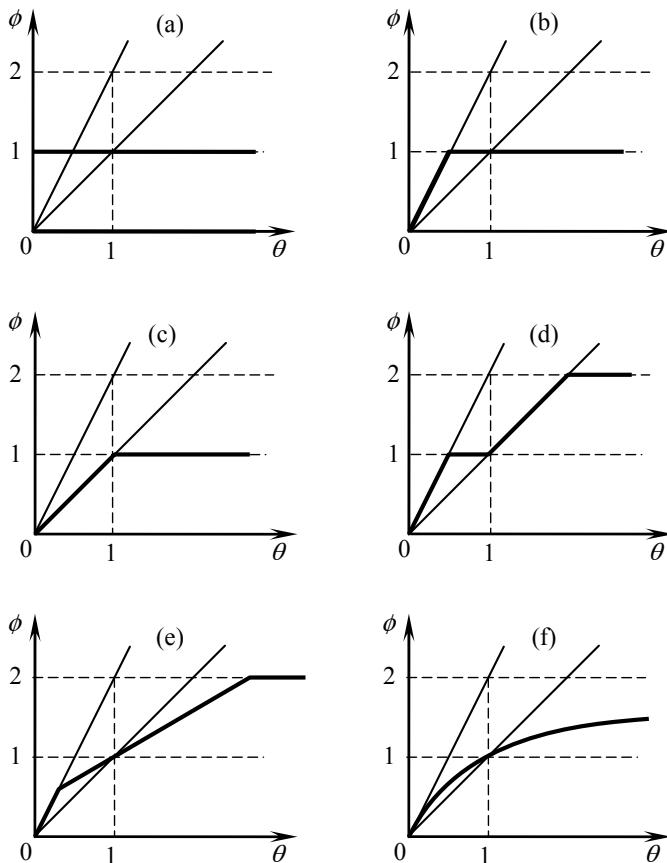


Figure 6.24. Representation of classical schemes and limiters in the (θ, ϕ) space. Upwind scheme and Lax-Wendroff scheme (a), Monotone Lax-Wendroff (b), minmod limiter (c), Superbee limiter (d), MC limiter (e), Van Leer's limiter (f)

6.6.5. Computational example

The performance of the various schemes and limiters are illustrated by an application to the linear advection equation. The pure advection of a square concentration profile in a uniform velocity field is simulated by solving equation [1.48] numerically. The parameters of the test case are given in Table 6.7. The Courant number is equal to 0.5, which is the configuration where the effect of the limiter (or the absence of limiter) is maximum because the quantity $(Cr - 1) Cr$, that is the coefficient of the limiting function, is maximum (see e.g. equation [6.86]

or equation [6.90]). Figure 6.25 shows the numerical solutions computed at $t = 50$ s using a 15 m wide initial, square concentration profile.

Symbol	Meaning	Value
u	Advection velocity	1 m/s
Δt	Computational time step	0.5 s
Δx	Cell width	1 m

Table 6.7. Pure advection of a square concentration profile in a uniform velocity field.
Parameters of the test case

The performance of the schemes with limiters is intermediate between that of the first-order, diffusive upwind scheme and the second-order, dispersive Lax-Wendroff scheme. The amplitude and the total variation of the solutions at $t = 50$ s is shown in Table 6.8. The non-monotone character of the Lax-Wendroff scheme is illustrated by the increase in the total variation between $t = 0$ s and $t = 50$ s.

The compressive Superbee limiter is the only one that allows the amplitude of the signal to be preserved after 100 time steps. The MC and Van Leer's limiter give very similar results. The monotone Lax-Wendroff scheme yields an asymmetrical solution, with a larger amplitude and total variation than the minmod limiter.

Solution method	Amplitude at $t = 50$ s	Total variation at $t = 50$ s
Analytical solution	1	2
Upwind scheme	0.837	1.673
Lax-Wendroff scheme	1.373	2.990
Monotone Lax-Wendroff scheme	0.996	1.992
Minmod limiter	0.981	1.962
Superbee limiter	1.000	1.999
MC limiter	0.999	1.998
Van Leer's limiter	0.998	1.997

Table 6.8. Amplitude and total variation of the numerical solution at $t = 50$ s

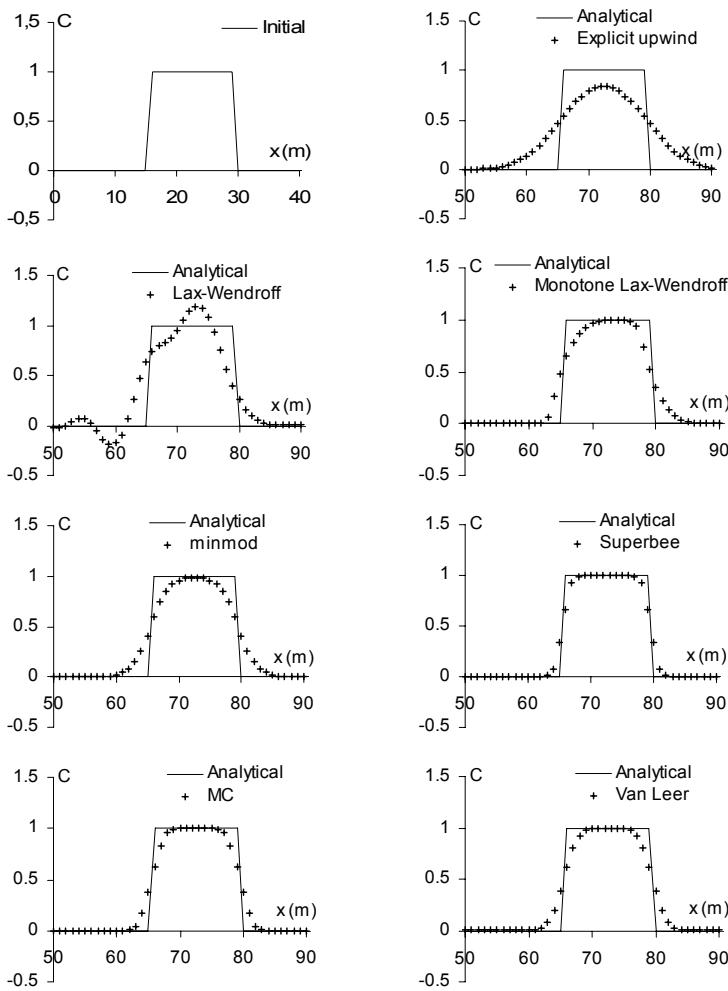


Figure 6.25. Pure advection of a square concentration profile (arbitrary units). Analytical and numerical solutions after 100 time steps with $Cr = 1/2$

6.7. The flux splitting technique

6.7.1. Principle of the approach

The flux splitting technique [STE 81], also known as the flux vector splitting technique or flux difference splitting technique, allows the TVD approach to be generalized to hyperbolic systems of conservation laws. The discretization can be

made conservative provided that a number of conditions are satisfied. The flux splitting technique is applicable to all upwind schemes, the formulation of which depends on the direction in which the waves propagate. In hyperbolic systems of conservation laws, the various waves propagate at different speeds and in different directions. The purpose of the flux splitting technique is precisely to account for the propagation direction of each of the waves in the formulation of the upwind terms. The flux splitting technique has been applied to a variety of hyperbolic systems. The original publications focused on the equations of gas dynamics [STE 81, DIC 85]. An application to the one-dimensional Saint Venant equations can be found in [ALC 92, HUB 00]. The purpose is to solve the non-conservation form [2.5]. The discretization can be made conservative as explained in section 6.7.2. For the sake of clarity, the source term is assumed to be zero hereafter. The non-conservation form [2.5] then simplifies into:

$$\frac{\partial U}{\partial t} + A \frac{\partial U}{\partial x} = 0 \quad [6.95]$$

The derivative of U with respect to time is classically discretized as:

$$\frac{\partial U}{\partial t} \approx \frac{U_i^{n+1} - U_i^n}{\Delta t} \quad [6.96]$$

while the derivative with respect to space is discretized as:

$$A \frac{\partial U}{\partial x} \approx A_{i-1/2}^+ \left(\frac{\partial U}{\partial x} \right)_{i-1/2}^{n+1/2} + A_{i+1/2}^- \left(\frac{\partial U}{\partial x} \right)_{i+1/2}^{n+1/2} \quad [6.97]$$

where the terms $(\partial U / \partial x)_{i-1/2}^{n+1/2}$ and $(\partial U / \partial x)_{i+1/2}^{n+1/2}$ are estimated over the intervals $[i-1, i]$ and $[i, i+1]$ respectively. The estimates may be obtained using a first-order, upwind formulation or one of the more complex TVD formulations seen in section 6.6. The superscript $n+1/2$ indicates that the terms are estimated between the time levels n and $n+1$. The estimate may be purely explicit, purely implicit or semi-implicit as in the Preissmann scheme. The flux splitting technique consists of estimating the matrices $A_{i-1/2}^+$ and $A_{i+1/2}^-$ so as to preserve the upwind character of the formulation, that is, by eliminating all the waves that do not propagate in the direction of the point i . This is achieved using the diagonal form [2.22]:

$$\frac{\partial W}{\partial t} + \Lambda \frac{\partial W}{\partial x} = 0 \quad [6.98]$$

where Λ is the diagonal matrix formed by the eigenvalues of A and W is the vector formed by the Riemann invariants. The matrix Λ is written as the sum of two matrices Λ^- and Λ^+ that contain only the negative and positive eigenvalues of A respectively:

$$\frac{\partial W}{\partial t} + (\Lambda^- + \Lambda^+) \frac{\partial W}{\partial x} = 0 \quad [6.99]$$

where Λ^- and Λ^+ are defined as:

$$\Lambda^- = \begin{bmatrix} \lambda^{(1)} & & & & & \\ & \ddots & & & & \\ & & \lambda^{(r)} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & 0 & & & & 0 \end{bmatrix}, \quad \lambda^{(1)} < \lambda^{(2)} < \dots < \lambda^{(r)} \leq 0$$

$$\Lambda^+ = \begin{bmatrix} 0 & & & & & \\ & \ddots & & & & \\ & & 0 & & & \\ & & & \lambda^{(r+1)} & & \\ & & & & \ddots & \\ & 0 & & & & \lambda^{(m)} \end{bmatrix}, \quad 0 < \lambda^{(r+1)} < \dots < \lambda^{(m)} \quad [6.100]$$

Equation [6.99] is discretized as follows:

$$\Lambda \frac{\partial W}{\partial x} \approx \Lambda_{i+1/2}^- \left(\frac{\partial W}{\partial x} \right)_{i+1/2}^{n+1/2} + \Lambda_{i-1/2}^+ \left(\frac{\partial W}{\partial x} \right)_{i-1/2}^{n+1/2} \quad [6.101]$$

where the terms $(\partial W / \partial x)_{i-1/2}^{n+1/2}$ and $(\partial W / \partial x)_{i+1/2}^{n+1/2}$ are estimated over the intervals $[i-1, i]$ and $[i, i+1]$ respectively. Multiplying equation [6.101] by the matrix K , introducing the product $K^{-1} K = I$ leads to:

$$\begin{aligned} K \Lambda K^{-1} K \frac{\partial W}{\partial x} &\approx K \Lambda_{i+1/2}^- K^{-1} K \left(\frac{\partial W}{\partial x} \right)_{i+1/2}^{n+1/2} \\ &+ K \Lambda_{i-1/2}^+ K^{-1} K \left(\frac{\partial W}{\partial x} \right)_{i-1/2}^{n+1/2} \end{aligned} \quad [6.102]$$

By definition (see equations [2.16] and [2.21]):

$$\left. \begin{array}{l} K \Lambda K^{-1} = A \\ K dW = dU \end{array} \right\} \quad [6.103]$$

Substituting equations [6.103] into equation [6.102] leads to:

$$A \frac{\partial U}{\partial x} \approx K \Lambda_{i+1/2}^- K^{-1} \left(\frac{\partial U}{\partial x} \right)_{i+1/2}^{n+1/2} + K \Lambda_{i-1/2}^+ K^{-1} \left(\frac{\partial U}{\partial x} \right)_{i-1/2}^{n+1/2} \quad [6.104]$$

Comparing equation [6.104] and [6.97] yields the following expressions for $A_{i-1/2}^+$ and $A_{i+1/2}^-$:

$$\left. \begin{array}{l} A_{i+1/2}^- \approx K \Lambda_{i+1/2}^- K^{-1} \\ A_{i-1/2}^+ \approx K \Lambda_{i-1/2}^+ K^{-1} \end{array} \right\} \quad [6.105]$$

6.7.2. Application to classical schemes

6.7.2.1. Explicit upwind scheme for the water hammer equations

The conservation form of the explicit upwind scheme uses the following approximations for the space derivatives:

$$\left. \begin{array}{l} \left(\frac{\partial U}{\partial x} \right)_{i-1/2}^{n+1/2} \approx \frac{U_i^n - U_{i-1}^n}{\Delta x_i} \\ \left(\frac{\partial U}{\partial x} \right)_{i+1/2}^{n+1/2} \approx \frac{U_{i+1}^n - U_i^n}{\Delta x_i} \end{array} \right\} \quad [6.106]$$

Substituting equations [6.96], [6.104] and [6.106] into equation [6.95] gives:

$$\frac{U_i^{n+1} - U_i^n}{\Delta t} + A_{i-1/2}^+ \frac{U_{i-1}^n - U_i^n}{\Delta x_i} + A_{i+1/2}^- \frac{U_i^n - U_{i+1}^n}{\Delta x_i} = 0 \quad [6.107]$$

which leads to:

$$U_i^{n+1} = U_i^n + \frac{A_{i-1/2}^+ \Delta t}{\Delta x_i} (U_{i-1}^n - U_i^n) + \frac{A_{i+1/2}^- \Delta t}{\Delta x_i} (U_i^n - U_{i+1}^n) \quad [6.108]$$

The expressions of the matrices A , K and K^{-1} derived in section 2.4.3 are recalled hereafter:

$$A = \begin{bmatrix} 0 & 1 \\ c^2 & 0 \end{bmatrix}, \quad K = \begin{bmatrix} 1 & 1 \\ -c & c \end{bmatrix}, \quad K^{-1} = \frac{1}{2c} \begin{bmatrix} c & -1 \\ c & 1 \end{bmatrix} \quad [6.109]$$

The matrices Λ , Λ^- and Λ^+ are:

$$\Lambda = \begin{bmatrix} -c & 0 \\ 0 & c \end{bmatrix}, \quad \Lambda^- = \begin{bmatrix} -c & 0 \\ 0 & 0 \end{bmatrix}, \quad \Lambda^+ = \begin{bmatrix} 0 & 0 \\ 0 & c \end{bmatrix} \quad [6.110]$$

Hence the matrices A^- and A^+ :

$$A^- = \frac{1}{2} \begin{bmatrix} -c & 1 \\ c^2 & -c \end{bmatrix}, \quad A^+ = \frac{1}{2} \begin{bmatrix} c & 1 \\ c^2 & c \end{bmatrix} \quad [6.111]$$

The variable vector U is defined as:

$$U = \begin{bmatrix} \rho A \\ \rho Q \end{bmatrix} \quad [6.112]$$

Substituting equations [6.111–112] into equation [6.108] leads to the following scheme:

$$\left. \begin{aligned} (\rho A)_i^{n+1} &= (\rho A)_i^n + \frac{c \Delta t}{2 \Delta x_i} \left[(\rho A)_{i-1}^n - 2(\rho A)_i^n + (\rho A)_{i+1}^n \right] \\ &\quad + \frac{\Delta t}{2 \Delta x_i} \left[(\rho Q)_{i-1}^n - (\rho Q)_{i+1}^n \right] \\ (\rho Q)_i^{n+1} &= (\rho Q)_i^n + \frac{c^2 \Delta t}{2 \Delta x_i} \left[(\rho A)_{i-1}^n - (\rho A)_{i+1}^n \right] \\ &\quad + \frac{c \Delta t}{2 \Delta x_i} \left[(\rho Q)_{i-1}^n - 2(\rho Q)_i^n + (\rho Q)_{i+1}^n \right] \end{aligned} \right\} \quad [6.113]$$

The linear dependence between the mass per unit length and the pressure and the approximation [2.75] are recalled:

$$\left. \begin{aligned} d(\rho A) &\approx \frac{A}{c^2} dp \\ d(\rho u A) &\approx \rho dQ \end{aligned} \right\} \quad [6.114]$$

Substituting approximations [6.114] into equations [6.113] leads to:

$$\left. \begin{aligned} p_i^{n+1} &= \frac{\text{Cr}}{2} (p_{i-1}^{n+1} + p_{i+1}^{n+1}) + (1 - \text{Cr}) p_i^n + \text{Cr} \frac{\rho c}{2A} (Q_{i-1}^{n+1} - Q_{i+1}^{n+1}) \\ Q_i^{n+1} &= \frac{\text{Cr}}{2} (Q_{i-1}^{n+1} + Q_{i+1}^{n+1}) + (1 - \text{Cr}) Q_i^n + \text{Cr} \frac{A}{2\rho c} (p_{i-1}^{n+1} - p_{i+1}^{n+1}) \end{aligned} \right\} \quad [6.115]$$

where the Courant number is defined as in equation [6.35], recalled hereafter:

$$\text{Cr} = \frac{c\Delta t}{\Delta x_i}$$

Note that equations [6.115] and [6.34] are equivalent when the grid is regular.

6.7.2.2. TVD scheme for the water hammer equations

The general expression for a scalar TVD scheme is provided in section 6.6.2. The generalization of the scheme to vector variables requires the evaluation of the derivatives for positive and negative wave speeds. Remember that TVD schemes can be recast in the following form

$$U_i^{n+1} = \begin{cases} U_i^n - \lambda \Delta t \left(\frac{\partial U}{\partial x} \right)_{i-1/2} & \text{if } \lambda \geq 0 \\ U_i^n - \lambda \Delta t \left(\frac{\partial U}{\partial x} \right)_{i+1/2} & \text{if } \lambda \leq 0 \end{cases} \quad [6.116]$$

For the sake of clarity, the time index is omitted in the derivatives. The derivatives may be estimated using an explicit formulation (in which case the time level n is used), or an implicit approach (in which case the superscript $n+1$ should be used), or any intermediate formulation. Comparing equations [6.116] and [6.86], using the definition $\text{Cr} = \lambda \Delta t / \Delta x$ leads to:

$$\left. \begin{aligned} \left(\frac{\partial U}{\partial x} \right)_{i-1/2} &\approx \frac{U_i - U_{i-1}}{\Delta x} \\ &\quad - \frac{|\text{Cr}| - 1}{2\Delta x} [(U_{i+1} - U_i)\phi_{i+1/2}^+ - (U_i - U_{i-1})\phi_{i-1/2}^+] \\ \left(\frac{\partial U}{\partial x} \right)_{i+1/2} &\approx \frac{U_{i+1} - U_i}{\Delta x} \\ &\quad + \frac{|\text{Cr}| - 1}{2\Delta x} [(U_{i+1} - U_i)\phi_{i+1/2}^- - (U_i - U_{i-1})\phi_{i-1/2}^-] \end{aligned} \right\} [6.117]$$

where the limiters used at the interfaces $i - 1/2$ and $i + 1/2$ depend on the sign of the wave propagation speed. The limiters ϕ^- and ϕ^+ are defined for the waves propagating in the direction of negative and positive x respectively:

$$\left. \begin{aligned} \phi_{i+1/2}^- &= \phi(\theta_{i+1/2}^-) \\ \phi_{i+1/2}^+ &= \phi(\theta_{i+1/2}^+) \end{aligned} \right\} [6.118]$$

where the monotony indicators are defined from equation [6.87] as:

$$\left. \begin{aligned} \theta_{i+1/2}^+ &= \frac{U_i^n - U_{i-1}^n}{U_{i+1}^n - U_i^n} \\ \theta_{i+1/2}^- &= \frac{U_{i+2}^n - U_{i+1}^n}{U_{i+1}^n - U_i^n} \end{aligned} \right\} [6.119]$$

Equation [6.117] is generalized to hyperbolic systems as follows:

1) In the particular case of the water hammer equations, both waves propagate in opposite directions, the absolute values of both speeds are identical. Therefore, the absolute value of the Courant number attached to both waves is the same. The same value of Cr is applied identically to all the components U_p of U in equation [6.117].

2) Several waves may propagate in the same direction in the general case. Each wave has a specific Courant number. The following options are available:

2.1) Apply the wave speed, the absolute value of which is the smallest. This option minimizes the anti-diffusion.

2.2) Apply equations [6.117] to each of the Riemann invariants individually and apply the following estimate for the derivative of U :

$$\left(\frac{\partial U}{\partial x} \right)_{i \pm 1/2} = K_{i \pm 1/2} \left(\frac{\partial W}{\partial x} \right)_{i \pm 1/2} [6.120]$$

Option 1) is used in the particular case of the water hammer equations. The limiter is computed for each of the components of the vector \mathbf{U} . The limiter function is generalized into a diagonal matrix and equation [6.117] is generalized into:

$$\left. \begin{aligned} \left(\frac{\partial \mathbf{U}}{\partial x} \right)_{i-1/2} &\approx \frac{\mathbf{U}_i - \mathbf{U}_{i-1}}{\Delta x} \\ &\quad - \frac{|\mathbf{C}\mathbf{r}| - 1}{2\Delta x} [\Phi_{i+1/2}^+(\mathbf{U}_{i+1} - \mathbf{U}_i) - \Phi_{i-1/2}^+(\mathbf{U}_i - \mathbf{U}_{i-1})] \\ \left(\frac{\partial \mathbf{U}}{\partial x} \right)_{i+1/2} &\approx \frac{\mathbf{U}_{i+1} - \mathbf{U}_i}{\Delta x} \\ &\quad + \frac{|\mathbf{C}\mathbf{r}| - 1}{2\Delta x} [\Phi_{i+1/2}^-(\mathbf{U}_{i+1} - \mathbf{U}_i) - \Phi_{i-1/2}^-(\mathbf{U}_i - \mathbf{U}_{i-1})] \end{aligned} \right\} [6.121]$$

where the coefficients of the limiter matrix are given by:

$$(\Phi_{p,q})_{i+1/2}^\pm = \begin{cases} \phi[(\theta_p)_{i+1/2}^\pm] & \text{if } p = q \\ 0 & \text{if } p \neq q \end{cases} [6.122]$$

where θ_p is the monotony indicator for the p th component of \mathbf{U} . The superscript + or – indicates that two monotony indicators must be defined: one in the direction of negative waves, another in the direction of positive waves.

6.7.2.3. Computational example

Discretization [6.121] is applied to the simulation of the instantaneous failure of a valve in a pipe. The parameters of the test case are given in Table 6.9.

Symbol	Meaning	Value
A	Cross-sectional area of the pipe	1 m^2
c	Sound speed	$1,000 \text{ m/s}$
L	Length of the pipe	500 m
p_0	Initial pressure	5.10^5 Pa for $x < 250 \text{ m}$, 10^5 Pa otherwise
T	Simulated time	0.15 s
Δt	Computational time step	0.005 s
Δx	Cell width	10 m
ρ	Density of water	$1,000 \text{ kg/m}^3$

Table 6.9. Sudden failure of a valve. Physical and numerical parameters

The cell width and the computational time step are chosen such that the Courant number is 0.5, a value for which the effect of profile limiting (when applied) is maximum. The pressure profiles computed at $t = 0.15$ s are represented in Figure 6.26.

Note that the monotone version of the Lax-Wendroff scheme yields an asymmetry in the fronts, which the other limiters do not do. The Superbee limiter remains the most compressive, while the MC and Van Leer limiters yield comparable results.

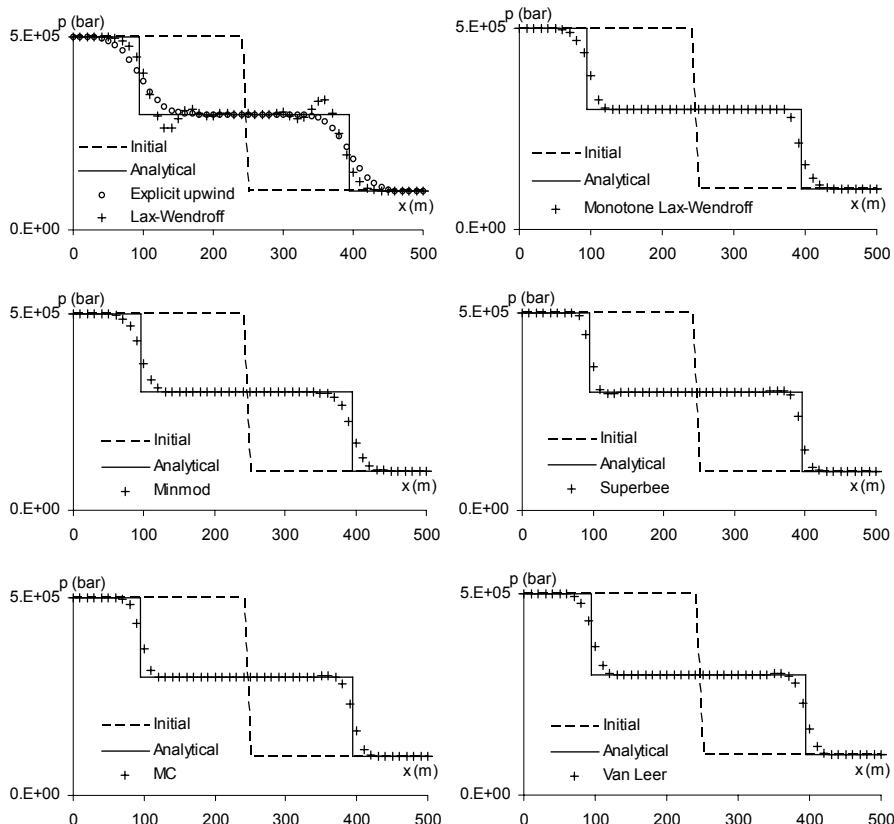


Figure 6.26. Sudden failure of a valve. Pressure profiles computed at $t = 0.15$ s by the upwind scheme, the Lax-Wendroff scheme, and TVD schemes with various limiters

6.8. Conservative discretizations: Roe's matrix

6.8.1. Rationale and principle of the approach

A number of techniques presented in the previous sections are based on the discretization of the equations in non-conservation form. This is the case of the Preissmann scheme seen in section 6.4 with the linearization technique [6.55] presented in section 6.4.2. This is also the case of the flux splitting technique seen in section 6.7. Both techniques require that the Jacobian matrix A of the flux F with respect to the conserved variable U be estimated. The computational examples provided in section 6.2.3 show however that the non-conservation form of the equations may give erroneous results when discontinuities appear in the solution. This is because the propagation speed of shocks is extremely sensitive to the formulation used to estimate the wave speed in the non-conservation form. If a wrong estimate is used, conservation may be violated. The approach presented here allows conservation to be preserved even though the equation is not solved in conservation form.

Consider first the conservation form [2.2] without source term. A balance over a control volume centered around the point i between the time level n and $n + 1$ yields:

$$U_i^{n+1} = U_i^n + \frac{\Delta t}{\Delta x_i} (F_{i-1/2}^{n+1/2} - F_{i+1/2}^{n+1/2}) \quad [6.123]$$

The linearized Preissmann scheme and flux splitting techniques use discretizations of the non-conservation form that can be written as:

$$U_i^{n+1} = U_i^n + \frac{\Delta t}{\Delta x_i} \left[A_{i-1/2}^{n+1/2} \Delta U_{i-1/2}^{n+1/2} - A_{i+1/2}^{n+1/2} \Delta U_{i+1/2}^{n+1/2} \right] \quad [6.124]$$

where ΔU denotes the variation in U between two adjacent computational points:

$$\Delta U_{i+1/2}^{n+1/2} = U_{i+1}^{n+1/2} - U_i^{n+1/2} \quad [6.125]$$

The superscript $n + 1/2$ indicates that the estimate is carried out between the time levels n and $n + 1$. It could be an explicit estimate, a fully implicit estimate or a semi-implicit estimate. Roe [ROE 81] provides a set of necessary conditions that should be satisfied by the estimate of the matrix A for conservation to be guaranteed:

$$\left. \begin{array}{l} A_{i-1/2}^{n+1/2} = A_{i-1/2}^{n+1/2}(U_{i-1}^{n+1/2}, U_i^{n+1/2}) \\ A_{i-1/2}^{n+1/2}(U, U) = \partial F / \partial U(U) \\ A_{i-1/2}^{n+1/2} \Delta U_{i-1/2}^{n+1/2} = F_i^{n+1/2} - F_{i-1}^{n+1/2} \\ \exists(K, \Lambda) \neq (0,0), \quad \Lambda = K^{-1} A_{i-1/2}^{n+1/2} K \end{array} \right\} [6.126]$$

The first and second conditions [6.126] express consistency. The third condition enforces conservation. The last condition states the hyperbolic character of the system of conservation laws. Substituting the third equation [6.126] into equation [6.124] leads to equation [6.123] and conservation is guaranteed.

6.8.2. Expression of Roe's matrix

6.8.2.1. Roe's method

In his publication [ROE 81], Roe provides a general method for the derivation of the matrix A. The method is based on the following reasoning.

The expression of A is a function of the values U_L and U_R of U on the left- and right-hand sides of the discontinuity. Conservation is guaranteed if:

$$A(U_L - U_R) = F_L - F_R \quad [6.127]$$

A “parameter vector” V is introduced, such that:

$$\left. \begin{array}{l} U_L - U_R = B(V_L - V_R) \\ F_L - F_R = C(V_L - V_R) \end{array} \right\} [6.128]$$

where B and C are matrices, the expression of which is to be determined from equation [6.128]. Comparing equations [6.127] and [6.128] leads to the following expression for A:

$$A = CB^{-1} \quad [6.129]$$

Application example: Roe [ROE 81] proposed the following parameter vector for the Euler equations:

$$V = \begin{bmatrix} \rho^{1/2} \\ \rho^{1/2} u \\ \rho^{1/2} H \end{bmatrix} \quad [6.130]$$

where the enthalpy H is defined as:

$$H = (e + p)\rho \quad [6.131]$$

The following expression is obtained for the matrix B:

$$B = \begin{bmatrix} 2\rho^{1/2} & 0 & 0 \\ \rho^{1/2}u & \rho^{1/2} & 0 \\ \rho^{1/2}H/\gamma & (\gamma-1)\rho^{1/2}u/\gamma & \rho^{1/2}/\gamma \end{bmatrix} \quad [6.132]$$

and the following expression is obtained for C:

$$C = \begin{bmatrix} \rho^{1/2}u & \rho^{1/2} & 0 \\ (\gamma-1)\rho^{1/2}H/\gamma & (\gamma+1)\rho^{1/2}u/\gamma & (\gamma-1)\rho^{1/2}/\gamma \\ 0 & \rho^{1/2}H & \rho^{1/2}u \end{bmatrix} \quad [6.133]$$

Note that the eigenvalues and eigenvectors of A may be obtained directly from the expressions of B and C without using equation [6.129]. By definition the eigenvalues of A verify:

$$|A - \lambda I| = 0 \quad [6.134]$$

Consequently:

$$|AB - \lambda B| = 0 \quad [6.135]$$

Using equation [6.129], condition [6.135] becomes:

$$|C - \lambda B| = 0 \quad [6.136]$$

Substituting equations [6.132–133] into equation [6.129] leads to:

$$\left. \begin{array}{l} \rho^{1/2} = \frac{\rho_L^{1/2} + \rho_R^{1/2}}{2} \\ u = \frac{\rho_L^{1/2}u_L + \rho_R^{1/2}u_R}{\rho_L^{1/2} + \rho_R^{1/2}} \\ H = \frac{\rho_L^{1/2}H_L + \rho_R^{1/2}H_R}{\rho_L^{1/2} + \rho_R^{1/2}} \end{array} \right\} \quad [6.137]$$

6.8.2.2. Expression in the base of eigenvectors

The expression of A may also be obtained by writing the variations ΔU and ΔF as linear combinations of the eigenvectors of A. The so-called wave strengths a_p ($p = 1, \dots, m$) are introduced:

$$\left. \begin{aligned} \Delta U &= \sum_{p=1}^m \alpha_p K^{(p)} \\ \Delta F &= \sum_{p=1}^m \alpha_p \lambda^{(p)} K^{(p)} \end{aligned} \right\} [6.138]$$

where the eigenvectors $K^{(p)}$ and the wave speeds $\lambda^{(p)}$ are estimated from a weighted average between U_L and U_R . Finding the expressions of the wave strengths allows the relationship between ΔU and ΔF to be determined.

Application example. Consider the Saint Venant equations. U, F and K are defined as:

$$U = \begin{bmatrix} A \\ Q \end{bmatrix}, F = \begin{bmatrix} Q \\ uQ + P / \rho \end{bmatrix}, K^{(1)} = \begin{bmatrix} 1 \\ u - c \end{bmatrix}, K^{(2)} = \begin{bmatrix} 1 \\ u + c \end{bmatrix} [6.139]$$

where A is the channel cross-sectional area, c is the speed of the waves in still water, P is the pressure force exerted on the channel cross-section, Q is the liquid discharge, u is the flow velocity and ρ is the density of water. The first equation [6.138] yields the following system:

$$\left. \begin{aligned} \alpha_1 + \alpha_2 &= \Delta A \\ (u - c)\alpha_1 + (u + c)\alpha_2 &= \Delta Q \end{aligned} \right\} [6.140]$$

Solving equations [6.140] for the wave strengths leads to:

$$\left. \begin{aligned} \alpha_1 &= \frac{u + c}{2c} \Delta A - \frac{\Delta Q}{2c} \\ \alpha_2 &= -\frac{u - c}{2c} \Delta A + \frac{\Delta Q}{2c} \end{aligned} \right\} [6.141]$$

The second equation [6.138] gives:

$$\left. \begin{aligned} (u - c)\alpha_1 + (u + c)\alpha_2 &= \Delta Q \\ (u - c)^2 \alpha_1 + (u + c)^2 \alpha_2 &= \Delta(uQ + P / \rho) \end{aligned} \right\} [6.142]$$

Note that the first equation [6.142] is equivalent to the second equation [6.140]. Solving equations [6.142] for c and u leads to:

$$\left. \begin{aligned} u &= \frac{A_L^{1/2}u_L + A_R^{1/2}u_R}{A_L^{1/2} + A_R^{1/2}} \\ c^2 &= \frac{g}{\rho} \frac{P_R - P_L}{A_R - A_L} \end{aligned} \right\} \quad [6.143]$$

6.9. Multidimensional problems

6.9.1. Explicit alternate directions

Most classical numerical schemes were originally developed in one dimension of space. As well as for historical reasons, the numerical properties of one-dimensional schemes (consistency, solution stability, etc.) are easier to study than those of multidimensional schemes. Alternate directions, sometimes referred to as dimension splitting or time splitting techniques, allow multidimensional problems to be solved using one-dimensional techniques. Alternate directions techniques are analyzed in [STR 68] and [GOU 77]. Such techniques are mainly used on Cartesian grids.

Consider the following multidimensional problem:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} + \frac{\partial G}{\partial y} + \frac{\partial H}{\partial z} = S \quad [6.144]$$

where U is the conserved variable, F , G and H are the flux vectors in the x -, y - and z -direction respectively, and S is the source term. The solution U^n at the time level n is assumed to be known at all points of the computational grid. The purpose is to compute the solution U^{n+1} at the time level $n+1$. The solution may be approximated as:

$$U^{n+1} = L_{\Delta t}^{(S)} L_{\Delta t}^{(z)} L_{\Delta t}^{(y)} L_{\Delta t}^{(x)} U^n \quad [6.145]$$

where $L_{\Delta t}^{(x)}$ is the numerical scheme (also referred to as “numerical operator”) that solves the conservation part of equation [6.144] in the x -direction over the time step Δt . The conservation part of equation [6.144] in the x -direction is:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} = 0 \quad [6.146]$$

and $L_{\Delta t}^{(x)}$ is the numerical scheme that solves equation [6.146] numerically over the time step Δt via a formula that can be written in the form:

$$U_i^{n+1} = L_{\Delta t}^{(x)} U_i^n \quad [6.147]$$

By definition of the time derivative, the values of U at the time levels n and $n + 1$ are related by:

$$\begin{aligned} U_i^{n+1} &\approx U_i^n + \Delta t \left(\frac{\partial U}{\partial t} \right)_i^{n+1/2} \\ &\approx U_i^n - \Delta t \frac{\partial F}{\partial x}(U_i^n) \end{aligned} \quad [6.148]$$

where $\partial F / \partial x(U_i^n)$ is the discretized form of $\partial F / \partial x$ at the computational point i . Comparing equations [6.147] and [6.148] leads to the following definition for the operator $L_{\Delta t}^{(x)}$:

$$L_{\Delta t}^{(x)} = I - \Delta t \frac{\partial F}{\partial x} \quad [6.149]$$

where I is the identity matrix. The operators $L_{\Delta t}^{(y)}$ and $L_{\Delta t}^{(z)}$ solve the conservation part of the equation in the y - and z -direction respectively and the operator $L_{\Delta t}^{(S)}$ accounts for the contribution of the source term by solving:

$$\frac{\partial U}{\partial t} = S \quad [6.150]$$

Equation [6.145] describes the following sequence. The conservation part of equation [6.144] is solved in the x -direction. The result is used as a starting point for the solution of the conservation part in the y -direction. The result of this computation is used as an initial state to solve the conservation part of the equation in the z -direction. The result of this step is used as an initial condition in the solution of equation [6.150].

The numerical solution is stable if each of the operators in sequence [6.145] is stable. The advection operators in the x -, y - and z -direction give stable solutions if the absolute value of the Courant number in the corresponding direction of space is smaller than one:

$$\max(|Cr_x|, |Cr_y|, |Cr_z|) \leq 1 \quad [6.151]$$

where Cr_x , Cr_y and Cr_z are respectively the Courant numbers in the x -, y - and z -direction. Sequence [6.145] is first-order accurate with respect to time.

When second- and higher-order schemes are used, the first-order time stepping sequence [6.145] may not be accurate enough. The accuracy of the solution may be improved by using the following sequence for the conservation part, as suggested in [STR 68]:

$$U^{n+1} = L_{\Delta t / 2}^{(z)} L_{\Delta t / 2}^{(y)} L_{\Delta t}^{(x)} L_{\Delta t / 2}^{(y)} L_{\Delta t / 2}^{(z)} U^n \quad [6.152]$$

where the subscripts $\Delta t / 2$ indicate that the operators are applied over half a time step. Equation [6.152] can be generalized to account for the influence of the source term as:

$$U^{n+1} = L_{\Delta t / 2}^{(S)} L_{\Delta t / 2}^{(z)} L_{\Delta t / 2}^{(y)} L_{\Delta t}^{(x)} L_{\Delta t / 2}^{(y)} L_{\Delta t / 2}^{(z)} L_{\Delta t / 2}^{(S)} U^n \quad [6.153]$$

Since the operators $L^{(y)}$ and $L^{(z)}$ are used over half a time step, the stability criterion becomes:

$$\max(|Cr_x|, |Cr_y / 2|, |Cr_z / 2|) \leq 1 \quad [6.154]$$

Sequence [6.153] allows spurious effects such as solution anisotropy to be reduced to a large extent compared to the first-order approach [6.145]. However, it is more time-consuming.

6.9.2. The ADI method

The Alternate Directions Implicit (ADI) method is iterative. It uses implicit one-dimensional schemes that allow large time steps to be used without making the numerical solution unstable. The schemes being one-dimensional, the linear algebraic systems to be solved remain diagonal and can be solved using fast system inversion techniques. Any implicit discretization of equation [6.144] can be written in the form:

$$U^{n+1} + \frac{\partial F(U^{n+1})}{\partial x} + \frac{\partial G(U^{n+1})}{\partial y} + \frac{\partial H(U^{n+1})}{\partial z} - S(U^{n+1}) = R(U^n) \quad [6.155]$$

where $R(U^n)$ contains all the terms that are functions of the known solution at the time level n . Equation [6.155] is solved as follows:

1) The equation is solved in the x -direction, all the remaining terms in the equation being assumed known. In the first iteration, the derivatives of G and H must be “guessed”. The simplest method is to use the value at the time level n . The following equation is solved:

$$U^{n+1,x} + \frac{\partial F(U^{n+1,x})}{\partial x} = -\frac{\partial G(U^n)}{\partial y} - \frac{\partial H(U^n)}{\partial z} + S(U^n) + R(U^n) \quad [6.156]$$

where $U^{n+1,x}$ is the solution of equation [6.156] over the time step Δt .

2) The solution $U^{n+1,x}$ is used as an initial condition for the equation in the y -direction:

$$U^{n+1,y} + \frac{\partial G(U^{n+1,y})}{\partial y} = -\frac{\partial F(U^{n+1,x})}{\partial x} - \frac{\partial H(U^{n+1,x})}{\partial z} + S(U^{n+1,x}) + R(U^{n+1,x}) \quad [6.157]$$

where $U^{n+1,y}$ is the solution of equation [6.157] over the time step Δt .

3) The solution $U^{n+1,y}$ is used as an initial condition for the equation in the z -direction:

$$U^{n+1,z} + \frac{\partial G(U^{n+1,z})}{\partial y} = -\frac{\partial F(U^{n+1,y})}{\partial x} - \frac{\partial H(U^{n+1,y})}{\partial z} + S(U^{n+1,y}) + R(U^{n+1,y}) \quad [6.158]$$

where $U^{n+1,z}$ is the solution of equation [6.158] over the time step Δt .

4) The solution $U^{n+1,z}$ is used as an initial condition for the contribution of the source term:

$$U^{n+1,S} - S(U^{n+1,S}) = -\frac{\partial G(U^{n+1,z})}{\partial y} - \frac{\partial F(U^{n+1,z})}{\partial x} - \frac{\partial H(U^{n+1,z})}{\partial z} + R(U^{n+1,z}) \quad [6.159]$$

where $U^{n+1,S}$ is the solution of equation [6.159] over the time step Δt .

5) The solution $U^{n+1,S}$ is used as an initial condition to solve the equation in the x -direction:

$$U^{n+1,x} + \frac{\partial F(U^{n+1,x})}{\partial x} = -\frac{\partial G(U^{n+1,S})}{\partial y} - \frac{\partial H(U^{n+1,S})}{\partial z} + S(U^{n+1,S}) + R(U^{n+1,S}) \quad [6.160]$$

Steps 2–5, that form a single iteration, must be repeated until convergence is achieved, that is:

- the value of U at a given step (e.g. step 5) between two successive iterations should not differ by more than a given threshold value ε specified by the modeler;
- the values of U between two successive steps within the same iteration should not differ by more than the threshold value ε .

In many industrial implementations of the ADI method, only the first condition is checked. In other applications, the first condition is not checked and the second condition is checked only between two steps (e.g. between steps 2 and 3) instead of the complete sequence. When this is the case, the solution may be abnormally sensitive to the orientation chosen for the main axes of the grid. The ADI method is known to converge slowly when applied to nonlinear systems. For this reason, most industrial implementations of the method use a maximum permissible number of iterations, after which the sequence 2)–5) will be stopped regardless of the convergence of the iterative process. The maximum permissible number of iterations may be left to the modeler's choice or hard-programmed in the software. The fact that the user is not informed of or has no control over the parameters that influence the accuracy of the method may lead to erroneous computational results, thus destroying the predictive power of the simulation.

6.9.3. Multidimensional schemes

Multidimensional schemes solve multidimensional problems by treating all the spatial derivatives within one single step. When a two-dimensional problem is to be solved, the operators $L^{(x)}$ and $L^{(y)}$ introduced in the previous section are applied as follows:

$$U^{n+1} = \left[L_{\Delta t}^{(x)} + L_{\Delta t}^{(y)} - I \right] U^n \quad [6.161]$$

where I is the identity matrix. Equation [6.161] is derived as follows. Extending equation [6.149] to the y -direction gives:

$$L_{\Delta t}^{(y)} = I - \Delta t \frac{\partial G}{\partial y} \quad [6.162]$$

The discretized form of the equation to be solved is:

$$U_i^{n+1} = U_i^n - \Delta t \frac{\partial F}{\partial x}(U_i^n) - \Delta t \frac{\partial G}{\partial y}(U_i^n) \quad [6.163]$$

It is easy to check that substituting equations [6.149] and [6.162] into equation [6.163] leads to equation [6.161]. Extending the reasoning to three dimensions of space leads to the following formula:

$$U^{n+1} = \left[L_{\Delta t}^{(x)} + L_{\Delta t}^{(y)} + L_{\Delta t}^{(z)} - 2I \right] U^n \quad [6.164]$$

Since the effects of the operators in each direction of space are added within a single step, the stability constraint attached to multidimensional schemes is usually more restrictive than for alternate direction techniques. Most two-dimensional explicit schemes are subjected to the following constraint:

$$\max(|Cr_x| + |Cr_y|) \leq 1 \quad [6.165]$$

the three-dimensional version of which is:

$$\max(|Cr_x| + |Cr_y| + |Cr_z|) \leq 1 \quad [6.166]$$

6.10. Summary

6.10.1. *What you should remember*

Finite difference methods are based on the discretization of space and time, which allows partial differential equations to be approximated in the form of differences between the values taken by the solution at predefined time and space coordinates. When the unknown value of the solution at the next time level can be expressed only as a function of the known value at the current time step, the method is said to be explicit. A discretization of the partial differential equation that provides a relationship between several unknowns at the next time level is said to be implicit. Explicit methods are subjected to stability constraints, while implicit methods are not in general.

The performance of a numerical method for hyperbolic conservation laws is conditioned by the Courant number, defined as the ratio of the distance covered by

the wave over a time step to the distance between two adjacent computational points.

Characteristic-based methods, covered in section 6.2, solve the governing equations in characteristic form. The Riemann invariants are interpolated at the feet of the characteristics from their values at the computational points. Linear interpolations lead to first-order, diffusive numerical schemes. Parabolic interpolations lead to second-order, dispersive schemes. Characteristic-based methods may fail to preserve the conservation properties of the solution, especially in the presence of shocks and when the Courant number is not uniform over the computational domain.

Upwind schemes for scalar laws, presented in section 6.3, may be applied to the conservation and non-conservation form of the equations. In such schemes the derivative with respect to space is discretized using the point located upstream of the point at which the time derivative is discretized.

The Preissmann scheme presented in section 6.4 is a conservative scheme. It is used by a number of commercially available software packages for free-surface flow modeling. The degree of consistency of the discretization can be adjusted in time and space via two weighting parameters ψ and θ . The parameters ψ that control the weighting in space is usually set to 1/2. When this is the case, setting $\theta=1/2$ leads to a purely dispersive scheme. Increasing θ induces numerical diffusion.

Centered schemes are dealt with in section 6.5. They are not sensitive to the direction in which the waves propagate. They are less diffusive than upwind schemes. The Crank-Nicholson scheme is an implicit scheme. Explicit centered schemes require the use of third- and higher-order Runge-Kutta time integration algorithms for stability reasons. Their range of stability is wider than that of classical explicit schemes.

TVD schemes for scalar laws are presented in section 6.6. They can be seen as weighted combinations between the upwind scheme and the second-order Lax-Wendroff scheme. The weighting between the two schemes is a function of the local variations in the slope of the variable. The contributions of the gradient of the variable are limited using a so-called limiter, for which many formulations have been proposed in the literature.

Flux splitting techniques are presented in section 6.7. Such techniques allow upwind and TVD schemes to be generalized to hyperbolic systems of conservation laws. They use the non-conservation form of the equations. The flux splitting approach consists of separating the Jacobian matrix of the flux into two matrices,

one that accounts for the waves propagating in the direction of negative x , the other accounting for the waves that propagate in the direction of positive x .

The flux splitting technique allows conservation to be preserved provided that the Jacobian matrix satisfies a number of criteria. Roe's linearization technique, presented in section 6.8, provides guidelines for the derivation of the Jacobian matrix.

Several options are available for the treatment of multidimensional problems (see section 6.9). The alternate direction technique consists of solving the governing equations in each direction of space successively. Multidimensional schemes treat all the directions of space within a single step.

6.10.2. Application exercises

6.10.2.1. Exercise 6.1: finite difference methods for scalar laws

Check the conclusions of Exercises 1.1 to 1.5 using finite difference methods. The following methods are advised:

- a characteristic-based method,
- an upwind scheme (conservative version),
- Preissmann's scheme,
- a TVD scheme.

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

6.10.2.2. Exercise 6.2: finite difference methods for hyperbolic systems

Check the conclusions derived in Exercises 2.2 and 2.5 by solving the equations numerically. The first-order MOC is advised for the sake of simplicity.

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

6.10.2.3. Exercise 6.3: finite difference methods for hyperbolic systems

Implement the numerical schemes used in the examples presented in sections 6.2.3, 6.2.4, 6.4.3, 6.6.5 and 6.7.2.3.

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

Chapter 7

Finite Volume Methods for Hyperbolic Systems

7.1. Principle

7.1.1. One-dimensional conservation laws

Finite volume methods solve the conservation form of conservation laws. Scalar hyperbolic laws are expressed in conservation form as in equation [1.1], recalled here:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} = S$$

As shown in section 1.1.2, equation [1.1] is derived from a balance over a control volume $[x_0, x_0 + \delta x] \times [t_0, t_0 + \delta t]$, the size of which is made infinitesimal. The balance equation [1.12] is recalled:

$$\begin{aligned} \int_{x_0}^{x_0 + \delta x} [U(x, t_0 + \delta t) - U(x, t_0)] dx &= \int_{t_0}^{t_0 + \delta t} [F(x_0, t) - F(x_0 + \delta x, t)] dt \\ &+ \int_{t_0}^{t_0 + \delta t} \int_{x_0}^{x_0 + \delta x} S(x, t) dx dt \end{aligned}$$

As shown in section 1.1.2, equation [1.1] is a particular case of equation [1.12]. In contrast with equation [1.1], equation [1.12] is not based on the assumption that

the solution is continuous and differentiable with respect to time and space. In finite volume methods, space is discretized into volumes, also called computational cells, over which balance equation [1.12] is solved (Figure 7.1).

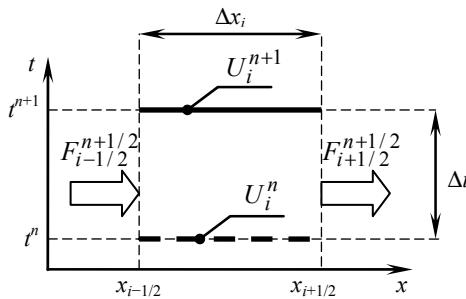


Figure 7.1. Discretization of space and time for a one-dimensional finite volume method

Applying balance equation [1.12] over the control volume sketched in Figure 7.1 leads to the following equation:

$$(U_i^{n+1} - U_i^n) \Delta x_i = (F_{i-1/2}^{n+1/2} - F_{i+1/2}^{n+1/2}) \Delta t + \Delta t \Delta x_i S_i^{n+1/2} \quad [7.1]$$

where Δx_i is the width of the cell i , Δt is the computational time step, U_i^n is the average value of U over the cell i at the time level n , $F_{i+1/2}^{n+1/2}$ is the average value of the flux F at the interface $i + 1/2$ between the cells i and $i + 1$ between the time levels n and $n + 1$ and $S_i^{n+1/2}$ is the average value of the source term over the cell i between the time levels n and $n + 1$. Dividing by Δx_i yields:

$$U_i^{n+1} = U_i^n + \frac{\Delta t}{\Delta x_i} (F_{i-1/2}^{n+1/2} - F_{i+1/2}^{n+1/2}) + \Delta t S_i^{n+1/2} \quad [7.2]$$

The average value of U over the cell i at the next time level can be determined provided that the fluxes at the interfaces between the cells and the source term can be estimated between the time levels n and $n + 1$. How to estimate the fluxes is dealt with in the next sections. If F and S are estimated using only the known solution at the time level n , the method is said to be explicit. If the estimates of F and S are functions of the unknown solution at the time level $n + 1$, the method is said to be implicit.

Hyperbolic systems of conservation laws are written in conservation form as in equation [2.2], recalled here (see Chapter 2):

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} = \mathbf{S}$$

Equation [2.2] is discretized by applying the scalar discretization [7.2] to each of the components of \mathbf{U} , \mathbf{F} and \mathbf{S} . The vector form of equation [7.2] is obtained:

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i^n + \frac{\Delta t}{\Delta x_i} (\mathbf{F}_{i-1/2}^{n+1/2} - \mathbf{F}_{i+1/2}^{n+1/2}) + \Delta t \mathbf{S}_i^{n+1/2} \quad [7.3]$$

7.1.2. Multidimensional conservation laws

Only two-dimensional problems are considered in what follows. The extension to three-dimensional problems will not be detailed hereafter. Two-dimensional systems can be written in the conservation form [5.12], recalled here:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} + \frac{\partial \mathbf{G}}{\partial y} = \mathbf{S}$$

where \mathbf{F} and \mathbf{G} are the fluxes in the x - and y -direction respectively. Space is discretized into polygonal cells, usually triangles or quadrangles. The edges of the cells are straight lines (Figure 7.2).

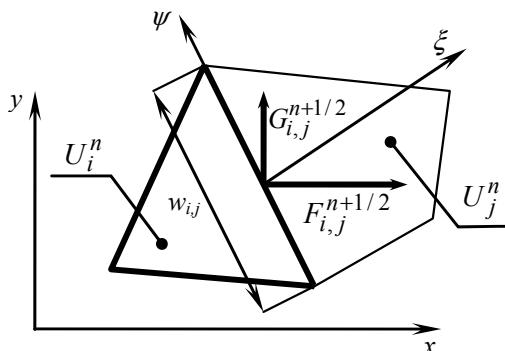


Figure 7.2. Discretization of space for the finite volume solution of a two-dimensional hyperbolic conservation law. Notation for a scalar law

The balance over the cell i can be written as:

$$\begin{aligned} (\mathbf{U}_i^{n+1} - \mathbf{U}_i^n) A_i + \sum_{j \in V(i)} \left[\mathbf{F}_{i,j}^{n+1/2} n_{i,j}^{(x)} + \mathbf{G}_{i,j}^{n+1/2} n_{i,j}^{(y)} \right] w_{i,j} \Delta t \\ = \mathbf{S}_i^{n+1/2} A_i \Delta t \end{aligned} \quad [7.4]$$

where A_i is the area of the cell, $V(i)$ is the set of the neighbor cells of the cell i , $\mathbf{F}_{i,j}^{n+1/2}$ and $\mathbf{G}_{i,j}^{n+1/2}$ are respectively the average values of the fluxes F and G at the interface (i,j) between the cells i and j between the time levels n and $n+1$, $n_{i,j}^{(x)}$ and $n_{i,j}^{(y)}$ are the x - and y -components of the normal unit vector attached to the interface (i,j) , positive from i to j , and $w_{i,j}$ is the width of the interface (i,j) . Note that $n_{i,j}^{(x)}$ and $n_{i,j}^{(y)}$ are respectively the cosine and sine of the angle between the normal unit vector of the interface and the x -axis. Equation [7.4] can be rewritten as:

$$\begin{aligned} \mathbf{U}_i^{n+1} = \mathbf{U}_i^n - \frac{\Delta t}{A_i} \sum_{j \in V(i)} \left[\mathbf{F}_{i,j}^{n+1/2} n_{i,j}^{(x)} + \mathbf{G}_{i,j}^{n+1/2} n_{i,j}^{(y)} \right] w_{i,j} \\ + \mathbf{S}_i^{n+1/2} \Delta t \end{aligned} \quad [7.5]$$

The fluxes F and G must be estimated at each interface (i,j) in the computational domain. The apparent complexity of equation [7.5] can be reduced to some extent by noting that the quantity between brackets in equation [7.5] is nothing other than the flux in the direction normal to the interface (i,j) . Consequently, the method is greatly simplified by solving the projection of the equations onto the normal unit vector. The following algorithm is used in practice:

1) For each interface (i,j) , the variables \mathbf{U}_i^n and \mathbf{U}_j^n are expressed in the local coordinate system (ξ, ψ) attached to the interface. The directions ξ and ψ are normal and tangent to the interface respectively. Scalar variables such as the pressure, water depth, entropy, internal energy, etc. are left unchanged by the transformation, while vector variables such as the velocity or unit discharge vector are transformed using a classical rotation formula:

$$\left. \begin{aligned} u_\xi &= n_{i,j}^{(x)} u + n_{i,j}^{(y)} v \\ u_\psi &= -n_{i,j}^{(y)} u + n_{i,j}^{(x)} v \end{aligned} \right\} \quad [7.6]$$

where u_ξ and u_ψ are respectively the components of the velocity in the ξ - and ψ -direction. Given the orientation of the ξ -axis, the quantities U_i and U_j are often denoted by U_L and U_R respectively. Transformation [7.6] can be written in vector form as:

$$\left. \begin{aligned} U_L &= P^{-1} U_i^n \\ U_R &= P^{-1} U_j^n \end{aligned} \right\} \quad [7.7]$$

where P is the conversion matrix from the global coordinate system (x, y) into the local coordinate system (ξ, ψ) .

2) For each interface (i, j) , compute the flux F_ξ in the direction normal to the interface using the values U_L and U_R . The various existing techniques for flux computation are detailed in the next sections and in Appendix C on Riemann solvers.

3) Convert the flux F_ξ to the global coordinate system (x, y) and apply the balance equation [7.5] that can be rewritten as:

$$U_i^{n+1} = U_i^n - \frac{\Delta t}{A_i} \sum_{j \in V(i)} P F_{\xi i,j}^{n+1/2} w_{i,j} + S_i^{n+1/2} \Delta t \quad [7.8]$$

Note that a number of discretization techniques, such as the “source term upwinding technique” also use the expression of the source term in the local coordinate system attached to the interface. In this case, steps 1) to 3) must also be applied to the source term S .

7.1.3. Application to the two-dimensional shallow water equations

The present section deals with the application of the technique described in section 7.1.2 to the two-dimensional shallow water equations dealt with in section 5.4. The two-dimensional shallow water equations can be written in the conservation form [5.12], recalled here:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} + \frac{\partial \mathbf{G}}{\partial y} = \mathbf{S}$$

by defining U, F, G and S as in equation [5.64], recalled hereafter:

$$\begin{aligned} \mathbf{U} &= \begin{bmatrix} h \\ hu \\ hv \end{bmatrix}, & \mathbf{F} &= \begin{bmatrix} hu \\ hu^2 + gh^2 / 2 \\ huv \end{bmatrix}, \\ \mathbf{G} &= \begin{bmatrix} hv \\ huv \\ hv^2 + gh^2 / 2 \end{bmatrix}, & \mathbf{S} &= \begin{bmatrix} 0 \\ (S_{0,x} - S_{f,x})gh \\ (S_{0,y} - S_{f,y})gh \end{bmatrix} \end{aligned}$$

For the sake of clarity, the discretization of the source term is not dealt with hereafter. Steps 1) to 3) of the algorithm in section 7.1.2 are applied.

1) Expression of U in the local coordinate system. h is left unchanged, while u and v are transformed into u_ξ and u_ψ as in equation [7.6]. \mathbf{U}_L and \mathbf{U}_R are given by:

$$\left. \begin{aligned} \mathbf{U}_L &= \begin{bmatrix} h_i^n \\ (hu_\xi)_i^n \\ (hu_\psi)_i^n \end{bmatrix} = \begin{bmatrix} h_i^n \\ \left[n_{i,j}^{(x)} u_i^n + n_{i,j}^{(y)} v_i^n \right] h_i^n \\ \left[n_{i,j}^{(y)} u_i^n + n_{i,j}^{(x)} v_i^n \right] h_i^n \end{bmatrix} = P^{-1} \mathbf{U}_i^n \\ \mathbf{U}_R &= \begin{bmatrix} h_j^n \\ (hu_\xi)_j^n \\ (hu_\psi)_j^n \end{bmatrix} = \begin{bmatrix} h_j^n \\ \left[n_{i,j}^{(x)} u_j^n + n_{i,j}^{(y)} v_j^n \right] h_j^n \\ \left[-n_{i,j}^{(y)} u_j^n + n_{i,j}^{(x)} v_j^n \right] h_j^n \end{bmatrix} = P^{-1} \mathbf{U}_j^n \end{aligned} \right\} \quad [7.9]$$

The matrix P^{-1} is derived from equation [7.9]:

$$P^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & n_{i,j}^{(x)} & n_{i,j}^{(y)} \\ 0 & -n_{i,j}^{(y)} & n_{i,j}^{(x)} \end{bmatrix} \quad [7.10]$$

hence the expression of P:

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & n_{i,j}^{(x)} & -n_{i,j}^{(y)} \\ 0 & n_{i,j}^{(y)} & n_{i,j}^{(x)} \end{bmatrix} \quad [7.11]$$

2) Solution of the equations in the local coordinate system. Since the interface is a straight line, the problem is locally one-dimensional and the governing equation becomes (remember that the source term is assumed to be zero):

$$\frac{\partial U_\xi}{\partial t} + \frac{\partial F_\xi}{\partial \xi} = 0 \quad [7.12]$$

with:

$$U_\xi = P^{-1}U = \begin{bmatrix} h \\ hu_\xi \\ hu_\psi \end{bmatrix}, \quad F_\xi^{n+1/2} = \begin{bmatrix} h \\ hu_\xi^2 + \frac{gh^2}{2} \\ hu_\xi u_\psi \end{bmatrix}_{i,j}^{n+1/2} \quad [7.13]$$

The flux at the interface is usually computed by solving a Riemann problem. Examples of exact or approximate Riemann solvers can be found in Appendix C.

3) Conversion of the flux F_ξ to the global coordinate system (x, y) and balance as in equation [7.8]:

$$\begin{bmatrix} h \\ hu \\ hv \end{bmatrix}_i^{n+1} = \begin{bmatrix} h \\ hu \\ hv \end{bmatrix}_i^n - \sum_{j \in V(i)} \frac{w_{i,j} \Delta t}{A_i} \begin{bmatrix} 1 & 0 & 0 \\ 0 & n_{i,j}^{(x)} & -n_{i,j}^{(y)} \\ 0 & n_{i,j}^{(y)} & n_{i,j}^{(x)} \end{bmatrix} \begin{bmatrix} h \\ hu_\xi^2 + \frac{gh^2}{2} \\ hu_\xi u_\psi \end{bmatrix}_{i,j}^{n+1/2} \quad [7.14]$$

7.2. Godunov's scheme

7.2.1. Principle

At the time of the publication of Godunov's scheme [GOD 59, GOD 99], the concept of finite volumes had not yet been formalized and Godunov presented his scheme as a finite difference scheme. Twenty years elapsed before the finite volume formalism was introduced by Van Leer in the development of the MUSCL scheme (see section 7.3).

Godunov originally saw his own scheme as a conservative generalization of the CIR scheme (see section 6.2.2). The CIR scheme is based on the characteristic form of the equations. As seen in section 6.2.2, the CIR scheme fails to preserve conservation in the presence of shocks. This is why Godunov tried to solve the conservation form of the equation by estimating the fluxes at the interface between the computational points. In Godunov's scheme, the flux $F_{\xi i,j}^{n+1/2}$ is computed from the solution of a Riemann problem in the local coordinate system attached to the interface (i, j) . The Riemann problem is defined as:

$$\left. \begin{aligned} \frac{\partial U_\xi}{\partial t} + \frac{\partial F_\xi}{\partial \xi} &= 0 \\ U(\xi, t^n) &= \begin{cases} U_L & \text{for } \xi < 0 \\ U_R & \text{for } \xi > 0 \end{cases} \end{aligned} \right\} \quad [7.15]$$

where the interface is located at $\xi = 0$. In Godunov's scheme, the left- and right-states of the Riemann problem are taken from the average values of the variable over the cell on the left- and right-hand side of the interface respectively. The Riemann problem may be solved exactly (see Chapter 4) or approximately (see Appendix C). In the case where the Riemann problem is solved exactly, the value of U and F is known for all ξ and t . However, only the value of the flux at $\xi = 0$ is of practical interest. Approximate Riemann solvers focus on the determination of the flux at the interface. Note that the self-similarity property of the solution of the Riemann problem is of direct interest to the calculation of F because the variable and the flux are independent of time at the location of the initial discontinuity.

Also note that:

- in the one-dimensional case, the cells i and j are aligned along the x -axis, therefore $j = i + 1$ and:

$$\left. \begin{aligned} U_L &= U_i^n \\ U_R &= U_{i+1}^n \end{aligned} \right\} \quad [7.16]$$

in this case the solution of the Riemann problem yields the flux $F_{i+1/2}^{n+1/2}$;

- in multidimensional problems, U_L and U_R must be determined as in equation [7.7].

7.2.2. Application to the scalar advection equation

7.2.2.1. Discretization

This section deals with the application of Godunov's scheme to the linear advection equation [1.39], recalled here:

$$\frac{\partial}{\partial t}(AC) + \frac{\partial}{\partial x}(QC) = 0$$

where A is the cross-sectional area of the channel, C is the concentration of the dissolved substance and Q is the liquid discharge. A and Q are assumed known a priori. They satisfy the continuity equation [1.46], recalled hereafter:

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0$$

As shown in Chapter 1, combining equations [1.39] and [1.46] leads to the non-conservation form [1.48]:

$$\frac{\partial C}{\partial t} + u \frac{\partial C}{\partial x} = 0$$

where $u = Q/A$. The characteristic form [1.50] follows:

$$C = \text{Const} \quad \text{for } \frac{dx}{dt} = u$$

Applying equation [7.2] to equation [1.39] gives:

$$(AC)_i^{n+1} = (AC)_i^n + \frac{\Delta t}{\Delta x_i} \left[(QC)_{i-1/2}^{n+1/2} - (QC)_{i+1/2}^{n+1/2} \right] \quad [7.17]$$

Equation [7.17] can be rewritten as:

$$C_i^{n+1} = \frac{A_i^n}{A_i^{n+1}} C_i^n + \frac{\Delta t}{A_i^{n+1} \Delta x_i} (Q_{i-1/2}^{n+1/2} C_{i-1/2}^{n+1/2} - Q_{i+1/2}^{n+1/2} C_{i+1/2}^{n+1/2}) \quad [7.18]$$

where A and Q are assumed known everywhere in the computational domain at all times. C_i^{n+1} can be computed provided that the values $C_{i-1/2}^{n+1/2}$ and $C_{i+1/2}^{n+1/2}$ at the

interfaces $i - 1/2$ and $i + 1/2$ can be computed. Only the calculation of $C_{i+1/2}^{n+1/2}$ is detailed hereafter, the procedure being identical for the remaining interfaces.

7.2.2.2. Flux calculation at internal interfaces

The concentration $C_{i+1/2}^{n+1/2}$ at the interface $i + 1/2$ is computed from the solution of the following Riemann problem:

$$\left. \begin{aligned} \frac{\partial}{\partial t}(AC) + \frac{\partial}{\partial t}(QC) &= 0 \\ C(x, t^n) &= \begin{cases} C_L = C_i^n & \text{for } x < x_{i+1/2} \\ C_R = C_{i+1}^n & \text{for } x > x_{i+1/2} \end{cases} \end{aligned} \right\} \quad [7.19]$$

The solution of problem [7.19] is described in detail in section 4.2.1. The concentration being invariant along the characteristic lines (see equation [1.50]), the concentration profile is given by:

$$C(x, t > t^n) = \begin{cases} C_L = C_i^n & \text{for } x < x_{i+1/2} + (t - t^n)u \\ C_R = C_{i+1}^n & \text{for } x > x_{i+1/2} + (t - t^n)u \end{cases} \quad [7.20]$$

where u is the average velocity at the interface $i + 1/2$ between the time levels n and $n + 1$ (see section 7.2.2.4 for suggested estimates of u). $C_{i+1/2}^{n+1/2}$ is given by:

$$C_{i+1/2}^{n+1/2} = \begin{cases} C_R = C_{i+1}^n & \text{if } u < 0 \\ C_L = C_i^n & \text{if } u > 0 \end{cases} \quad [7.21]$$

Note that when $u = 0$, the profile does not move. Equation [7.21] leads to indeterminacy. This however is not a problem in practice because the flux is zero when $u = 0$. The final estimate for the flux becomes:

$$Q_{i+1/2}^{n+1/2} C_{i+1/2}^{n+1/2} = \begin{cases} Q_{i+1/2}^{n+1/2} C_{i+1}^n & \text{if } Q_{i+1/2}^{n+1/2} < 0 \\ 0 & \text{if } Q_{i+1/2}^{n+1/2} = 0 \\ Q_{i+1/2}^{n+1/2} C_i^n & \text{if } Q_{i+1/2}^{n+1/2} > 0 \end{cases} \quad [7.22]$$

Equation [7.22] can be condensed into the following expression:

$$Q_{i+1/2}^{n+1/2} C_{i+1/2}^{n+1/2} = \frac{|Q_{i+1/2}^{n+1/2}| + Q_{i+1/2}^{n+1/2}}{2} C_i^n + \frac{|Q_{i+1/2}^{n+1/2}| - Q_{i+1/2}^{n+1/2}}{2} C_{i+1}^n \quad [7.23]$$

7.2.2.3. Boundary conditions

Boundary conditions are needed at every boundary where the characteristics enter the domain. If the discharge $Q_{1/2}^{n+1/2}$ at the left-hand boundary of the domain is positive (Figure 7.3a), a boundary condition must be prescribed at the interface 1/2. Conversely, if the discharge $Q_{M+1/2}^{n+1/2}$ at the right-hand boundary of the domain is negative (Figure 7.3b), a boundary condition must be prescribed at the interface $M+1/2$. A negative discharge $Q_{1/2}^{n+1/2}$ (Figure 7.3c) and a positive discharge $Q_{M+1/2}^{n+1/2}$ (Figure 7.3d) do not require any boundary condition.

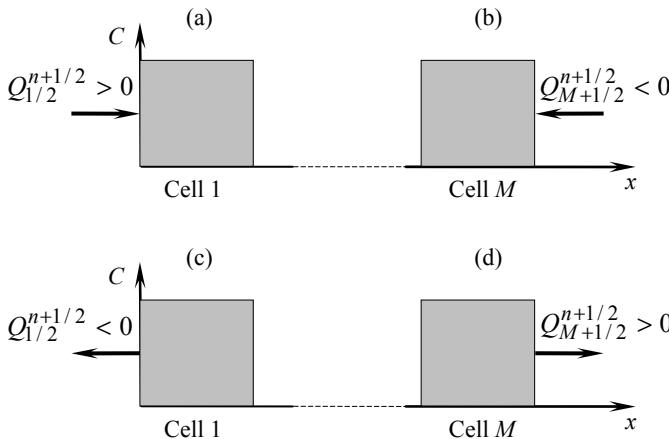


Figure 7.3. Boundary conditions to be prescribed for the various possible flow configurations

Two types of boundary conditions are used in practical solute transport applications:

1) Prescribed concentration C_b at the boundary. For a left-hand boundary the flux is computed as:

$$Q_{1/2}^{n+1/2} C_{1/2}^{n+1/2} = Q_{1/2}^{n+1/2} C_b \quad [7.24]$$

2) Prescribed flux F_b at the boundary. The flux at the left-hand boundary becomes:

$$Q_{1/2}^{n+1/2} C_{1/2}^{n+1/2} = F_b \quad [7.25]$$

Note that the fluxes at outflowing boundaries are given by:

$$\left. \begin{array}{l} Q_{1/2}^{n+1/2} = Q_{1/2}^{n+1/2} C_1^n & \text{if } Q_{1/2}^{n+1/2} < 0 \\ Q_{M+1/2}^{n+1/2} = Q_{M+1/2}^{n+1/2} C_M^n & \text{if } Q_{M+1/2}^{n+1/2} > 0 \end{array} \right\} \quad [7.26]$$

7.2.2.4. Calculation of the liquid discharge at the cell interfaces

Various options are available for the calculation of the discharge $Q_{i+1/2}^{n+1/2}$. The following approaches are proposed:

1) If A and Q are computed using a finite volume method, A is computed over the computational cells and Q is computed at the cell interfaces, which makes its use in equation [7.23] straightforward.

2) If A and Q are available from other techniques such as finite difference methods, they are not available at the interfaces between the cells but at computational points. The point value of A and Q at a given point i may be viewed as their average value over the control volume centered around i . The average discharge $Q_{i+1/2}^{n+1/2}$ over the computational time step may be obtained from an average of the point values in time and space:

$$Q_{i+1/2}^{n+1/2} = \frac{Q_i^n + Q_i^{n+1} + Q_{i+1}^n + Q_{i+1}^{n+1}}{4} \quad [7.27]$$

3) When coupling occurs between hydrodynamics and transport, the variations in A and Q are influenced by those in C . This is the case in sediment transport models, where the suspended sediment concentration influences the erosion or deposition rate, therefore acting on the cross-sectional area. The coupling between the flow and transport processes usually requires an iterative process, whereby A and Q serve as a starting point for the calculation of C over the time step. The result of the transport calculation is used to update the hydrodynamic equation that is solved again over the time step. The result of the updated hydrodynamic equation is used to carry out a new transport calculation. This iterative process is repeated until convergence is achieved. Such a procedure being time-consuming, an explicit coupling is often

preferred. The discharge Q at the interface is computed using only the point values at the beginning of the time step:

$$Q_{i+1/2}^{n+1/2} = \frac{Q_i^n + Q_{i+1}^n}{2} \quad [7.28]$$

7.2.2.5. Algorithm

The algorithm is the following:

- 1) Compute the liquid discharge at each interface between the computational cells. If the discharge was computed using a finite volume method, it may be used directly. If the discharge was computed using a finite difference method, equations [7.27] or [7.28] may be used.
- 2) Compute the flux QC at each interface using equation [7.23] at internal interfaces and equations [7.24], [7.25] or [7.26] depending on the type of the boundary condition.
- 3) Carry out the mass balance using equation [7.17] or [7.18].

7.2.3. Application to the inviscid Burgers equation

7.2.3.1. Discretization

The inviscid Burgers equation is derived in section 1.4. Its conservation form is given by equation [1.69], recalled hereafter:

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} \left(\frac{u^2}{2} \right) = 0$$

The equation may be written in non-conservation form as in equation [1.66]:

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = 0$$

The characteristic form of the equation is given by equation [1.68]:

$$u = \text{Const} \quad \text{for } \frac{dx}{dt} = u$$

Applying the general discretization [7.2] to the conservation form [1.69] leads to:

$$u_i^{n+1} = u_i^n + \frac{\Delta t}{\Delta x_i} \left[(u_{i-1/2}^{n+1/2})^2 - (u_{i+1/2}^{n+1/2})^2 \right] \quad [7.29]$$

7.2.3.2. Flux computation at internal interfaces

The value $u_{i+1/2}^{n+1/2}$ at the interface $i + 1/2$ is obtained from the solution of the following Riemann problem:

$$\left. \begin{aligned} \frac{\partial u}{\partial t} + \frac{\partial}{\partial x} \left(\frac{u^2}{2} \right) &= 0 \\ u(x, t^n) &= \begin{cases} u_L = u_i^n & \text{for } x < x_{i+1/2} \\ u_R = u_{i+1}^n & \text{for } x > x_{i+1/2} \end{cases} \end{aligned} \right\} \quad [7.30]$$

The solution of the Riemann problem for the inviscid Burgers equation is studied in section 4.2.2. The solution may be a shock or a rarefaction wave depending on u_L and u_R (see Table 4.1). Recall that:

- if $u_L < u_R$, a rarefaction wave appears. If $u_L > 0$ the wave travels to the right and u takes the value u_L at $x_{i+1/2}$. If $u_R < 0$ the wave travels to the left and u takes the value u_R at $x_{i-1/2}$. If u_L and u_R do not have the same sign, u is zero at $x_{i+1/2}$;

- if $u_L > u_R$, a shock appears. The speed of the shock is the average of the speeds u_L and u_R (see equations [3.28] and [4.16]). The direction in which the shock propagates depends on the sum $u_L + u_R$. Note that the configuration $u_L + u_R = 0$ leads to indeterminacy because $u(x, t^n)$ is undefined in equation [7.30]. However, determining the solution completely is not needed in this case because the shock is stationary. Using the Rankin-Hugoniot condition [3.28] necessarily leads to a zero flux, which corresponds to $u = 0$ at the interface.

The various possible configurations are summarized in Table 7.1. The value of u at the interface may be computed using the following formula that accounts for the various possible wave configurations:

$$u_{i+1/2}^{n+1/2} = \frac{|\varepsilon_{i+1/2}| + \varepsilon_{i+1/2}}{2} u_i^n + \frac{|\varepsilon_{i+1/2}| - \varepsilon_{i+1/2}}{2} u_{i+1}^n \quad [7.31]$$

where $\varepsilon_{i+1/2}$ is given by:

$$\varepsilon_{i+1/2} = \begin{cases} -1 & \text{if } u_i^n + u_{i+1}^n < 0 \\ 0 & \text{if } u_i^n + u_{i+1}^n = 0 \\ +1 & \text{if } u_i^n + u_{i+1}^n > 0 \end{cases} \quad [7.32]$$

Configuration	Wave pattern	Value of u at $x_{i+1/2}$
$u_L < u_R < 0$	Rarefaction wave heading to the left	u_L
$u_L < 0, u_R > 0$	Rarefaction wave centered around x_0	0
$0 < u_L < u_R$	Rarefaction wave heading to the right	u_R
$u_L > u_R, u_L + u_R < 0$	Shock heading to the left	u_L
$u_L > u_R, u_L + u_R = 0$	Stationary shock	Undefined, 0 acceptable
$u_L > u_R, u_L + u_R < 0$	Shock heading to the right	u_R

Table 7.1. Solution of the Riemann problem for the inviscid Burgers equation.
Solution at the interface for the various possible configurations

7.2.3.3. Boundary conditions

Boundary conditions must be prescribed when the characteristics enter the computational domain. No boundary conditions are needed when the characteristics leave the computational domain. Consequently, only positive velocities may be prescribed at the left-hand boundary, while only negative velocities may be prescribed at the right-hand boundary. These conditions are necessary but not sufficient in that an inflowing boundary condition may be overridden by an outgoing wave, as shown in the following example. Assume that the boundary condition $u_b > 0$ is to be prescribed at the left-hand boundary of the domain. If u_1^n is such that $u_b + u_1^n < 0$, a shock appears. The propagation speed of the shock is negative and the condition u_b cannot be prescribed at the boundary.

The following, general procedure allows the problem to be handled.

1) Define the following Riemann problem at the left-hand boundary:

$$\left. \begin{aligned} \frac{\partial u}{\partial t} + \frac{\partial}{\partial x} \left(\frac{u^2}{2} \right) &= 0 \\ u(x, t^n) &= \begin{cases} u_L = u_b & \text{for } x < x_{1/2} \\ u_R = u_1^n & \text{for } x > x_{1/2} \end{cases} \end{aligned} \right\} \quad [7.33]$$

The solution of equation [7.33] is:

$$u_{1/2}^{n+1/2} = \frac{|\varepsilon_{1/2}| + \varepsilon_{1/2}}{2} u_b + \frac{|\varepsilon_{1/2}| - \varepsilon_{1/2}}{2} u_1^n \quad [7.34]$$

2) Define the following Riemann problem at the right-hand boundary:

$$\left. \begin{aligned} \frac{\partial u}{\partial t} + \frac{\partial}{\partial x} \left(\frac{u^2}{2} \right) &= 0 \\ u(x, t^n) &= \begin{cases} u_L = u_M^n & \text{for } x < x_{M+1/2} \\ u_R = u_b & \text{for } x > x_{M+1/2} \end{cases} \end{aligned} \right\} \quad [7.35]$$

The solution of problem [7.35] is:

$$u_{M+1/2}^{n+1/2} = \frac{|\varepsilon_{M+1/2}| + \varepsilon_{M+1/2}}{2} u_M^n + \frac{|\varepsilon_{M+1/2}| - \varepsilon_{M+1/2}}{2} u_b \quad [7.36]$$

7.2.3.4. Algorithm

The algorithm of Godunov's scheme as applied to the inviscid Burgers equation can be summarized as follows:

- 1) Compute the value of u at the internal interfaces using equations [7.31–32].
- 2) Compute the value of u at the left- and right-hand boundary using equation [7.34] and [7.36] respectively.
- 3) Apply balance equation [7.29] to all the computational cells.

7.2.4. Application to the water hammer equations

7.2.4.1. Discretization

The water hammer equations studied in section 2.4 may be written in conservation form as in equation [2.2], recalled hereafter:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} = S$$

For the sake of clarity, the cross-sectional area A and the speed of sound are assumed to be constant. The pipe is assumed to be horizontal and friction is neglected. Then, U , F and S are given by (see equation [2.68]):

$$U = \begin{bmatrix} \rho A \\ \rho Q \end{bmatrix}, \quad F = \begin{bmatrix} \rho Q \\ Ap \end{bmatrix}, \quad S = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad [7.37]$$

The equation may also be written in the non-conservation form [2.5]:

$$\frac{\partial U}{\partial t} + A \frac{\partial U}{\partial x} = 0 \quad [7.38]$$

where the Jacobian matrix A of F with respect to U is given by equation [2.69]:

$$A = \begin{bmatrix} 0 & 1 \\ c^2 & 0 \end{bmatrix}$$

The speed of sound c is constant. The characteristic form [2.82] of the equation is recalled under the assumption of a zero source term:

$$\left. \begin{array}{l} \frac{d}{dt}(p - \rho cu) = 0 \quad \text{for } \frac{dx}{dt} = -c \\ \frac{d}{dt}(p + \rho cu) = 0 \quad \text{for } \frac{dx}{dt} = c \end{array} \right\} \quad [7.39]$$

Noting that c and ρ are constant, equation [7.39] is simplified into:

$$\left. \begin{array}{ll} p - \rho cu = C_1 & \text{for } \frac{dx}{dt} = -c \\ p + \rho cu = C_2 & \text{for } \frac{dx}{dt} = c \end{array} \right\} \quad [7.40]$$

where C_1 and C_2 are constants to be determined from the initial conditions. The conservation form is discretized as in equation [7.3] under the assumption of a zero source term:

$$U_i^{n+1} = U_i^n + \frac{\Delta t}{\Delta x_i} (F_{i-1/2}^{n+1/2} - F_{i+1/2}^{n+1/2}) \quad [7.41]$$

Equation [7.41] can be rewritten as:

$$\left. \begin{aligned} (\rho A)_i^{n+1} &= (\rho A)_i^n + \frac{\Delta t}{\Delta x_i} \left[(\rho Q)_{i-1/2}^{n+1/2} - (\rho Q)_{i+1/2}^{n+1/2} \right] \\ (\rho Q)_i^{n+1} &= (\rho Q)_i^n + \frac{\Delta t}{\Delta x_i} \left[(Ap)_{i-1/2}^{n+1/2} - (Ap)_{i+1/2}^{n+1/2} \right] \end{aligned} \right\} \quad [7.42]$$

Noting that $d(Ap) = c^2 d(\rho A)$, dividing the first and second equations [7.42] by A and Q respectively leads to:

$$\left. \begin{aligned} p_i^{n+1} &= p_i^n + \frac{\rho c^2}{A} \frac{\Delta t}{\Delta x_i} (Q_{i-1/2}^{n+1/2} - Q_{i+1/2}^{n+1/2}) \\ Q_i^{n+1} &= Q_i^n + \frac{A}{\rho} \frac{\Delta t}{\Delta x_i} (p_{i-1/2}^{n+1/2} - p_{i+1/2}^{n+1/2}) \end{aligned} \right\} \quad [7.43]$$

7.2.4.2. Flux calculation at internal interfaces

The flux at the interface $i + 1/2$ is computed from the solution of the following the Riemann problem:

$$\left. \begin{aligned} \frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} &= 0 \\ U(x, t^n) &= \begin{cases} U_L = U_i^n & \text{for } x < x_{i+1/2} \\ U_R = U_{i+1}^n & \text{for } x > x_{i+1/2} \end{cases} \end{aligned} \right\} \quad [7.44]$$

The solution of the Riemann problem for the water hammer equations is examined in detail in section 4.3.2. An intermediate region of constant state is separated from the left and right states by two contact discontinuities propagating in opposite directions at speeds $-c$ and $+c$. The interface $i + 1/2$ belongs to the intermediate region of constant state. The value of U and F at $x_{i+1/2}$ is determined using the first Riemann invariant [7.40] between U_R and the intermediate region and the second Riemann invariant [7.40] between U_L and the intermediate region of constant state. The following system is obtained:

$$\left. \begin{aligned} p_{i+1/2}^{n+1/2} - \rho c u_{i+1/2}^{n+1/2} &= p_R - \rho c u_R = p_{i+1}^n - \rho c u_{i+1}^n \\ p_{i+1/2}^{n+1/2} + \rho c u_{i+1/2}^{n+1/2} &= p_L + \rho c u_L = p_i^n + \rho c u_i^n \end{aligned} \right\} \quad [7.45]$$

Solving the system for p and u at the interface leads to:

$$\left. \begin{aligned} p_{i+1/2}^{n+1/2} &= \frac{p_i^n + p_{i+1}^n}{2} + \rho c \frac{u_i^n - u_{i+1}^n}{2} \\ u_{i+1/2}^{n+1/2} &= \frac{u_i^n + u_{i+1}^n}{2} + \frac{p_i^n - p_{i+1}^n}{2\rho c} \end{aligned} \right\} [7.46]$$

Consequently:

$$\left. \begin{aligned} p_{i+1/2}^{n+1/2} &= \frac{1}{2}(p_i^n + p_{i+1}^n) + \frac{\rho c}{2A}(Q_i^n - Q_{i+1}^n) \\ Q_{i+1/2}^{n+1/2} &= \frac{1}{2}(Q_i^n + Q_{i+1}^n) + \frac{A}{2\rho c}(p_i^n - p_{i+1}^n) \end{aligned} \right\} [7.47]$$

Substituting equation [7.47] into equation [7.43] leads to the final expression:

$$\left. \begin{aligned} p_i^{n+1} &= \frac{Cr}{2}(p_{i-1}^{n+1} + p_{i+1}^{n+1}) + (1 - Cr)p_i^n + Cr \frac{\rho c}{2A}(Q_{i-1}^{n+1} - Q_{i+1}^{n+1}) \\ Q_i^{n+1} &= \frac{Cr}{2}(Q_{i-1}^{n+1} + Q_{i+1}^{n+1}) + (1 - Cr)Q_i^n + Cr \frac{A}{2\rho c}(p_{i-1}^{n+1} - p_{i+1}^{n+1}) \end{aligned} \right\} [7.48]$$

where the Courant number is defined as:

$$Cr = \frac{c\Delta t}{\Delta x_i}$$

7.2.4.3. Treatment of boundary conditions

The following types of boundary conditions are dealt with hereafter: prescribed pressure and prescribed discharge. The calculation of the flux is detailed only for the left-hand boundary, the transposition to the right-hand boundary being straightforward.

– Prescribed pressure p_b . The pressure at the interface 1/2 is defined as:

$$p_{1/2}^{n+1/2} = p_b [7.49]$$

The discharge is obtained using the first Riemann invariant between the cell 1 and the interface 1/2. The first relationship [7.40] is rewritten as:

$$p_{1/2}^{n+1/2} - \rho c u_{1/2}^{n+1/2} = p_1^n - \rho c u_1^n [7.50]$$

Substituting equation [7.49] into equation [7.50] yields:

$$u_{1/2}^{n+1/2} = \frac{p_b - p_1^n}{\rho c} + u_1^n \quad [7.51]$$

Multiplying by the cross-sectional area A gives:

$$Q_{1/2}^{n+1/2} = \frac{A}{\rho c} (p_b - p_1^n) + Q_1^n \quad [7.52]$$

Reasoning by symmetry, the following formulae are obtained for the right-hand boundary:

$$\left. \begin{aligned} p_{M+1/2}^{n+1/2} &= p_b \\ Q_{M+1/2}^{n+1/2} &= \frac{A}{\rho c} (p_M^n - p_b) + Q_M^n \end{aligned} \right\} \quad [7.53]$$

– Prescribed discharge Q_b at the boundary. The discharge is known directly from the boundary condition:

$$Q_{1/2}^{n+1/2} = Q_b \quad [7.54]$$

and the pressure is obtained from the first relationship [7.40]:

$$p_{1/2}^{n+1/2} - \rho c \frac{Q_b}{A} = p_1^n - \rho c u_1^n \quad [7.55]$$

Equation [7.55] leads to:

$$p_{1/2}^{n+1/2} = p_1^n + \frac{\rho c}{A} (Q_b - Q_1^n) \quad [7.56]$$

Using the second relationship [7.40] at the right-hand boundary leads to:

$$p_{M+1/2}^{n+1/2} = p_M^n + \frac{\rho c}{A} (Q_M^n - Q_b) \quad [7.57]$$

7.2.4.4. Algorithm

The algorithm for the solution of the water hammer equation by Godunov's scheme can be summarized as follows:

- 1) Compute the values of p and Q at the internal interfaces using equations [7.47].
- 2) Compute the values of p and Q at the boundaries of the domain using equation [7.52] or [7.54] at the left-hand boundary and using equation [7.53] or [7.57] at the right-hand boundary.
- 3) Apply balance equation [7.43] to all the computational cells in the domain.

7.3. Higher-order Godunov-type schemes

7.3.1. Rationale and principle

7.3.1.1. Historical perspective

Godunov's scheme [GOD 59] is first-order. First-order schemes are characterized by a strong numerical diffusion, the effect of which is to smooth out the numerical solutions, in particular in the neighborhood of shocks, contact discontinuities or sharp gradient transitions. The quality of the numerical solution can be increased only by increasing the number of computational cells, with consequences on the computational effort required. At the end of the 1970s Van Leer [VAN 77, VAN 79] introduced a new formalism for the development of higher-order conservative schemes. The purpose was twofold. Firstly, the numerical method had to be conservative, a necessary condition for the treatment of weak solutions. Secondly, the scheme had to be monotone in order to eliminate spurious oscillations in the vicinity of sharp gradients. Van Leer introduced the concepts of reconstruction and slope limiting, the principle of which is outlined in the next sections. These concepts were used by Van Leer to develop the now widely used MUSCL scheme. At the beginning of the 1980s, Colella and Woodward proposed the more complex but much more accurate PPM [COL 84]. Considerable research effort has been devoted to the development of higher-order schemes since then, leading to numerous higher-order Godunov-type schemes (see e.g. [TOR 97], [LEV 02], [GUI 03a]).

Today's formalism of Godunov-type schemes uses the following steps: (i) reconstruction, (ii) profile limiting, (iii) solution of a generalized Riemann problem, (iv) flux computation and (v) balance over the computational cells. These steps are detailed in sections 7.3.1.2 to 7.3.1.5.

7.3.1.2. Reconstruction of the flow variable

Higher-order Godunov-type schemes are derived from the following remark: the accuracy of the numerical solution is conditioned by that of the method used for the computation of the fluxes. The accuracy of the flux computation can be increased

only by more accurately locating the gradients in the variable than the original Godunov scheme does. The lack of accuracy of Godunov's scheme mainly stems from the fact that the Riemann problem is defined using the average value of the variable over the computational cells, that is, assuming a zero gradient in the computational cells. Van Leer proposed that the accuracy of the scheme should be increased by reconstructing the variations of U within a given cell using the average values of U over the neighboring cells.

The values U_i^n of U over the cells being known at the time level n , a reconstructed profile, denoted by $\tilde{U}_i^n(x)$, is defined over each cell (Figure 7.4). The reconstruction must satisfy conservation, that is, its average value over the cell i should be equal to U_i^n :

$$\int_{x_{i-1/2}}^{x_{i+1/2}} \tilde{U}_i^n(x) dx = \Delta x_i U_i^n \quad [7.58]$$

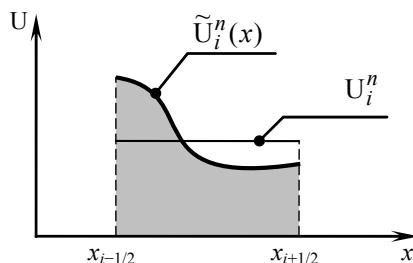


Figure 7.4. Reconstruction of the flow variable within the computational cell i . The area below the reconstructed profile (gray-shaded region) is equal to the area below the horizontal line that indicates the average value

7.3.1.3. Slope limiting

The reconstructed profile $\tilde{U}_i^n(x)$ must be corrected prior to the computation of the fluxes at the interfaces between the computational cells. As shown in [VAN 77] and [COL 84] the following necessary conditions should be fulfilled for the TVD character of the scheme to be guaranteed:

- The reconstructed profile $\tilde{U}_i^n(x)$ must be monotone over the cell i .
- The value $\tilde{U}_i^n(x_{i-1/2})$ at the interface $i - 1/2$ must lie between U_{i-1}^n and U_i^n .

- The value $\tilde{U}_i^n(x_{i+1/2})$ at the interface $i + 1/2$ must lie between U_i^n and U_{i+1}^n .

These conditions are summarized as follows:

$$\left. \begin{aligned} \min(U_{i-1}^n, U_i^n) &\leq \tilde{U}_i^n(x_{i-1/2}) \leq \max(U_{i-1}^n, U_i^n) \\ \min(U_i^n, U_{i+1}^n) &\leq \tilde{U}_i^n(x_{i+1/2}) \leq \max(U_i^n, U_{i+1}^n) \end{aligned} \right\} [7.59]$$

In the notation [7.59], the min, max and \leq operators are applied to each of the components of U separately.

The reconstructed profile is corrected if necessary. Applying the monotony conditions amounts to minimizing, or limiting, the average gradient of the variable over the computational cell, hence the term “slope limiting” often used to refer to the correction. The effect of slope limiting on the reconstructed profile is illustrated in Figure 7.5. Note that the corrected profile should also satisfy the conservation constraint [7.58].

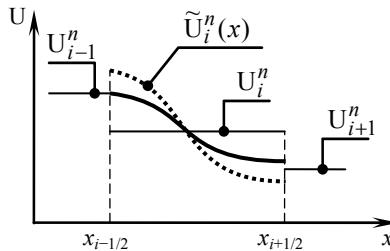


Figure 7.5. Effect of slope limiting on a profile to meet the monotony conditions [7.59]. Reconstructed profile before the correction (dashed line), after the correction (solid line)

7.3.1.4. Solution of the Riemann problem at the interfaces between the cells

The limited profiles are used to define generalized Riemann problems at the interfaces between the computational cells (Figure 7.6). Such problems take the form:

$$\left. \begin{aligned} \frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} &= 0 \\ U(x, t^n) &= \begin{cases} \tilde{U}_i^n(x) & \text{for } x < x_{i+1/2} \\ \tilde{U}_{i+1}^n(x) & \text{for } x > x_{i+1/2} \end{cases} \end{aligned} \right\} [7.60]$$

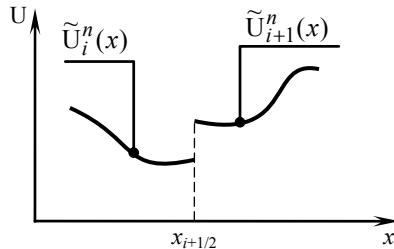


Figure 7.6. Generalized Riemann problem at the interface $i+1/2$

Such problems can in general not be solved analytically. They must be converted into equivalent Riemann problems [GUI 03a] that can be solved exactly or approximately, or they can be solved using so-called predictor-corrector methods [TOR 97]. Appendix C gives an overview of a number of approximate Riemann solvers available in the literature.

7.3.1.5. Flux computation and balance

Solving the Riemann problem at the interfaces between the computational cells allows the fluxes to be computed. The fluxes are used in balance equation [7.3], recalled hereafter:

$$U_i^{n+1} = U_i^n + \frac{\Delta t}{\Delta x_i} (F_{i-1/2}^{n+1/2} - F_{i+1/2}^{n+1/2}) + \Delta t S_i^{n+1/2}$$

7.3.2. Example: the MUSCL scheme

7.3.2.1. Reconstruction

The Monotonic Upwind Scheme for Conservation Laws (MUSCL) was developed by Van Leer [VAN 77]. Several options for the reconstruction of the conserved variable were proposed in the original publication. Only the most commonly used approach is presented here. The procedure is detailed for a scalar variable. It is generalized to vector variables by applying the reconstruction and limiting steps to each of the components of the vector variable.

The MUSCL scheme uses a linear reconstruction in the form (see Figure 7.7):

$$\tilde{U}_i^n = U_i^n + (x - x_i) a_i^n \quad [7.61]$$

where a_i^n is the slope of the profile over the cell i and $x_i = (x_{i-1/2} + x_{i+1/2})/2$ is the abscissa of the centre of the cell i . It is easy to check that equation [7.61] satisfies the conservation property [7.58] regardless of the value of a_i^n . The slope is computed as the average slope between the cells $i - 1$ and $i + 1$ (Figure 7.7):

$$a_i^n = \frac{U_{i+1}^n - U_{i-1}^n}{x_{i+1} - x_{i-1}} = 2 \frac{U_{i+1}^n - U_{i-1}^n}{\Delta x_{i-1} + 2\Delta x_{i-1} + \Delta x_{i+1}} \quad [7.62]$$

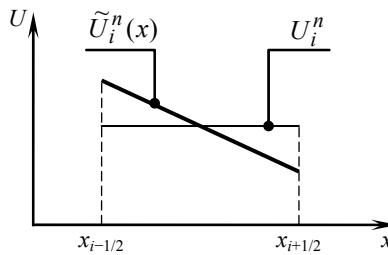


Figure 7.7. The MUSCL reconstruction

7.3.2.2. Slope limiting

The slope must be limited if at least one of the following situations occurs:

- 1) The cell i is a local extremum. This is true if:

$$(U_i^n - U_{i-1}^n)(U_{i+1}^n - U_i^n) \leq 0 \quad [7.63]$$

In such a case the slope a_i^n is set to zero. Tests 2) and 3) hereafter do not need to be carried out.

2) The first condition [7.59] is not satisfied because the value at the interface $i - 1/2$ does not lie between the average values in the cells $i - 1$ and i . The slope a_i^n is adjusted to the largest possible value that allows the first condition [7.59] to be satisfied:

$$a_i^n = 2 \frac{U_i^n - U_{i-1}^n}{\Delta x_i} \quad [7.64]$$

3) The second condition [7.59] is not satisfied because the value at the interface $i + 1/2$ does not lie between the average values in the cells i and $i + 1$. The slope a_i^n

is adjusted to the largest possible value that allows the second condition [7.59] to be satisfied:

$$a_i^n = 2 \frac{U_{i+1}^n - U_i^n}{\Delta x_i} \quad [7.65]$$

7.3.2.3. Solution of the generalized Riemann problem

Several approaches are available for the solution of the generalized Riemann problem. The most widely used approach, referred to as the MUSCL-Hancock approach [TOR 97], is presented hereafter. Another approach, proposed by Savic and Holly [SAV 93], uses the average value of the Riemann invariants over the domain of dependence. A third approach, referred to as the EigenVector-based Reconstruction (EVR), allows for faster computations while leading to more stable solutions than the MUSCL-Hancock scheme [GUI 03a, SOA 07]. This latter approach is detailed in section 7.4.

In the MUSCL-Hancock approach, a first approximation is obtained for the fluxes by solving the following Riemann problem:

$$\left. \begin{aligned} \frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} &= 0 \\ \mathbf{U}(x, t^n) &= \begin{cases} \tilde{\mathbf{U}}_i^n(x_{i+1/2}) & \text{for } x < x_{i+1/2} \\ \tilde{\mathbf{U}}_{i+1}^n(x_{i+1/2}) & \text{for } x > x_{i+1/2} \end{cases} \end{aligned} \right\} \quad [7.66]$$

Solving the Riemann problem [7.66] yields fluxes that are used to compute the solution using equation [7.3] over half a time step:

$$\left. \begin{aligned} \mathbf{U}_{i+1/2,L}^{n+1/2} &= \tilde{\mathbf{U}}_i^n(x_{i+1/2}) \\ &\quad + \frac{\Delta t}{2\Delta x_i} \left[\mathbf{F}(\tilde{\mathbf{U}}_i^n(x_{i-1/2})) - \mathbf{F}(\tilde{\mathbf{U}}_i^n(x_{i+1/2})) \right] \\ \mathbf{U}_{i+1/2,R}^{n+1/2} &= \tilde{\mathbf{U}}_{i+1}^n(x_{i+1/2}) \\ &\quad + \frac{\Delta t}{2\Delta x_{i+1}} \left[\mathbf{F}(\tilde{\mathbf{U}}_{i+1}^n(x_{i+1/2})) - \mathbf{F}(\tilde{\mathbf{U}}_{i+1}^n(x_{i+3/2})) \right] \end{aligned} \right\} \quad [7.67]$$

The values $U_{i+1/2,L}^{n+1/2}$ and $U_{i+1/2,R}^{n+1/2}$ are taken as the left and right states of a Riemann problem, the solution of which is used as a final estimate for the fluxes in equation [7.3].

7.4. EVR approach

7.4.1. Principle of the approach

The EVR method allows the generalized Riemann problem to be solved by converting it into an Equivalent Riemann Problem (ERP). The method was first introduced for the simulation of two-phase flows in pipes [GUI 01b]. It was then generalized to hyperbolic systems of conservation laws [GUI 03a] but no name was given to the method. The name EVR appeared when an application of the method to two-dimensional free surface flow simulations with dry beds was published [SOA 07].

The advantage of the EVR is a less computationally demanding algorithm than the MUSCL-Hancock method, with a similar precision. The method is also more robust than the classical MUSCL-Hancock approach in the presence of dry beds, with no oscillations or instabilities near wetting and drying fronts [SOA 07]. The method is presented in a one-dimensional context hereafter. However, its generalization to multiple dimensions does not introduce any particular difficulty.

Assume that the vector variable \mathbf{U} has been reconstructed in each cell of the computational domain at time step n . The reconstructed profile in cell i at time level n is denoted by $\tilde{\mathbf{U}}_i^n(x)$. At interface $i + 1/2$ between the cells i and $i + 1$, the generalized Riemann problem is given by equation [7.60], recalled here:

$$\left. \begin{aligned} \frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} &= 0 \\ \mathbf{U}(x, t^n) &= \begin{cases} \tilde{\mathbf{U}}_i^n(x) & \text{for } x < x_{i+1/2} \\ \tilde{\mathbf{U}}_{i+1}^n(x) & \text{for } x > x_{i+1/2} \end{cases} \end{aligned} \right\}$$

The purpose is to determine the left and right states for an ERP that yields the same average value of the flux \mathbf{F} at interface $i + 1/2$ over the time step Δt . In other words, \mathbf{U}_L and \mathbf{U}_R are sought such that the Riemann problem:

$$\left. \begin{aligned} \frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} &= 0 \\ \mathbf{U}(x, t^n) &= \begin{cases} \mathbf{U}_L & \text{for } x < x_{i+1/2} \\ \mathbf{U}_R & \text{for } x > x_{i+1/2} \end{cases} \end{aligned} \right\} \quad [7.68]$$

leads to the same flux $\mathbf{F}_{i+1/2}^{n+1/2}$ as the original Riemann problem [7.60].

The average flux $F_{i+1/2}^{n+1/2}$ between time levels n and $n + 1$ is estimated from a linearization of the flux function $F(U)$:

$$F_{i+1/2}^{n+1/2} \approx F(U_{i+1/2}^{n+1/2}) \quad [7.69]$$

where $U_{i+1/2}^{n+1/2}$ is an estimate of the average of U at interface $i + 1/2$ between time levels n and $n + 1$. The two Riemann problems [7.60] and [7.67] are equivalent if they lead to the same average $U_{i+1/2}^{n+1/2}$. This allows necessary conditions to be written for the left and right states of the equivalent Riemann problem. This is done by writing the hyperbolic system in non-conservation form, equation [7.38]:

$$\frac{\partial U}{\partial t} + A \frac{\partial U}{\partial x} = 0$$

The underlying idea of the EVR approach is that the reconstructed profiles in the cells i and $i + 1$ can be expressed in the base K of eigenvectors $K^{(p)}$ of the Jacobian matrix A :

$$\left. \begin{aligned} \tilde{U}_i^n(x) &= \sum_{p=1}^m K_i^{n(p)} \tilde{\alpha}_i^{(p)}(x) \\ \tilde{U}_{i+1}^n(x) &= \sum_{p=1}^m K_{i+1}^{n(p)} \tilde{\alpha}_{i+1}^{(p)}(x) \end{aligned} \right\} \quad [7.70]$$

where coefficients $\alpha_i^{(p)}$ and $\alpha_{i+1}^{(p)}$ are called the wave strengths by analogy with Roe's approximate solver (see Appendix C). In reconstruction [7.70], the two eigenvectors $K_i^{n(p)}$ and $K_{i+1}^{n(p)}$ are constant over each cell, while the wave strengths are functions of x . The solution $U(x, t)$ is sought in the form:

$$U(x, t) = \begin{cases} \sum_{p=1}^m K_i^{n(p)} \alpha(x, t) & \text{for } x < x_{i+1/2} \\ \sum_{p=1}^m K_{i+1}^{n(p)} \alpha(x, t) & \text{for } x > x_{i+1/2} \end{cases} \quad [7.71]$$

Substituting equation [7.71] into equation [7.38] leads to:

$$\left. \begin{aligned} \sum_{p=1}^m K_i^{n(p)} \frac{\partial \alpha^{(p)}}{\partial t} + A_i^n \sum_{p=1}^m K_i^{n(p)} \frac{\partial \alpha^{(p)}}{\partial x} &= 0 \\ \sum_{p=1}^m K_{i+1}^{n(p)} \frac{\partial \alpha^{(p)}}{\partial t} + A_{i+1}^n \sum_{p=1}^m K_{i+1}^{n(p)} \frac{\partial \alpha^{(p)}}{\partial x} &= 0 \end{aligned} \right\} \quad [7.72]$$

left-multiplying the first and second equations [7.72] by the inverse of matrices K_i^n and K_{i+1}^n respectively, we obtain:

$$\left. \begin{aligned} \frac{\partial \alpha^{(p)}}{\partial t} + \lambda_i^n \alpha^{(p)} \frac{\partial \alpha^{(p)}}{\partial x} &= 0 \\ \frac{\partial \alpha^{(p)}}{\partial t} + \lambda_{i+1}^n \alpha^{(p)} \frac{\partial \alpha^{(p)}}{\partial x} &= 0 \end{aligned} \right\} \quad [7.73]$$

that is:

$$\left. \begin{aligned} \frac{d\alpha^{(p)}}{dt} &= 0 & \text{for } \frac{dx}{dt} = \lambda_i^n, x < x_{i+1/2} \\ \frac{d\alpha^{(p)}}{dt} &= 0 & \text{for } \frac{dx}{dt} = \lambda_{i+1}^n, x > x_{i+1/2} \end{aligned} \right\} \quad [7.74]$$

Consequently, the wave strength $\alpha^{(p)}$ is a constant along the p th characteristic.

Equations [7.74] are used to express U at the interface as a combination of the wave strengths:

$$\begin{aligned} U_{i+1/2}^{n+1/2} &= \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} U(x_{i+1/2}, t) dt \\ &= \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} \sum_{p=1}^m K^{(p)} \alpha^{(p)}(x_{i+1/2}, t) dt \\ &= \sum_{p=1}^m K^{(p)} \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} \alpha^{(p)}(x_{i+1/2}, t) dt \end{aligned} \quad [7.75]$$

The sum is broken down into:

$$\begin{aligned} U_{i+1/2}^{n+1/2} &= \sum_{p, \lambda^{(p)} < 0}^m K^{(p)} \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} \alpha^{(p)}(x_{i+1/2}, t) dt \\ &\quad + \sum_{p, \lambda^{(p)} > 0}^m K^{(p)} \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} \alpha^{(p)}(x_{i+1/2}, t) dt \end{aligned} \quad [7.76]$$

The waves with positive wave speeds originate from cell i , while those with negative wave speeds come from cell $i + 1$. Equation [7.76] becomes:

$$\begin{aligned} U_{i+1/2}^{n+1/2} &= \sum_{p, \lambda^{(p)} < 0}^m K_{i+1}^{n(p)} \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} \alpha^{(p)}(x_{i+1/2}, t) dt \\ &\quad + \sum_{p, \lambda^{(p)} > 0}^m K_i^{n(p)} \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} \alpha^{(p)}(x_{i+1/2}, t) dt \end{aligned} \quad [7.77]$$

The average value of the wave strength $\alpha^{(p)}$ at interface $i + 1/2$ between t^n and t^{n+1} is equal to its average value over the domain of dependence of the p th wave:

$$\frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} \alpha^{(p)} dt = \begin{cases} \frac{1}{\lambda_i^{n(p)} \Delta t} \int_{x_{i+1/2} - \lambda_i^n \Delta t}^{x_{i+1/2}} \alpha^{(p)}(x, t^n) dx & \text{if } \lambda_i^{n(p)} > 0 \\ \frac{1}{\lambda_{i+1}^{n(p)} \Delta t} \int_{x_{i+1/2} - \lambda_{i+1}^n \Delta t}^{x_{i+1/2}} \alpha^{(p)}(x, t^n) dx & \text{if } \lambda_{i+1}^{n(p)} < 0 \end{cases} \quad [7.78]$$

Substituting equation [7.78] into equation [7.77], we have:

$$\begin{aligned} U_{i+1/2}^{n+1/2} &= \sum_{p, \lambda^{(p)} < 0}^m K_{i+1}^{n(p)} \frac{1}{\lambda_{i+1}^{n(p)} \Delta t} \int_{x_{i+1/2} - \lambda_{i+1}^n \Delta t}^{x_{i+1/2}} \tilde{\alpha}_{i+1}^{n(p)}(x) dx \\ &\quad + \sum_{p, \lambda^{(p)} > 0}^m K_i^{n(p)} \frac{1}{\lambda_i^{n(p)} \Delta t} \int_{x_{i+1/2} - \lambda_i^n \Delta t}^{x_{i+1/2}} \tilde{\alpha}_i^{n(p)}(x) dx \end{aligned} \quad [7.79]$$

Moreover, the left and right states U_L and U_R of the equivalent Riemann problem are sought in the form:

$$\left. \begin{aligned} U_L &= \sum_{p=1}^m K_i^{n(p)} \alpha_L^{(p)} \\ U_R &= \sum_{p=1}^m K_{i+1}^{n(p)} \alpha_R^{(p)} \end{aligned} \right\} \quad [7.80]$$

A necessary condition for equations [7.68] and [7.60] to yield the same average value at interface $i + 1/2$ over the time step is:

$$\left. \begin{aligned} \alpha_L^{(p)} &= \frac{1}{\lambda_i^{n(p)} \Delta t} \int_{x_{i+1/2} - \lambda_i^n \Delta t}^{x_{i+1/2}} \tilde{\alpha}_i^{n(p)}(x) dx && \text{if } \lambda_i^{n(p)} \geq 0 \\ \alpha_R^{(p)} &= \frac{1}{\lambda_{i+1}^{n(p)} \Delta t} \int_{x_{i+1/2} - \lambda_{i+1}^n \Delta t}^{x_{i+1/2}} \tilde{\alpha}_{i+1}^{n(p)}(x) dx && \text{if } \lambda_{i+1}^{n(p)} \leq 0 \end{aligned} \right\} \quad [7.81]$$

In cell i , the wave strengths with negative wave speeds may be chosen arbitrarily. Conversely, in cell $i + 1$, the wave strengths with positive wave speeds may be chosen arbitrarily. In practice, the average value of the wave strength over the computational cell is shown to give satisfactory results [GUI 01b, GUI 03a, SOA 07]. The final formulae of the wave strengths are:

$$\left. \begin{aligned} \alpha_L^{(p)} &= \begin{cases} \frac{1}{\lambda_i^{n(p)} \Delta t} \int_{x_{i+1/2} - \lambda_i^n \Delta t}^{x_{i+1/2}} \tilde{\alpha}_i^{n(p)}(x) dx & \text{if } \lambda_i^{n(p)} \geq 0 \\ \alpha_i^n & \text{if } \lambda_i^{n(p)} < 0 \end{cases} \\ \alpha_R^{(p)} &= \begin{cases} \frac{1}{\lambda_{i+1}^{n(p)} \Delta t} \int_{x_{i+1/2} - \lambda_{i+1}^n \Delta t}^{x_{i+1/2}} \tilde{\alpha}_{i+1}^{n(p)}(x) dx & \text{if } \lambda_{i+1}^{n(p)} \leq 0 \\ \alpha_{i+1}^n & \text{if } \lambda_{i+1}^{n(p)} > 0 \end{cases} \end{aligned} \right\} \quad [7.82]$$

7.4.2. Application to the one-dimensional shallow water equations

The one-dimensional shallow water equations arise as a particular case of the Saint Venant equations [2.2], [2.118] under the assumption of a wide, prismatic rectangular channel. In such a case, the definition of U, F and S is modified into:

$$U = \begin{bmatrix} h \\ q \end{bmatrix}, \quad F = \begin{bmatrix} q \\ q^2/h + gh^2/2 \end{bmatrix}, \quad S = \begin{bmatrix} 0 \\ (S_0 - S_f)gh \end{bmatrix} \quad [7.83]$$

These equations can also be seen as the one-dimensional restriction of equations [5.64]. The Jacobian matrix A is given by equation [2.119]:

$$A = \begin{bmatrix} 0 & 1 \\ c^2 - u^2 & 2u \end{bmatrix}$$

and the eigenvectors are given by equation [2.125]:

$$\mathbf{K}^{(1)} = \begin{bmatrix} 1 \\ u - c \end{bmatrix}, \quad \mathbf{K}^{(2)} = \begin{bmatrix} 1 \\ u + c \end{bmatrix}$$

The wave strengths are defined as:

$$\mathbf{U} = \alpha^{(1)} \mathbf{K}^{(1)} + \alpha^{(2)} \mathbf{K}^{(2)} = \mathbf{K} \alpha \quad [7.84]$$

where α is the vector formed by the wave strengths. Substituting equations [7.83] and [2.125] into equation [7.84] leads to:

$$\left. \begin{aligned} \alpha^{(1)} + \alpha^{(2)} &= h \\ (u - c)\alpha^{(1)} + (u + c)\alpha^{(2)} &= q \end{aligned} \right\} \quad [7.85]$$

Solving the system for $\alpha^{(1)}$ and $\alpha^{(2)}$ gives:

$$\alpha^{(1)} = \alpha^{(2)} = \frac{h}{2} \quad [7.86]$$

Consequently, reconstructing h suffices to reconstruct \mathbf{U} completely. Reconstructing the unit discharge q is not necessary. Assume now that h has been reconstructed using the MUSCL approach. The reconstructed profile can be expressed as:

$$\tilde{h}_i^n(x) = h_i^n + (x - x_i) a_i^n \quad [7.87]$$

where x_i is the abscissa of the center of cell i . The left state of the equivalent Riemann problem at interface $i + 1/2$ is:

$$\begin{aligned} \begin{bmatrix} h_L \\ q_L \end{bmatrix} &= \frac{h^{(1)}}{2} \begin{bmatrix} 1 \\ u_i^n - c_i^n \end{bmatrix} + \frac{h^{(2)}}{2} \begin{bmatrix} 1 \\ u_i^n + c_i^n \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} \frac{h^{(1)} + h^{(2)}}{2} \\ \frac{h^{(1)} + h^{(2)}}{2} u_i^n + \frac{h^{(2)} - h^{(1)}}{2} c_i^n \end{bmatrix} \end{aligned} \quad [7.88]$$

where $h^{(1)}$ and $h^{(2)}$ are respectively the average values of h over the dependence domains of the waves $\lambda^{(1)} = u - c$ and $\lambda^{(2)} = u + c$. They are given by equation [7.82] as:

$$h^{(1)} = \begin{cases} \frac{1}{(u_i^n - c_i^n)\Delta t} \int_{x_{i+1/2} - (u_i^n - c_i^n)\Delta t}^{x_{i+1/2}} h_i^n(x) dx & \text{if } u_i^n - c_i^n \geq 0 \\ h_i^n & \text{if } u_i^n - c_i^n < 0 \end{cases} \quad [7.89]$$

$$h^{(2)} = \begin{cases} \frac{1}{(u_i^n + c_i^n)\Delta t} \int_{x_{i+1/2} - (u_i^n + c_i^n)\Delta t}^{x_{i+1/2}} \tilde{h}_i^n(x) dx & \text{if } u_i^n + c_i^n \geq 0 \\ h_i^n & \text{if } u_i^n + c_i^n < 0 \end{cases}$$

Substituting equation [7.87] into equation [7.89] and noting that $x_{i+1/2} = x_i + \Delta x/2$, we have:

$$h^{(1)} = \begin{cases} h_i^n + \frac{a_i^n}{2} [\Delta x_i - (u_i^n - c_i^n)\Delta t] & \text{if } u_i^n - c_i^n \geq 0 \\ h_i^n & \text{if } u_i^n - c_i^n < 0 \end{cases} \quad [7.90]$$

$$h^{(2)} = \begin{cases} h_i^n + \frac{a_i^n}{2} [\Delta x_i - (u_i^n + c_i^n)\Delta t] & \text{if } u_i^n + c_i^n \geq 0 \\ h_i^n & \text{if } u_i^n + c_i^n < 0 \end{cases}$$

Applying a similar reasoning to cell $i + 1$, the right state of the equivalent Riemann problem is found to be:

$$\begin{bmatrix} h_R \\ q_R \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \frac{h^{(1)} + h^{(2)}}{2} \\ \frac{h^{(1)} + h^{(2)}}{2} u_{i+1}^n + \frac{h^{(2)} - h^{(1)}}{2} c_{i+1}^n \end{bmatrix} \quad [7.91]$$

where the wave strengths are given by:

$$h^{(1)} = \begin{cases} h_{i+1}^n - \frac{a_{i+1}^n}{2} [\Delta x_i + (u_{i+1}^n - c_{i+1}^n)\Delta t] & \text{if } u_{i+1}^n - c_{i+1}^n \leq 0 \\ h_{i+1}^n & \text{if } u_{i+1}^n - c_{i+1}^n > 0 \end{cases} \quad [7.92]$$

$$h^{(2)} = \begin{cases} h_{i+1}^n - \frac{a_{i+1}^n}{2} [\Delta x_i + (u_{i+1}^n + c_{i+1}^n)\Delta t] & \text{if } u_{i+1}^n + c_{i+1}^n \leq 0 \\ h_{i+1}^n & \text{if } u_{i+1}^n + c_{i+1}^n > 0 \end{cases}$$

As a conclusion, the EVR approach requires only one variable reconstruction (the water depth h), against two reconstructions for the MUSCL-Hancock approach. Moreover, the EVR approach uses a first-order, explicit time marching approach, against two steps for the second-order in time MUSCL-Hancock.

7.5. Summary

7.5.1. *What you should remember*

Finite volume schemes solve the conservation form of conservation laws. The discretized quantity is the average value of the conserved variable over the computational cells.

The change in the conserved variable from one time level to the next is computed using a balance equation over the computational cells. This requires that the fluxes be estimated at the interfaces between the cells.

Godunov-type schemes use Riemann problems to estimate the fluxes at the interfaces between the cells. The Riemann problem may be solved exactly or approximately.

Higher-order Godunov-type schemes use a reconstruction procedure to estimate the variations of the conserved variable in the computational cells. The variable in a given cell is reconstructed using the average values in the neighboring cells. This allows the gradients to be located more accurately, thus leading to more accurate estimates of the fluxes.

The second-order MUSCL scheme presented in section 7.3 uses a linear reconstruction of the conserved variable. The slope of the profile within the cell i is computed as the average slope between the cells $i - 1$ and $i + 1$. Spurious oscillations are eliminated from the solution by limiting the slope of the reconstructed profile, as indicated in section 7.3.2.2.

The reconstructed variables lead to generalized Riemann problems at the cell interfaces. Such generalized Riemann problems can be solved using the two-step MUSCL-Hancock approach (valid only for a linear reconstruction) presented in section 7.3.2.3 or using the EVR approach presented in section 7.4, with a particular application to the Saint Venant equations in section 7.4.2. When applied to the shallow water equations, the EVR approach has the advantage that only the water depth needs to be reconstructed and the Riemann problem needs to be solved only once (compared to two reconstructed variables and two solutions of the Riemann problem for the MUSCL-Hancock approach). The EVR approach allows the

computational effort to be reduced substantially, with an increased robustness of the numerical solution in the presence of dry beds.

7.5.2. Application exercises

Apply Godunov's scheme to the water hammer equations, the Saint Venant equations and the Euler equations. Use simplifying assumptions on the geometry (e.g. horizontal pipe/channel, negligible friction, etc.) so that the source term is assumed to be zero. Solve the Riemann problems dealt with in Chapter 4 and comment on the accuracy of the numerical solution.

Indications and searching tips for the solution of this exercise can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

Chapter 8

Finite Element Methods for Hyperbolic Systems

8.1. Principle for one-dimensional scalar laws

8.1.1. Weak form

The conservation form [1.1] of a scalar hyperbolic conservation law (see section 1.1.1) is recalled:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} = S$$

This equation may also be written as in equation [3.20], recalled here:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} - S = 0$$

Assume that equation [1.1] is to be solved over a computational domain $[0, L]$, with initial and boundary conditions defined so as to guarantee solution existence and uniqueness (see section 1.2.2 for details).

The first step consists of multiplying equation [3.20] with a so-called weighting function $w(x, t)$, also called a test function:

$$\left(\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} - S \right) w = 0 \quad \begin{cases} \forall w(x, t) \\ \forall (x, t) \end{cases} \quad [8.1]$$

Equation [8.1] is integrated over the solution domain $[0, L]$ between times t_1 and t_2 :

$$\int_{t_1}^{t_2} \int_0^L \left(\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} - S \right) w \, dx \, dt = 0 \quad [8.2]$$

Equation [8.2] is a particular case of the weak form [3.21] seen in section 3.4.1. Equation [8.2] is obtained from equation [3.21] by setting $x_1 = 0$ and $x_2 = L$. Finite element methods seek a solution to equation [8.2].

Note that the non-conservation form [1.17] of the equation (see section 1.1.3 for more details) may also be solved. Reproducing the above reasoning for equation [1.17] yields:

$$\int_{t_1}^{t_2} \int_0^L \left(\frac{\partial U}{\partial t} + \lambda \frac{\partial U}{\partial x} - S' \right) w \, dx \, dt = 0 \quad [8.3]$$

As mentioned in Chapter 3, equations [1.1], [1.17], [8.2] and [8.3] are equivalent as long as the solution U is continuous. When the solution is discontinuous, the choice of the formulation and the choice of the test function w may have important consequences on the behavior of the solution (see section 8.5).

Solving the weak forms [8.2] or [8.3] amounts to solving the original conservation law in an average sense over the solution domain. The averaging is a function of the weighting function w used. For this reason, the approach is sometimes referred to as the “weighted residuals” method.

8.1.2. Discretization of space and time

8.1.2.1. Principle

Finite element methods [HER 07] are similar to finite difference and finite volume approaches in that they solve a discretized version of the governing equations. Space is discretized into pre-defined computational points (called nodes) and time is discretized into pre-defined time levels (Figure 8.1). In contrast with finite differences (see section 6.1.1), finite element methods do not seek the computational solution at the computational nodes only. The finite element solution is defined at all points of the domain $[0, L]$ for a given time. It is sought in the form:

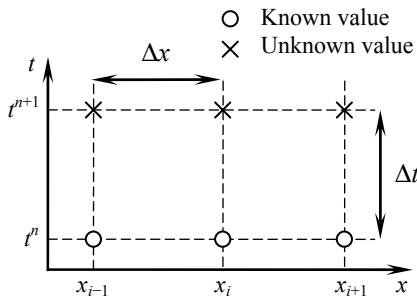


Figure 8.1. Discretization of space and time for a one-dimensional problem

$$U(x, t^n) = \sum_{i=1}^M U_i^n s_i(x) \quad [8.4]$$

where functions $s_i(x)$ are so-called shape functions. They are defined *a priori*. The U_i^n are coefficients allowing for a linear combination of the shape functions. The solution is known completely at time level n if all the coefficients U_i^n can be computed.

The flux and source term are sought in the form:

$$\left. \begin{aligned} F(x, t^n) &= \sum_{i=1}^M F_i^n s_i(x) \\ S(x, t^n) &= \sum_{i=1}^M S_i^n s_i(x) \end{aligned} \right\} \quad [8.5]$$

In classical finite element approaches, s_i is equal to 1 at node i and is zero at all other nodes. It is non-zero over the interval $[x_{i-1}, x_{i+1}]$ (Figure 8.2).

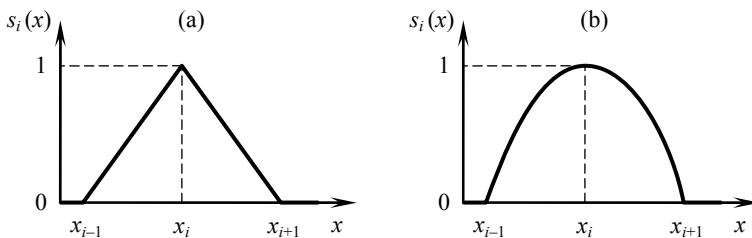


Figure 8.2. Two examples of shape functions: piecewise linear (a), piecewise parabolic (b)

8.1.2.2. Discretization of the conservation form

Let $t_1 = t^n$ and $t_2 = t^{n+1}$ in equation [8.2]. Swapping the time and space integrals leads to:

$$\int_0^L [U^{n+1}(x) - U^n(x)] w \, dx + \int_0^{L t^{n+1}} \left(\frac{\partial F}{\partial x} - S \right) w \, dt \, dx = 0 \quad [8.6]$$

The time integral of $(\partial F / \partial x - S)w$ is approximated as:

$$\int_{t^n}^{t^{n+1}} \left(\frac{\partial F}{\partial x} - S \right) w \, dt \approx \left[(1-\theta) \left(\frac{\partial F}{\partial x} - S \right)^n + \theta \left(\frac{\partial F}{\partial x} - S \right)^{n+1} \right] \Delta t \, w \quad [8.7]$$

where θ is an implicitation parameter between 0 and 1. Seeking the solution in the form [8.4–5] and substituting equation [8.7] into equation [8.6], we obtain

$$\begin{aligned} & \int_0^L \sum_{i=1}^M (U_i^{n+1} - U_i^n) s_i w \, dx + \Delta t \int_0^L \frac{\partial}{\partial x} \sum_{i=1}^M [(1-\theta)F_i^n + \theta F_i^{n+1}] s_i w \, dx \\ & + \Delta t \int_0^L \sum_{i=1}^M [(1-\theta)S_i^n + \theta S_i^{n+1}] s_i w \, dx = 0 \end{aligned} \quad [8.8]$$

Swapping the sum operators and partial derivatives yields

$$\begin{aligned} & \sum_{i=1}^M (U_i^{n+1} - U_i^n) \int_0^L s_i w \, dx + \Delta t \sum_{i=1}^M [(1-\theta)F_i^n + \theta F_i^{n+1}] \int_0^L \frac{\partial s_i}{\partial x} w \, dx \\ & + \Delta t \sum_{i=1}^M [(1-\theta)S_i^n + \theta S_i^{n+1}] \int_0^L s_i w \, dx = 0 \end{aligned} \quad [8.9]$$

The shape functions s_i and the weighting functions w being known, their integrals over the computational domain are known too. Equation [8.9] may be rewritten in the form:

$$\begin{aligned} & \sum_{i=1}^M \left\{ U_i^{n+1} - U_i^n + [(1-\theta)S_i^n + \theta S_i^{n+1}] \Delta t \right\} C_{i,w} \\ & + \sum_{i=1}^M [(1-\theta)F_i^n + \theta F_i^{n+1}] \Delta t D_{i,w} = 0 \end{aligned} \quad [8.10]$$

where coefficients $A_{i,w}$ and $B_{i,w}$ are given by:

$$\left. \begin{aligned} C_{i,w} &= \int_0^L s_i(x) w(x) dx \\ D_{i,w} &= \int_0^L \frac{\partial s_i(x)}{\partial x} w(x) dx \end{aligned} \right\} [8.11]$$

Equation [8.10] involves the unknown nodal values U_i^{n+1} , F_i^{n+1} and S_i^{n+1} at time $n+1$. Since F and S are known functions of U , equation [8.10] may be rewritten so as to involve the single unknown U_i^{n+1} .

Denoting by M the number of nodes in the computational domain, M equations [8.10] can be written. To do so, it is sufficient to choose M different weighting functions w_j ($j = 1, 2, \dots, M$) and determine $C_{i,j}$ and $D_{i,j}$ obtained for each w_j ($j = 1, 2, \dots, M$).

Computation of the coefficients $C_{i,j}$ and $D_{i,j}$ is facilitated if the support of the weighting functions is narrow. A classical choice consists of using weighting functions w_j that are zero except over the interval $[x_{j-1}, x_{j+1}]$ (see section 8.1.3 for typical examples). In such a case, the coefficients $C_{i,j}$ and $D_{i,j}$ are non-zero only for $j = i - 1, j = i$ or $j = i + 1$. Equation [8.10] is rewritten in the form:

$$\begin{aligned} &\sum_{i=1}^M C_{i,j} (U_i^{n+1} + \theta \Delta t S_i^{n+1}) + D_{i,j} \theta \Delta t F_i^{n+1} \\ &= \sum_{i=1}^M C_{i,j} [U_i^n - (1 - \theta) \Delta t S_i^n] - (1 - \theta) \Delta t D_{i,j} F_i^n \end{aligned} [8.12]$$

where $C_{i,j}$ and $D_{i,j}$ are defined as:

$$\left. \begin{aligned} C_{i,j} &= \int_0^L s_i(x) w_j(x) dx \\ D_{i,j} &= \int_0^L \frac{\partial s_i(x)}{\partial x} w_j(x) dx \end{aligned} \right\} [8.13]$$

Writing M equations [8.12] for $j = 1, \dots, M$, leads to an $M \times M$ system of algebraic equations that can be solved uniquely for the nodal values U_i^n .

If F and S are nonlinear functions of U , system [8.12] is nonlinear and must be solved using iterative techniques. If F and S are linear functions of U , [8.12] becomes linear and can be rewritten in the form:

$$RU^{n+1} = b \quad [8.14]$$

where the components of vector U^{n+1} are the nodal unknowns U_i^{n+1} and the j th row of vector b is the right-hand side member of equation [8.12]. Matrix R is often referred to as the mass matrix, or rigidity matrix. Its expression is given in section 8.2 in a number of cases.

8.1.2.3. Discretization of the non-conservation form

An alternative option consists of discretizing the non-conservation form [8.3] of the governing equation. This leads to a system involving only the nodal unknowns U_i^{n+1} . Substituting $t_1 = t^n$, $t_2 = t^{n+1}$ and $w = w_j$ in equation [8.3], swapping the time and space integrals leads to:

$$\int_0^{t^{n+1}} [(U^{n+1}(x) - U^n(x))w_j] dx + \int_0^{t^{n+1}} \left(\lambda \frac{\partial U}{\partial x} - S' \right) w_j dt dx = 0 \quad [8.15]$$

As in section 8.1.2.2, the time integral of $(\lambda \partial U / \partial x - S')w_j$ is approximated as:

$$\begin{aligned} \int_{t^n}^{t^{n+1}} \left(\lambda \frac{\partial U}{\partial x} - S' \right) w_j dt &\approx \left[(1-\theta) \left(\lambda^{n+1/2} \frac{\partial U^n}{\partial x} - S^n \right) \right. \\ &\quad \left. + (1-\theta) \left(\lambda^{n+1/2} \frac{\partial U^{n+1}}{\partial x} - S^{n+1} \right) \right] w_j \Delta t \end{aligned} \quad [8.16]$$

Linearizing S with respect to U leads to:

$$S_i^{n+1} \approx S_i^n + \left(\frac{\partial S}{\partial U} \right)_i^{n+1/2} (U_i^{n+1} - U_i^n) \quad [8.17]$$

Reasoning along the same line as in section 8.1.2.2, substituting equation [8.17] into equation [8.16] yields:

$$\begin{aligned} &\sum_{i=1}^M (C_{i,j} + \theta D_{i,j} \lambda_i^{n+1/2} \Delta t) U_i^{n+1} - C_{i,j} \left(\frac{\partial S}{\partial U} \right)_i^{n+1/2} U_i^{n+1} \Delta t \\ &= \sum_{i=1}^M \left[C_{i,j} - (1-\theta) D_{i,j} \lambda_i^{n+1/2} \Delta t \right] U_i^n + C_{i,j} \left[S_i^n - \left(\frac{\partial S}{\partial U} \right)_i^{n+1/2} U_i^n \right] \Delta t \end{aligned} \quad [8.18]$$

where coefficients C_{ij} and D_{ij} are defined as in equation [8.13]. The term $(\partial S / \partial U)_i^{n+1/2}$ may be estimated explicitly from the known value at time level n . It may also be computed iteratively from a linear combination of the values at time levels n and $n + 1$.

Although the non-conservation form of the equation may seem easier to discretize because it leads to a linear system, it must be used with care with conservation laws with discontinuous solutions. As shown in section 8.4.2, the estimate of the wave speed $\lambda_i^{n+1/2}$ strongly influences the accuracy of the numerical solution.

8.1.3. Classical shape and test functions

8.1.3.1. Galerkin technique

In the Galerkin technique, the shape and weighting functions are taken from the same function space. The simplest possible option is to use $w_i = s_i$. In the case of piecewise linear functions (Figure 8.2a), the following expressions are used for s and w :

$$s_i(x) = w_i(x) = \begin{cases} 0 & \text{if } x \leq x_{i-1} \\ \frac{x - x_{i-1}}{x_i - x_{i-1}} & \text{if } x_{i-1} \leq x \leq x_i \\ \frac{x - x_{i+1}}{x_i - x_{i+1}} & \text{if } x_i \leq x \leq x_{i+1} \\ 0 & \text{if } x \geq x_{i+1} \end{cases} \quad [8.19]$$

The derivative $\partial s_i / \partial x$ is given by:

$$\frac{\partial s_i}{\partial x}(x) = \begin{cases} 0 & \text{if } x \leq x_{i-1} \\ \frac{1}{x_i - x_{i-1}} & \text{if } x_{i-1} \leq x \leq x_i \\ \frac{-1}{x_{i+1} - x_i} & \text{if } x_i \leq x \leq x_{i+1} \\ 0 & \text{if } x \geq x_{i+1} \end{cases} \quad [8.20]$$

Substituting equations [8.19–20] into equation [8.13] yields the following expression for C_{ij} :

$$C_{i,j} = \begin{cases} 0 & \text{if } j < i - 1 \\ (x_i - x_{i-1})/6 & \text{if } j = i - 1 \\ (x_{i+1} - x_{i-1})/3 & \text{if } j = i \\ (x_{i+1} - x_i)/6 & \text{if } j = i + 1 \\ 0 & \text{if } j > i + 1 \end{cases} \quad [8.21]$$

while D_{ij} is given by:

$$D_{i,j} = \begin{cases} 0 & \text{if } j < i - 1 \\ 1/2 & \text{if } j = i - 1 \\ 0 & \text{if } j = i \\ -1/2 & \text{if } j = i + 1 \\ 0 & \text{if } j > i + 1 \end{cases} \quad [8.22]$$

8.1.3.2. Petrov-Galerkin techniques

Petrov-Galerkin (PG) techniques use shape and weighting functions taken from distinct function spaces (see Figure 8.3 for examples). This allows the formulation to be upwinded by giving more weight to upstream nodes and less weight to downstream nodes (Figure 8.3a). This approach is similar to that of upwind finite difference schemes (see section 6.3).

An extreme case (Figure 8.3b) is obtained with a test function w_i that is one over the cell immediately upstream of node i and that is zero everywhere else.

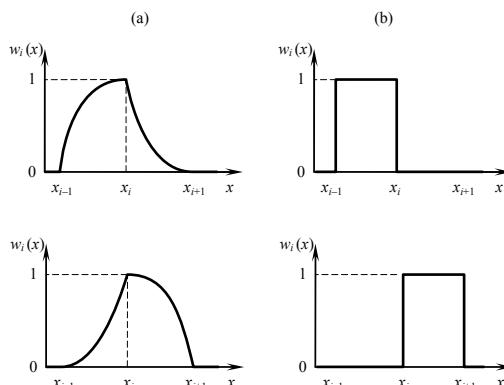


Figure 8.3. Petrov-Galerkin technique. Typical test functions. General case (a), particular case given by equations [8.23–24] (b). Top: function for a positive wave speed; bottom: functions for a negative wave speed

In this case, for a positive λ , w_i is defined as follows:

$$w_i(x) = \begin{cases} 0 & \text{for } x \leq x_{i-1} \\ 1 & \text{for } x_{i-1} < x \leq x_i \\ 0 & \text{for } x > x_i \end{cases} \quad [8.23]$$

while the following definition is obtained for a negative λ :

$$w_i(x) = \begin{cases} 0 & \text{for } x < x_i \\ 1 & \text{for } x_i \leq x < x_{i+1} \\ 0 & \text{for } x \geq x_{i+1} \end{cases} \quad [8.24]$$

Using equation [8.19] for the shape functions s_i leads to the following expressions for $C_{i,j}$ and $D_{i,j}$ (note that λ is assumed to be positive):

$$C_{i,j} = \begin{cases} 0 & \text{if } j < i \\ (x_i - x_{i-1})/2 & \text{if } j = i \\ (x_{i+1} - x_i)/2 & \text{if } j = i+1 \\ 0 & \text{if } j > i+1 \end{cases} \quad [8.25]$$

$$D_{i,j} = \begin{cases} 0 & \text{if } j < i \\ 1 & \text{if } j = i \\ -1 & \text{if } j = i+1 \\ 0 & \text{if } j > i+1 \end{cases}$$

8.1.3.3. A particular case: the SUPG approach

SUPG stands for Streamline Upwind Petrov-Galerkin. In the SUPG approach, the test function w_i is derived from the shape function s_i as:

$$w_i(x) = s_i(x) + a_i \lambda \frac{\partial s_i}{\partial x} \quad [8.26]$$

where the so-called stabilizing coefficient a_i is a function of the cell size ($a_i > 0$). With equation [8.26], the shape function is distorted by giving more weight to upstream nodes and less to downstream nodes. The consequence is scheme upwinding, with the expected result that solution monotony should be enhanced compared to the original Galerkin technique.

Equation [8.26] is applied to two types of functions hereafter:

1) For triangular shape functions s_i (Figure 8.4a), $\partial s_i / \partial x$ is piecewise constant. It is positive between nodes $i - 1$ and i , negative between nodes i and $i + 1$. A constant quantity, the sign of which is the same as λ , is added to s_i on the left-hand side of node i . In contrast, a constant quantity that has the sign of λ is subtracted from s_i on the right-hand side of the node. The resulting function w_i (Figure 8.4) is given by:

$$w_i(x) = \begin{cases} 0 & \text{for } x < x_{i-1} \\ \frac{x - x_{i-1}}{x_i - x_{i-1}} + \frac{a\lambda}{x_i - x_{i-1}} & \text{for } x_{i-1} < x < x_i \\ \frac{x - x_{i+1}}{x_{i+1} - x_i} - \frac{a\lambda}{x_{i+1} - x_i} & \text{for } x_i < x < x_{i+1} \\ 0 & \text{for } x > x_{i+1} \end{cases} \quad [8.27]$$

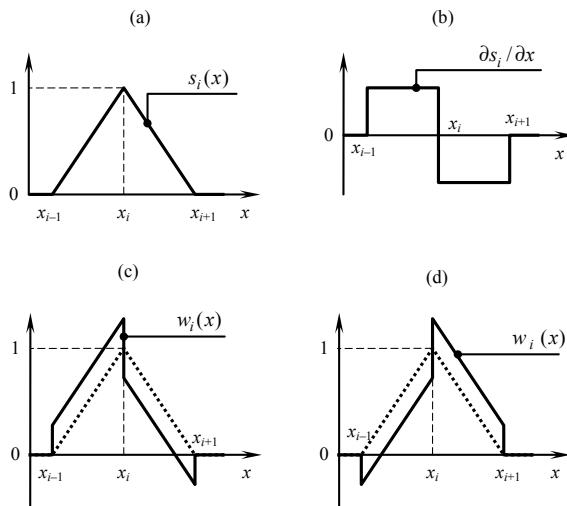


Figure 8.4. Applying the SUPG approach [8.26] to a triangular shape function. (a) shape function, (b) space derivative of the shape function, (c) SUPG test function for a positive λ , (d) SUPG test function for a negative λ

2) For parabolic shape functions s_i (Figure 8.4b), $\partial w_i / \partial x$ is linear and decreases linearly between nodes $i - 1$ and $i + 1$. It is positive at node $i - 1$, negative at node $i + 1$. The resulting SUPG test function is shown in Figure 8.5.

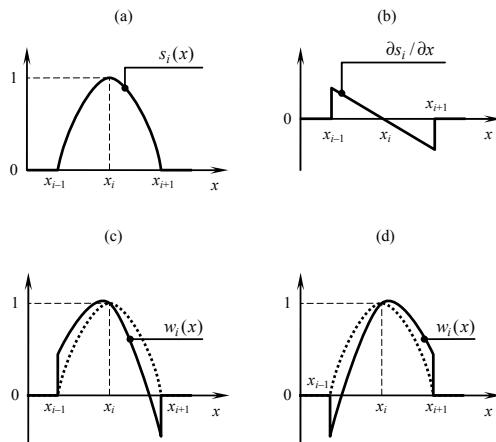


Figure 8.5. Applying the SUPG approach [8.26] to a parabolic shape function:
 (a) shape function, (b) space derivative of the shape function, (c) SUPG test function for a positive λ , (d) SUPG test function for a negative λ

It is easy to check that in the case of triangular shape functions, coefficients $C_{i,j}$ in equation [8.12] are given by:

$$C_{i,j} = \begin{cases} \frac{x_i - x_{i-1}}{6} - \frac{a\lambda}{2} & \text{if } j = i - 1 \\ \frac{x_{i+1} - x_{i-1}}{6} & \text{if } j = i \\ \frac{x_{i+1} - x_i}{6} + \frac{a\lambda}{2} & \text{if } j = i + 1 \\ 0 & \text{otherwise} \end{cases} \quad [8.28]$$

while coefficients $D_{i,j}$ are given by:

$$D_{i,j} = \begin{cases} \frac{1}{2} - \frac{a\lambda}{x_i - x_{i-1}} & \text{if } j = i - 1 \\ \frac{a\lambda}{x_i - x_{i-1}} + \frac{a\lambda}{x_{i+1} - x_i} & \text{if } j = i \\ -\frac{1}{2} - \frac{a\lambda}{x_{i+1} - x_i} & \text{if } j = i + 1 \\ 0 & \text{otherwise} \end{cases} \quad [8.29]$$

8.2. One-dimensional hyperbolic systems

8.2.1. Weak formulation

The conservation form [2.2] of hyperbolic systems of conservation laws is recalled:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} = \mathbf{S}$$

The weak form of equation [2.2] is:

$$\int_{t_1}^{t_2} \int_0^L \left(\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} - \mathbf{S} \right) w \, dx \, dt = 0 \quad [8.30]$$

The non-conservation form [2.5] is also recalled:

$$\frac{\partial \mathbf{U}}{\partial t} + \mathbf{A} \frac{\partial \mathbf{U}}{\partial x} = \mathbf{S}'$$

where \mathbf{A} is the Jacobian matrix of \mathbf{F} with respect to \mathbf{U} . The weak form of equation [2.5] is:

$$\int_{t_1}^{t_2} \int_0^L \left(\frac{\partial \mathbf{U}}{\partial t} + \mathbf{A} \frac{\partial \mathbf{U}}{\partial x} - \mathbf{S}' \right) w \, dx \, dt = 0 \quad [8.31]$$

Note that the characteristic form [2.24] may also be used:

$$\frac{\partial W_p}{\partial t} + \lambda^{(p)} \frac{\partial W_p}{\partial x} = S_p'', \quad p = 1, \dots, m$$

where W_p is the p th Riemann invariant. The weak form of equation [2.24] is:

$$\int_{t_1}^{t_2} \int_0^L \left(\frac{\partial W_p}{\partial t} + \lambda^{(p)} \frac{\partial W_p}{\partial x} - S_p'' \right) w \, dx \, dt = 0 \quad [8.32]$$

Each of these forms has advantages and drawbacks:

- The conservation form is more satisfactory from a theoretical point of view because it remains valid for discontinuous solutions. However, discretizing equation [8.30] for a nonlinear flux function \mathbf{F} leads to a nonlinear system that must

be solved for the (unknown) nodal values U_i^{n+1} (see section 8.1.2.2). To do so, iterative system inversion techniques must be used, with an increased computational effort.

– The non-conservation form leads to a system in nodal values U_i^{n+1} . However, its performance is highly sensitive to the estimate of the Jacobian matrix A when the solution is discontinuous (see section 8.4.2 for an illustration on a scalar case).

– The characteristic form [8.32] is easy to program because the Riemann invariants can be tracked independently of each other (see section 8.1). Moreover, upwind techniques such as the SUPG approach are easily programmed. However, the characteristic form is not adapted to problems involving shocks. In contrast, it is well-suited to the solution of linear conservation laws or hyperbolic systems such as the water hammer equations or the linear advection equation (see section 8.5.1).

8.2.2. Application to the non-conservation form

8.2.2.1. Galerkin technique

Reasoning as in section 8.1.2.3, equation [8.31] is transformed into:

$$\int_0^L [U^{n+1}(x) - U^n(x)] w_j \, dx + \int_0^{t^{n+1}} \int_{t^n}^L \left(A \frac{\partial U}{\partial x} - S' \right) w_j \, dt \, dx = 0 \quad [8.33]$$

The time integral of $(A \partial U / \partial x - S')w_j$ is approximated as:

$$\begin{aligned} \int_{t^n}^{t^{n+1}} \left(A \frac{\partial U}{\partial x} - S' \right) w_j \, dt &\approx \left[(1-\theta) \left(A^{n+1/2} \frac{\partial U^n}{\partial x} - S'^n \right) \right. \\ &\quad \left. + \theta \left(A^{n+1/2} \frac{\partial U^{n+1}}{\partial x} - S'^{n+1} \right) \right] \Delta t w_j \end{aligned} \quad [8.34]$$

Reasoning along the same line as in section 8.1.2.2, equation [8.33] is transformed into:

$$\begin{aligned} &\sum_{i=1}^M (C_{i,j} + \theta D_{i,j} A_i^{n+1/2} \Delta t) U_i^{n+1} - C_{i,j} \left(\frac{\partial S'}{\partial U} \right)_i^{n+1/2} U_i^{n+1} \Delta t \\ &= \sum_{i=1}^M [C_{i,j} - (1-\theta) D_{i,j} A_i^{n+1/2} \Delta t] U_i^n + C_{i,j} \left[S'^n_i - \left(\frac{\partial S'}{\partial U} \right)_i^{n+1/2} U_i^n \right] \Delta t \end{aligned} \quad [8.35]$$

where coefficients C_{ij} and D_{ij} are defined as in equation [8.13]. Note that $\partial S'/\partial U$ is the Jacobian matrix of S' with respect to U .

If shocks are present in the solution, the estimate of $A_i^{n+1/2}$ may exert a significant influence on the quality of the solution.

8.2.2.2. Petrov-Galerkin technique

Recall that the Petrov-Galerkin technique (including the SUPG approach) uses shape and test functions taken from distinct function spaces. In general, the test functions are asymmetric, which introduces upwinding. This is why Petrov-Galerkin techniques are often diffusive, thus allowing artificial oscillations near steep fronts to be eliminated from the solution.

However, upwinding can be applied only if a propagation direction can be identified. In the case of hyperbolic systems, however, the direction in which the waves propagate is not unique. Using fractional steps allows the hyperbolic systems to be broken into several terms. For some of them, a single propagation direction can be identified.

The technique is illustrated for the Saint Venant equations in a rectangular, prismatic channel. For the sake of clarity, the channel is considered frictionless and horizontal. The Saint Venant equations are simplified into the so-called one-dimensional shallow water equations:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} = 0 \\ U = \begin{bmatrix} h \\ q \end{bmatrix} \quad F = \begin{bmatrix} q \\ q^2/h + gh^2/2 \end{bmatrix} \quad [8.36]$$

The flux function F is broken into two parts F_1 and F_2 :

$$F_1 = \begin{bmatrix} q \\ q^2/h \end{bmatrix}, \quad F_2 = \begin{bmatrix} 0 \\ gh^2/2 \end{bmatrix} \quad [8.37]$$

Upwinding can be applied to the discretization of the flux F_1 . The time splitting technique [STR 68, GOU 77] is applied as follows:

- 1) In a first step, the following equation is solved:

$$\frac{\partial U}{\partial t} + \frac{\partial F_1}{\partial x} = 0 \quad [8.38]$$

Note that this equation can be written in non-conservation form as:

$$\frac{\partial U}{\partial t} + A_1 \frac{\partial U}{\partial x} = 0 \quad [8.39]$$

where matrix A_1 is given by:

$$A_1 = \begin{bmatrix} 0 & 1 \\ -u^2 & 2u \end{bmatrix} \quad [8.40]$$

This matrix has a double eigenvalue $\lambda^{(1)} = \lambda^{(2)} = u$. Consequently, equation [8.38] (that is only a part of the governing equation [8.36]) is characterized by the single propagation speed $\lambda = u$. This wave speed is used in equation [8.26] if a SUPG technique is to be applied.

2) The solution of equation [8.38] (or equation [8.39], depending on which form of the equation is to be solved) is used as an initial condition to solve the following equation over the computational time step:

$$\frac{\partial U}{\partial t} + \frac{\partial F_2}{\partial x} = 0 \quad [8.41]$$

Equation [8.41] may be solved in conservation form as:

$$\frac{\partial U}{\partial t} + A_2 \frac{\partial U}{\partial x} = 0 \quad [8.42]$$

where matrix A_2 is defined as:

$$A_2 = \begin{bmatrix} 0 & 0 \\ c^2 & 0 \end{bmatrix} \quad [8.43]$$

Steps 1) and 2) are repeated sequentially every time step.

This technique has the advantage that upwinding techniques (such as the SUPG technique) can be used in the first step of the time splitting procedure. Moreover, the approach is easily generalized to multidimensional problems.

8.3. Extension to multidimensional problems

8.3.1. Weak form of the equations

For the sake of clarity, only scalar two-dimensional problems are dealt with hereafter. The conservation form of two-dimensional scalar laws is given by equation [5.1], recalled here:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} + \frac{\partial G}{\partial y} = S$$

where F and G are respectively the fluxes in the x and y directions. The non-conservation form [5.2] is also recalled:

$$\frac{\partial U}{\partial t} + \lambda_x \frac{\partial U}{\partial x} + \lambda_y \frac{\partial U}{\partial y} = S'$$

where λ_x , λ_y and S' are given by equation [5.3], also recalled hereafter:

$$\left. \begin{aligned} \lambda_x &= \frac{\partial F}{\partial U} \\ \lambda_y &= \frac{\partial G}{\partial U} \\ S' &= S - \left(\frac{\partial F}{\partial x} \right)_{U=\text{Const}} - \left(\frac{\partial G}{\partial y} \right)_{U=\text{Const}} \end{aligned} \right\}$$

In practice, [5.1] or [5.2] is solved over a finite two-dimensional domain Ω . As in section 8.1, the weak form of the governing equations is solved. The weak form of the conservation form [5.1] is:

$$\int_{t^n}^{t^{n+1}} \int_{\Omega} \left(\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} + \frac{\partial G}{\partial y} - S \right) w \, d\Omega \, dt \quad [8.44]$$

while the weak form of the non-conservation form [5.2] is given by:

$$\int_{t^n}^{t^{n+1}} \int_{\Omega} \left(\frac{\partial U}{\partial t} + \lambda_x \frac{\partial F}{\partial x} + \lambda_y \frac{\partial G}{\partial y} - S' \right) w \, d\Omega \, dt \quad [8.45]$$

As in the case of one-dimensional problems, the solution is sought as a linear combination of pre-defined shape functions s_i over the domain Ω . The weighting functions w_i are also defined over the domain Ω . Space discretization aspects are covered in section 8.3.2. Classical shape and weighting functions are described in section 8.3.3.

8.3.2. Discretization of space

In finite element techniques, space is discretized into elements formed by the nodes. These elements form an unstructured grid. The elements are classically triangular or quadrangular (Figure 8.6). Figure 8.6 illustrates the meshing of a two-dimensional domain Ω . The sketch on the left-hand side of the figure shows the computational grid obtained using only triangular elements. The right-hand side sketch in the Figure shows the computational mesh obtained by allowing triangular elements to merge into quadrangular elements. The triangular and triangular-quadrangular meshes have respectively 362 and 211 elements.

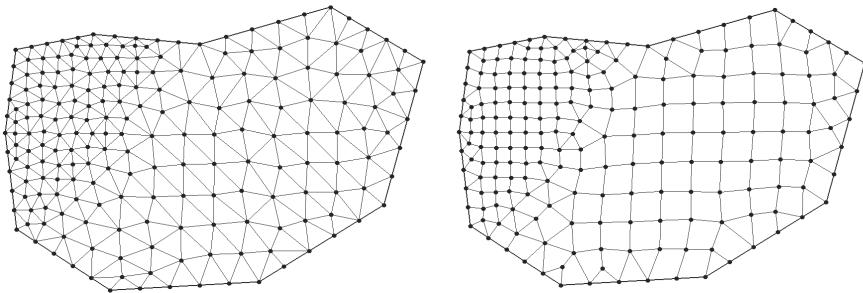


Figure 8.6. Two examples of finite element meshes for a two-dimensional domain Ω .
 Left: purely triangular grid (362 elements); right: mixed triangular-quadrangular elements (211 elements)

8.3.3. Classical shape and test functions

The shape functions used for multidimensional finite element methods are classically defined such that s_i is equal to one at the node i and takes a zero value at

all other nodes. Figure 8.7 illustrates the example of a piecewise linear function defined over a triangular grid.

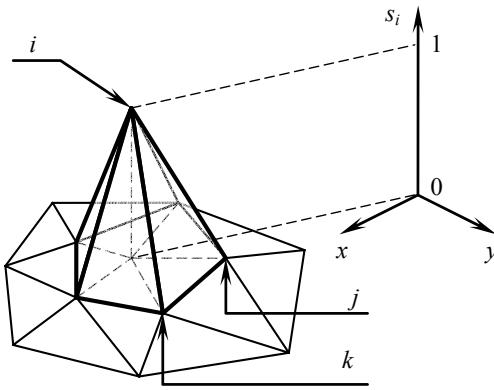


Figure 8.7. Piecewise linear shape function over a two-dimensional triangular grid.
The shape function s_i is equal to 1 at node i and is 0 at all other nodes

For a triangular element (i, j, k) to which the node i belongs, the shape function s_i is defined as:

$$s_i(x, y) = 1 + E_{i,j,k}(x - x_i) + F_{i,j,k}(y - y_i) \quad [8.46]$$

where the slopes $E_{i,j,k}$ and $F_{i,j,k}$ are given by:

$$\left. \begin{aligned} E_{i,j,k} &= \frac{x_i - x_j}{(x_j - x_i)^2 + (y_j - y_i)^2} + \frac{x_i - x_k}{(x_k - x_i)^2 + (y_k - y_i)^2} \\ F_{i,j,k} &= \frac{y_i - y_j}{(x_j - x_i)^2 + (y_j - y_i)^2} + \frac{y_i - y_k}{(x_k - x_i)^2 + (y_k - y_i)^2} \end{aligned} \right\} \quad [8.47]$$

The function s_i is zero over any element of which node i is not a corner node.

Galerkin's technique uses test functions w_i that are identical to the shape functions s_i . In the SUPG technique, the test function w_i is obtained from s_i by generalizing formula [8.26] to multiple dimensions as:

$$w_i = s_i + a_i \vec{\lambda} \cdot \overrightarrow{\text{Grads}}_i \quad [8.48]$$

where the vector $\vec{\lambda}$ is formed by the components λ_x , λ_y and $\overrightarrow{\text{Grad}}$ denotes the gradient operator. Applying equation [8.48] to equation [8.46] yields:

$$w_i(x, y) = 1 + \lambda_x E_{i,j,k} + \lambda_y F_{i,j,k} + E_{i,j,k}(x - x_i) + F_{i,j,k}(y - y_i) \quad [8.49]$$

As in the one-dimensional case, the test function w_i gives an increased weight to the nodes upstream of node i in the discretized equation, while the weight of the nodes downstream of i is reduced.

8.4. Discontinuous Galerkin techniques

8.4.1. Principle of the method

The classical finite element techniques presented in sections 8.1 to 8.3 solve the weak forms of conservation laws. However, due to the non-uniqueness of weak solutions, conservation is not guaranteed in the general case. This is illustrated by the application examples presented in section 8.5, where the solution of simple, non-linear scalar laws such as the inviscid Burgers equation is shown to give erroneous shock propagation speeds.

Discontinuous Galerkin (DG) techniques allow conservation to be guaranteed via a slight modification of the weak form of the governing equations. They can be seen as a combination of finite volume and finite element techniques that retains the advantages of both methods. They were applied to hyperbolic conservation laws in [COC 89b], to one-dimensional hyperbolic systems in [COC 89a], multidimensional conservation laws in [COC 90], multidimensional systems in [COC 98] and convection problems in [COC 01]. Applications to the shallow water equations can be found in [DAW 02] and [KES 09], with an extension to two-dimensional transport in [AIZ 02, AIZ 03]. An application to morphological modeling can be found in [KUB 06]. An improved slope limitation for one- and two-dimensional systems can be found in [GHO 09]. The method is presented for one-dimensional systems hereafter.

Assume that a hyperbolic system in conservation form [2.2], recalled here:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} = \mathbf{S}$$

is to be solved over the domain $\Omega = [0, L]$. As in the previous sections, equation [2.2] is multiplied by a test function w and integrated over the solution domain Ω :

$$\int_0^L \frac{\partial U}{\partial t} w dx + \int_0^L \frac{\partial F}{\partial x} w dx = \int_0^L S w dx \quad [8.50]$$

Using integration by parts in the middle integral, equation [8.50] becomes:

$$\int_0^L \frac{\partial U}{\partial t} w dx + [wF]_0^L - \int_0^L \frac{\partial w}{\partial x} F dx = \int_0^L S w dx \quad [8.51]$$

As in finite volume techniques, the domain Ω is discretized into M computational cells. Writing equation [8.51] for the cell i leads to:

$$\int_{x_{i-1/2}}^{x_{i+1/2}} \frac{\partial U}{\partial t} w dx + [wF]_{x_{i-1/2}}^{x_{i+1/2}} - \int_{x_{i-1/2}}^{x_{i+1/2}} \frac{\partial w}{\partial x} F dx = \int_{x_{i-1/2}}^{x_{i+1/2}} S w dx \quad [8.52]$$

where the square brackets denote the variation of the function between the brackets, $[f]_a^b = f(b) - f(a)$. The solution U in the cell i is sought in the form:

$$\tilde{U}_i^n(x) = \sum_{p=0}^P (U_p)_i^n s_p(x) \quad [8.53]$$

where the $s_p(x)$ are shape functions and $(U_p)_i^n$ is a constant vector. P is fixed arbitrarily. Classically, the shape functions are polynomials in x . In this case, the method is said to be of order $P + 1$ (that is, $P = 0$ yields a first-order method, $P = 1$ yields a second-order method, etc.). Substituting equation [8.53] into equation [8.52] and using a first-order approximation for the time derivative leads to:

$$\sum_{p=1}^P \int_{x_{i-1/2}}^{x_{i+1/2}} s_p w dx \frac{(U_p)_i^{n+1} - (U_p)_i^n}{\Delta t} = -[wF]_{x_{i-1/2}}^{x_{i+1/2}} + \int_{x_{i-1/2}}^{x_{i+1/2}} \left(\frac{\partial w}{\partial x} F + S w \right) dx \quad [8.54]$$

The solution is known completely over the cell i if all the coefficients $(U_p)_i^n$, $p = 1, \dots, P$ can be computed. Therefore, it is necessary to define P different weighting functions w over the cell i in order to form a $P \times P$ system that can be solved uniquely for the coefficients $(U_p)_i^n$. In the Galerkin approach, the test functions are identical to the shape functions. Consequently, equation [8.54] is rewritten successively for $w = s_0, w = s_1, \dots, w = s_P$. The following system is obtained:

$$\sum_{p=1}^P R_{p,j} (\mathbf{U}_p)_i^{n+1} = \sum_{p=1}^P R_{p,j} (\mathbf{U}_p)_i^n - \Delta t [s_j F]_{x_{i-1/2}}^{x_{i+1/2}} + \Delta t \int_{x_{i-1/2}}^{x_{i+1/2}} \left(\frac{\partial s_j}{\partial x} F + s_j S \right) dx, \quad j = 1, \dots, P \quad [8.55]$$

with

$$R_{p,j} = R_{j,p} = \int_{x_{i-1/2}}^{x_{i+1/2}} s_p(x) s_j(x) dx \quad [8.56]$$

As in Godunov-type methods (see Chapter 7), the flux F at each interface $x_{i-1/2}$ is computed by solving a Riemann problem. The left and right states of the Riemann problem are defined from the reconstructions [8.53] over the cells $i-1$ and i :

$$\begin{cases} \mathbf{U}_L = \tilde{\mathbf{U}}_{i-1}^n(x_{i-1/2}) \\ \mathbf{U}_R = \tilde{\mathbf{U}}_i^n(x_{i-1/2}) \end{cases} \quad [8.57]$$

The variation $[w_j F]$ and the two integrals on the right-hand side of equation [8.55] must be estimated. Their estimate, as well as the computation of the coefficients $R_{p,j}$, is examined in the following section.

8.4.2. Legendre polynomial-based reconstruction

Computation of the various terms in equations [8.55–8.56] is simplified if reconstruction [8.53] uses Legendre polynomials:

$$s_p(x) = \frac{1}{2^p} \sum_{k=0}^p \binom{p}{k}^2 \left(2 \frac{x - x_i}{\Delta x_i} - 1 \right)^{p-k} \left(2 \frac{x - x_i}{\Delta x_i} + 1 \right)^k \quad [8.58]$$

where x_i is the abscissa of the center of the cell i . For $p = 0, 1, 2$, we have:

$$s_0(x) = 1, \quad s_1(x) = 2 \frac{x - x_i}{\Delta x_i}, \quad s_2(x) = 6 \left(\frac{x - x_i}{\Delta x_i} \right)^2 - \frac{1}{2} \quad [8.59]$$

Polynomials [8.58] have a number of interesting properties:

- Property (P8.1). The Legendre polynomials form a family of orthogonal functions over the interval $[x_{i-1/2}, x_{i+1/2}]$. Consequently, $R_{j,p} = R_{p,j}$ is non-zero only if $j = p$.
- Property (P8.2). The polynomial takes particular values $s_p(x_{i-1/2}) = (-1)^p$ and $s_p(x_{i+1/2}) = 1$.
- Property (P8.3). The integral of $s_p(x)$ over the cell i is zero for all p , except for $p = 0$. This property can be seen as a direct consequence of property (P8.1) by noting that the integral of $s_p(x)$ over the cell i is equal to $R_{0,p}$ by definition.

As a direct consequence of property (P8.3), the integral of the reconstructed profile over the cell i is given by:

$$\int_{x_{i-1/2}}^{x_{i+1/2}} U(x, t^n) dx = \sum_{p=0}^P (U_p)_i^n \int_{x_{i-1/2}}^{x_{i+1/2}} s_p(x) dx = \Delta x (U_0)_i^n \quad [8.60]$$

Using properties (P8.1) and (P8.2), equation [8.55] is rewritten as:

$$\begin{aligned} (U_j)_i^{n+1} &= (U_j)_i^n + \frac{\Delta t}{R_{j,j}} \left[(-1)^j F_{i-1/2} - F_{i+1/2} \right] \\ &\quad + \frac{\Delta t}{R_{j,j}} \int_{x_{i-1/2}}^{x_{i+1/2}} \left(\frac{\partial s_j}{\partial x} F + s_j S \right) dx, \quad j = 1, \dots, P \end{aligned} \quad [8.61]$$

The integral in equation [8.61] may be estimated using for instance a second-order approximation (but other quadrature rules may be applied, see e.g. [KES 09]):

$$\int_{x_{i-1/2}}^{x_{i+1/2}} f(x) dx \approx \frac{f(x_{i-1/2}) + 4f(x_i) + f(x_{i+1/2})}{6} \Delta x_i \quad [8.62]$$

where f is defined as $f = \partial s_j / \partial x F + s_j S$. If an explicit formulation is chosen, f is easily determined at $x_{i-1/2}$, x_i and $x_{i+1/2}$ from the reconstructed profile of U at the known time level n . Indeed, from property (P8.2):

$$\left. \begin{aligned} U_i(x_{i-1/2}, t^n) &= \sum_{p=1}^P (-1)^p (U_p)_i^n \\ U_i(x_{i+1/2}, t^n) &= \sum_{p=1}^P (U_p)_i^n \end{aligned} \right\} [8.63]$$

The second-order DG method is obtained for $P = 1$. Writing equation [8.61] for $p = 0$ and $p = 1$ gives the following two formulae:

$$\left. \begin{aligned} (U_0)_i^{n+1} &= (U_0)_i^n + \frac{\Delta t}{\Delta x_i} \left(F_{i-1/2} - F_{i+1/2} + \int_{x_{i-1/2}}^{x_{i+1/2}} S \, dx \right) \\ (U_1)_i^{n+1} &= (U_1)_i^n + 3 \frac{\Delta t}{\Delta x_i} \left[-F_{i-1/2} - F_{i+1/2} + \int_{x_{i-1/2}}^{x_{i+1/2}} \left(\frac{2F}{\Delta x_i} + s_1 S \right) dx \right] \end{aligned} \right\} [8.64]$$

where the integrals are approximated using e.g. equation [8.62]. Note that equations [8.64] are obtained using the following equalities:

$$\frac{\partial s_0}{\partial x} = 1, R_{0,0} = \Delta x_i, \frac{\partial s_1}{\partial x} = \frac{2}{\Delta x_i}, R_{1,1} = \frac{\Delta x_i}{3} [8.65]$$

Remembering from equation [8.60] that $(U_0)_i^n$ represents the average value of U over the cell i at time level n , the first equation [8.64] is formally equivalent to the finite volume equation [7.3]. In other words, the DG technique retains the conservation properties of finite volume methods.

8.4.3. Limiting

DG methods of arbitrary high order may produce oscillations in the computed variables near steep fronts. If the flux function F is nonlinear, the oscillatory character of the solution may lead to nonlinear instability. This can be avoided by limiting the variations in the reconstructed profiles so as to avoid under- or overshooting.

The coefficients $(U_p)_i^n$ must be limited component by component. At the end of the limiting process, each component of the reconstructed edge value $\tilde{U}_i^n(x_{i-1/2})$ should lie between the corresponding components of $(U_0)_{i-1}^n$ and $(U_0)_i^n$. Conversely, each component of the edge value $\tilde{U}_i^n(x_{i+1/2})$ should lie between the corresponding components of $(U_0)_i^n$ and $(U_0)_{i+1}^n$.

If the cell i is a local extremum for a given component k , the profile for the k th component of the reconstructed profile is taken constant over the cell, that is, the k th component of $\tilde{U}_i^n(x)$ is set to the value of the k th component of the vector $(U_0)_i^n$. U_p must be limited as [KRI 07]:

$$\left[(U_p)_i^n \right]_{\text{lim}} = \text{minmod} \left[(U_p)_i^n, \frac{(U_{p-1})_i^n - (U_{p-1})_{i-1}^n}{2p-1}, \frac{(U_{p-1})_{i+1}^n - (U_{p-1})_i^n}{2p-1} \right]$$

[8.66]

where the minmod function of two arguments a and b is zero if a and b have opposite signs, and retains the argument that has the smaller modulus if $ab > 0$:

$$\begin{aligned} \text{minmod}(a, b) &= \frac{\text{sgn}(a) + \text{sgn}(b)}{2} \min(|a|, |b|) \\ \text{minmod}(a, b, c) &= \text{minmod}[\text{minmod}(a, b), c] \end{aligned}$$

[8.67]

Equation [8.66] is to be applied component-wise by descending values of p , from $p = P$ to $p = 1$. The limiting is stopped when the first value of p such that:

$$\left[(U_p)_i^n \right]_{\text{lim}} = (U_p)_i^n$$

[8.68]

is reached. For a second-order DG technique, equation [8.66] simplifies to:

$$\left[(U_1)_i^n \right]_{\text{lim}} = \text{min mod} \left[(U_1)_i^n, (U_0)_i^n - (U_0)_{i-1}^n, (U_0)_{i+1}^n - (U_0)_i^n \right]$$

[8.69]

8.4.4. Runge-Kutta time stepping

Runge-Kutta Discontinuous Galerkin (RKDG) techniques are obtained as generalizations of the explicit method presented in the previous sections. In fact, equations [8.55] and its particular expressions [8.61] and [8.64] are provided for a first-order approximation of the time derivative. They can be recast in the form [6.73], recalled here:

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i^n + \Delta t \mathbf{M} \mathbf{U}_i^n$$

where \mathbf{M} is a matrix operator as defined in section 6.5.2. Runge-Kutta time stepping may be applied as in equation [6.76], recalled here:

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i^n + \sum_{k=1}^M \frac{\Delta t^k}{k!} \mathbf{M}^k \mathbf{U}_i^n$$

In the particular case of a second-order Runge-Kutta (RK2) stepping, the formula simplifies into:

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i^n + \Delta t \mathbf{M} \mathbf{U}_i^n + \frac{\Delta t^2}{2} \mathbf{M}^2 \mathbf{U}_i^n = \mathbf{U}_i^n + \Delta t \mathbf{M} \left(\mathbf{U}_i^n + \frac{\Delta t}{2} \mathbf{M} \mathbf{U}_i^n \right) \quad [8.70]$$

Noting that the quantity $\mathbf{U}_i^{n+1} = \mathbf{U}_i^n + \Delta t / 2 \mathbf{M} \mathbf{U}_i^n$ is obtained by applying the numerical method [6.73] over half a time step, equation [8.70] can be translated into the algorithmic form as follows:

– step 1: use the DG technique over half a time step to compute an intermediate value $\mathbf{U}_i^{n+1/2}$ at the intermediate time level $n + 1/2$:

$$\mathbf{U}_i^{n+1/2} = \mathbf{U}_i^n + \frac{\Delta t}{2} \mathbf{M} \mathbf{U}_i^n \quad [8.71]$$

– step 2: use the intermediate value $\mathbf{U}_i^{n+1/2}$ to update the estimate of the operator \mathbf{M} and proceed to the next time level:

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i^n + \Delta t \mathbf{M} \mathbf{U}_i^{n+1/2} \quad [8.72]$$

The second-order RKDG method is stable for a Courant number smaller than 1/3. The third-order RKDG method is stable for $|Cr|$ smaller than 0.209, while the fourth-order RKDG method is stable for $|Cr|$ smaller than 0.145 [COC 01].

8.5. Application examples

8.5.1. The linear advection equation

8.5.1.1. Discretized equation

In this section, the linear advection equation is solved using the Galerkin technique (equations [8.21–22]), the Petrov-Galerkin approach with piecewise constant test functions (equation [8.25]), the SUPG technique (equations [8.28–29]) and the DG technique (equations [8.64] with limiting [8.66–67] and RK2 time stepping [8.71–72]). The conservation and non-conservation forms of the linear advection equation are recalled:

$$\begin{aligned} \frac{\partial U}{\partial t} + \frac{\partial}{\partial x}(\lambda U) &= 0 && \text{(conservation form)} \\ \frac{\partial U}{\partial t} + \lambda \frac{\partial U}{\partial x} &= 0 && \text{(non - conservation form)} \end{aligned} \quad [8.73]$$

For the sake of simplicity, λ is assumed to be constant and positive over the solution domain. The element size Δx is taken as constant. Noticing that $C_{i,j}$ and $D_{i,j}$ are non-zero only for $j = i - 1, j = i$ or $j = i + 1$, equation [8.18] simplifies to:

$$\begin{aligned} (C_{j-1,j} + \theta D_{j-i,j} \lambda \Delta t) U_{j-1}^{n+1} + (C_{j,j} + \theta D_{j,j} \lambda \Delta t) U_j^{n+1} \\ + (C_{j+1,j} + \theta D_{j+i,j} \lambda \Delta t) U_{j+1}^{n+1} = [C_{j-1,j} - (1-\theta) D_{j-i,j} \lambda \Delta t] U_{j-1}^n \\ + [C_{j,j} - (1-\theta) D_{j,j} \lambda \Delta t] U_j^n + [C_{j+1,j} - (1-\theta) D_{j+i,j} \lambda \Delta t] U_{j+1}^n \end{aligned} \quad [8.74]$$

Substituting equations [8.21–22] into equation [8.74] and dividing by Δx leads to a system in the form:

$$b U_{j-1}^{n+1} + c U_j^{n+1} + d U_{j+1}^{n+1} = e_i^n \quad [8.75]$$

where coefficients b, c, d and e are functions of the technique used.

– The SUPG approach with triangular shape functions leads to:

$$\left. \begin{aligned} b &= \left[\frac{1}{6} + \frac{a\lambda}{2\Delta x} - \theta \left(\frac{1}{2} + \frac{a\lambda}{\Delta x} \right) Cr \right] \\ c &= \left(\frac{2}{3} + \theta Cr \right) \\ d &= \left[\frac{1}{6} - \frac{a\lambda}{2\Delta x} + \theta \left(\frac{1}{2} - \frac{a\lambda}{\Delta x} \right) Cr \right] \\ e_i^n &= \left[\frac{1}{6} + \frac{a\lambda}{2\Delta x} + (1-\theta) \left(\frac{1}{2} + \frac{a\lambda}{\Delta x} \right) Cr \right] U_{j-1}^n + \left(\frac{2}{3} - \theta Cr \right) U_j^n \\ &\quad + \left[\frac{1}{6} - \frac{a\lambda}{2\Delta x} - \theta \left(\frac{1}{2} - \frac{a\lambda}{\Delta x} \right) Cr \right] U_{j-1}^n \end{aligned} \right\} [8.76]$$

where $Cr = \lambda \Delta t / \Delta x$ is the Courant number and a is the stabilizing coefficient in equation [8.26]. Note that the Galerkin discretization is recovered with $a = 0$.

– The Petrov-Galerkin approach [8.25] leads to:

$$\left. \begin{aligned} b &= \frac{1}{2} - \theta Cr \\ c &= \frac{1}{2} + \theta Cr \\ d &= 0 \\ e_i^n &= \left[\frac{1}{2} + (1-\theta) Cr \right] U_{j-1}^n + \left[\frac{1}{2} - (1-\theta) Cr \right] U_j^n \end{aligned} \right\} [8.77]$$

– The second-order DG technique [8.64, 8.66–67, 8.71–72] is simplified by noting that the solution of the Riemann problem at the interface $i - 1/2$ for a positive λ is equal to the reconstructed value $\tilde{U}_{i-1}^n(x_{i-1/2})$. Consequently:

$$\left. \begin{aligned} F_{i-1/2} &= \lambda \tilde{U}_{i-1}^n(x_{i-1/2}) = \left[(U_0)_{i-1}^n + (U_1)_{i-1}^n \right] \lambda \\ F_{i+1/2} &= \lambda \tilde{U}_i^n(x_{i+1/2}) = \left[(U_0)_i^n + (U_1)_i^n \right] \lambda \end{aligned} \right\} [8.78]$$

Moreover, we have:

$$\int_{x_{i-1/2}}^{x_{i+1/2}} \frac{2F}{\Delta x_i} dx = 2\lambda(U_0)_i^n \quad [8.79]$$

Substituting equations [8.78] and [8.79] into equations [8.64] leads to:

$$\left. \begin{aligned} (U_0)_i^{n+1} &= (U_0)_i^n + Cr[(U_0)_{i-1} + (U_1)_{i-1} - (U_0)_i - (U_1)_i] \\ (U_1)_i^{n+1} &= (U_1)_i^n + 3Cr[-(U_0)_{i-1} + (U_0)_i - (U_1)_{i-1} - (U_1)_i] \end{aligned} \right\} \quad [8.80]$$

NOTE.— Equation [8.75] may be written only at internal nodes ($i = 2, \dots, M-1$). However, $(U_1)_M^{n+1}$ is known from the upstream boundary condition, which adds an equation to the system. The missing equation for the downstream node ($i = M$) is supplied by applying the Petrov-Galerkin discretization [8.77].

8.5.1.2. Test case description and results

The following problem is solved: a constant value U_b is applied at the upstream end ($x = 0$) of the computational domain of length L . The initial value U_0 of U in the domain is uniform. The parameters of the test case are shown in Table 8.1, the computational results are illustrated by Figures 8.8 to 8.11. Note that the combination of λ , Δt and Δx adopted in this test yields a Courant number $Cr = 2$.

Recall that the Galerkin and Petrov-Galerkin techniques presented in this chapter are semi-implicit and that Courant numbers larger than unity can be handled without inducing stability problems.

Symbol	Meaning	Value
U_0	Initial value of U	0
U_b	Upstream boundary condition	1
L	Length of the domain	100 m
T	Simulated time	25 s
Δt	Computational time step	1 s
Δx	Cell size	1 m
λ	Wave speed	2.0 m/s
θ	Scheme implicitation parameter (Galerkin, Petrov-Galerkin and SUPG schemes)	0.50, 0.55, 0.65 and 1.00
ρ	SUPG's $a\lambda/\Delta x$ parameter	1

Table 8.1. Linear advection. Parameters of the test case

As shown in Figure 8.8, Galerkin's technique with triangular functions yields oscillations behind the computed front when θ is chosen close to 0.5. Increasing θ to 0.65 allows the oscillations to be eliminated almost completely.

For $\theta=1.0$, the computed profile is monotonic, but the front is smeared over almost 15 cells. This was to be expected because increasing θ yields an increased numerical diffusion, the effect of which is to damp the shorter wavelengths that are responsible for the oscillations in the profile. The longer wavelengths are preserved, hence the smearing of the front.

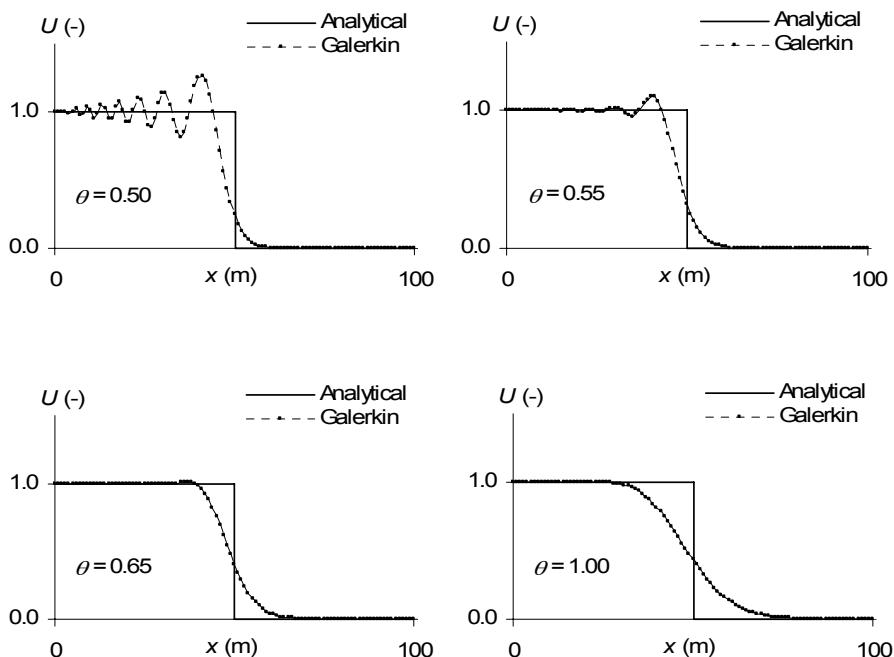


Figure 8.8. Linear advection equation solved using Galerkin's technique.
Numerical and analytical solutions at $t = 25$ s

The Petrov-Galerkin technique [8.18, 8.25] gives similar results (Figure 8.9), except that the oscillations take place over a shorter distance than in the case of the Galerkin technique. They are almost damped for $\theta=0.65$. The increased monotonic character of the Petrov-Galerkin technique is due to the stronger upwinding brought about by the piecewise constant test functions.

The SUPG technique (Figure 8.10) yields an increased damping of the oscillations compared to the Galerkin and Petrov-Galerkin discretizations. The oscillations behind the front are damped within a shorter distance than with the Petrov-Galerkin technique.

Owing to stability constraints, the Runge-Kutta discontinuous Galerkin technique is run with a much smaller time step than the first three methods. The time step chosen for the present application is $\Delta t = 0.125$ s, which corresponds to a Courant number $Cr = 0.25$ (remember that the stability limit is $Cr = 1/3$).

The profile computed at $T = 25$ seconds is compared with the analytical solution in Figure 8.11. Two numerical profiles are shown: the profile obtained without limiter and the profile obtained when limiting is applied. The profile obtained without limiting exhibits oscillations ahead of the front. In contrast, the limited profile is monotonic.

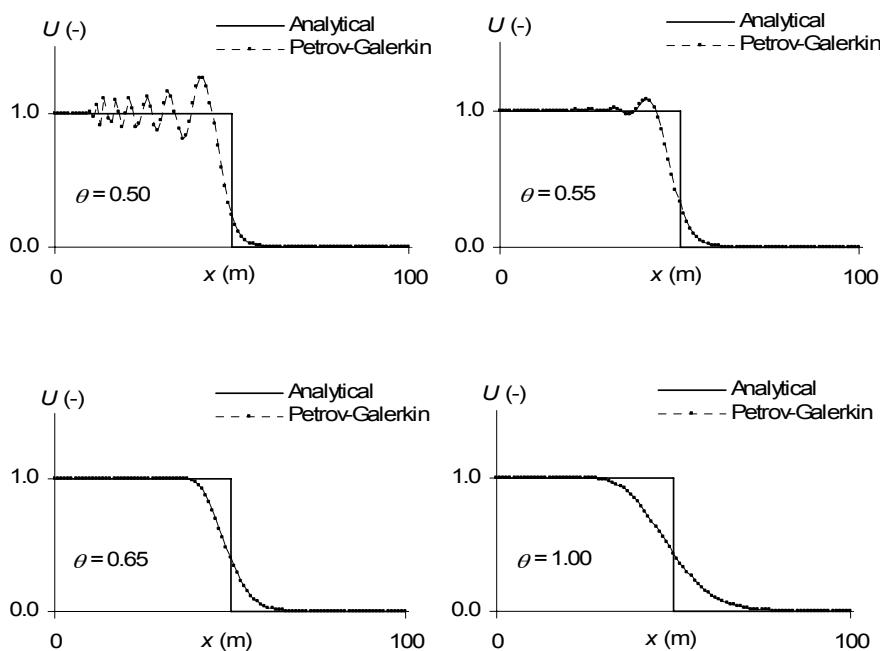


Figure 8.9. Linear advection equation solved using the Petrov-Galerkin technique with piecewise constant weighting functions. Numerical and analytical solutions at $t = 25$ s

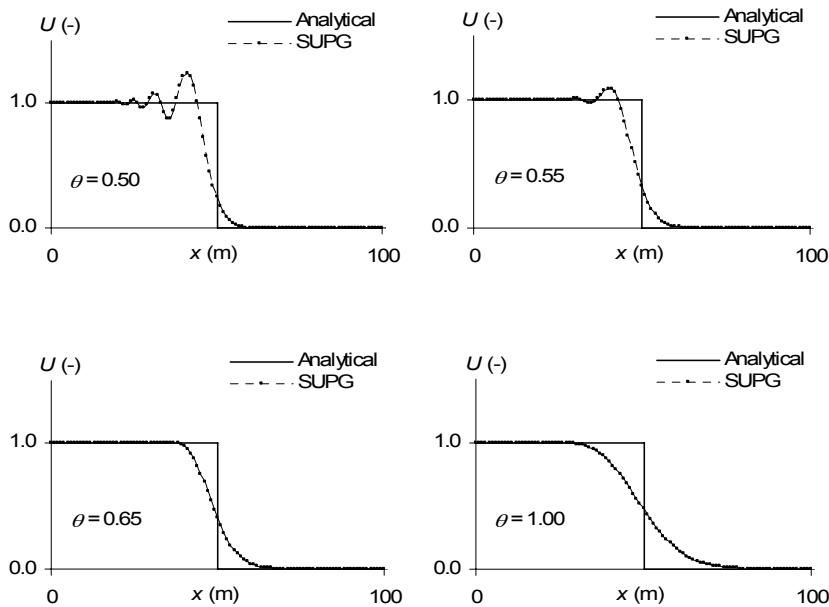


Figure 8.10. Linear advection equation solved using the SUPG technique.
Numerical and analytical solutions at $t = 25$ s

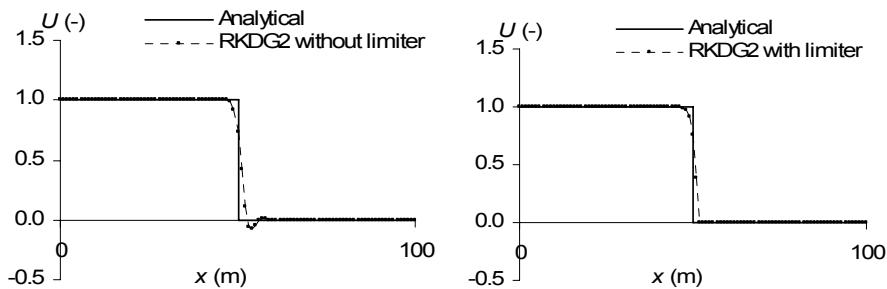


Figure 8.11. Linear advection equation solved using the RKDG2 technique.
Numerical and analytical solutions at $t = 25$ s

8.5.2. The inviscid Burgers equation

8.5.2.1. Solution by classical Galerkin techniques: explicit estimate of C_r

The purpose is to solve the non-conservation form of the inviscid Burgers equation [1.66], recalled here:

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = 0$$

using the Galerkin, Petrov-Galerkin and SUPG techniques. The application example is limited to positive values of u .

Equations [8.76–77] may be used to compute the solution, provided that the wave propagation speed λ is replaced with u in the calculation of the coefficients b to e . The question then arises of how u , that is not uniform over the solution domain, should be estimated. In what follows, the following space weighting, inspired from Preissmann's scheme (see section 6.4.1), is used:

$$Cr = \left[(1 - \psi) u_{i-1}^n + \psi u_i^n \right] \frac{\Delta t}{\Delta x} \quad [8.81]$$

Note that this expression is valid only for positive values of u and should be used in the computation of the coefficients for the node i .

The parameters used in the simulation are shown in Table 8.2.

Symbol	Meaning	Value
u_0	Initial condition	1 m/s
u_b	Prescribed velocity at the left-hand boundary	2 m/s
L	Domain length	100 m
T	Simulated time	50 s
Δt	Computational time step	1 s
Δx	Cell width	1 m
θ	Scheme implicitation parameter (Galerkin, Petrov-Galerkin and SUPG schemes)	0.50, 0.55, 0.65 and 1.00
ρ	Value of $a\lambda/\Delta x$ (SUPG scheme)	1.
ψ	Centering parameter for Cr in equation [8.81]	0.0, 0.5 and 1.0

Table 8.2. The inviscid Burgers equation. Parameters of the test case

The performance of the various schemes is illustrated in Figures 8.12 to 8.14. In these figures, the centering parameter in equation [8.81] is set to $\psi = 1/2$.

The solution computed by Galerkin's technique (Figure 8.12) exhibits strong oscillations for values of θ close to 0.5. We can check that for $\theta = 0.5$, the solution becomes unstable. Increasing θ allows the oscillations to be reduced and the shock speed to be better estimated. However, even the extreme value $\theta = 1.0$ does not allow the correct shock speed to be recovered in the numerical solution.

The Petrov-Galerkin method with a piecewise constant test function leads to a much smoother profile than the Galerkin technique (Figure 8.13). The shock speed is also better estimated. This, however, is achieved at the expense of a strongly smeared front.

The SUPG technique (Figure 8.14) combines the advantages of the previous two methods, with reduced oscillations compared to the Galerkin technique and a steeper front than in the Petrov-Galerkin technique. The drawback is a strongly underestimated shock speed in the numerical profile.

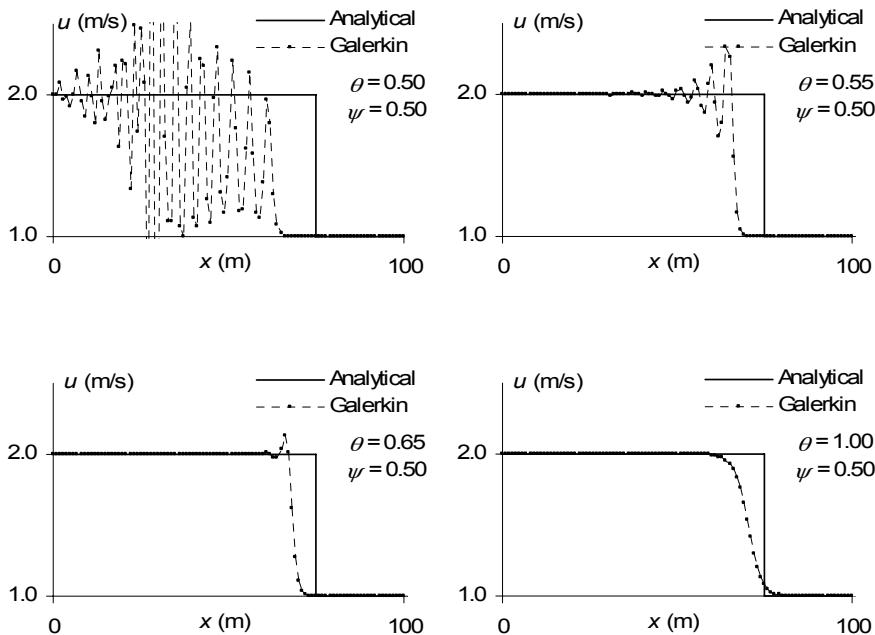


Figure 8.12. The inviscid Burgers equation. Solution using Galerkin's technique at $t = 50$ s

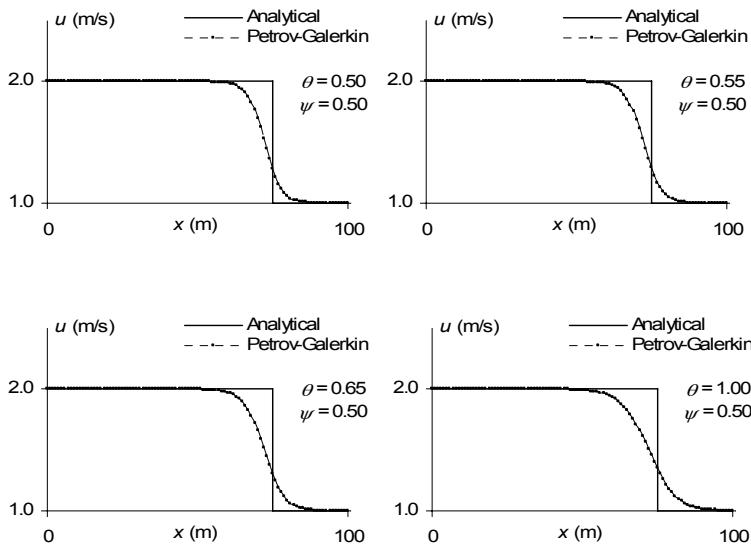


Figure 8.13. The inviscid Burgers equation. Solution using the Petrov-Galerkin technique at $t = 50$ s

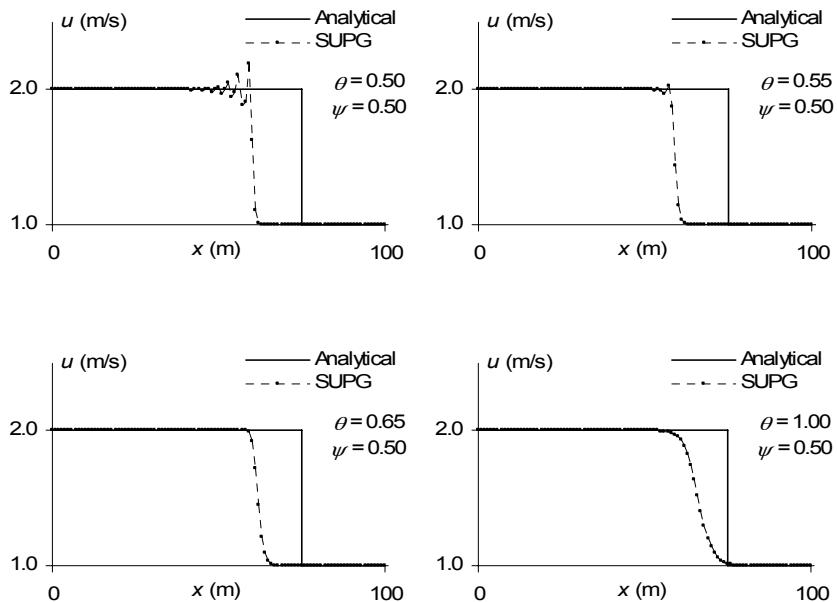


Figure 8.14. The inviscid Burgers equation. Solution using the SUPG technique at $t = 50$ s

The shock speed may be adjusted via the centering parameter ψ . In the present test case, u is larger on the left-hand side of the shock than on the right-hand side. Consequently, decreasing ψ in equation [8.81] should be expected to lead to a larger interpolated value, and thus a larger shock speed. Conversely, increasing φ should be expected to yield a smaller shock speed.

This is confirmed by Figures 8.15 and 8.16, where the influence of ψ is studied for two different values of θ .

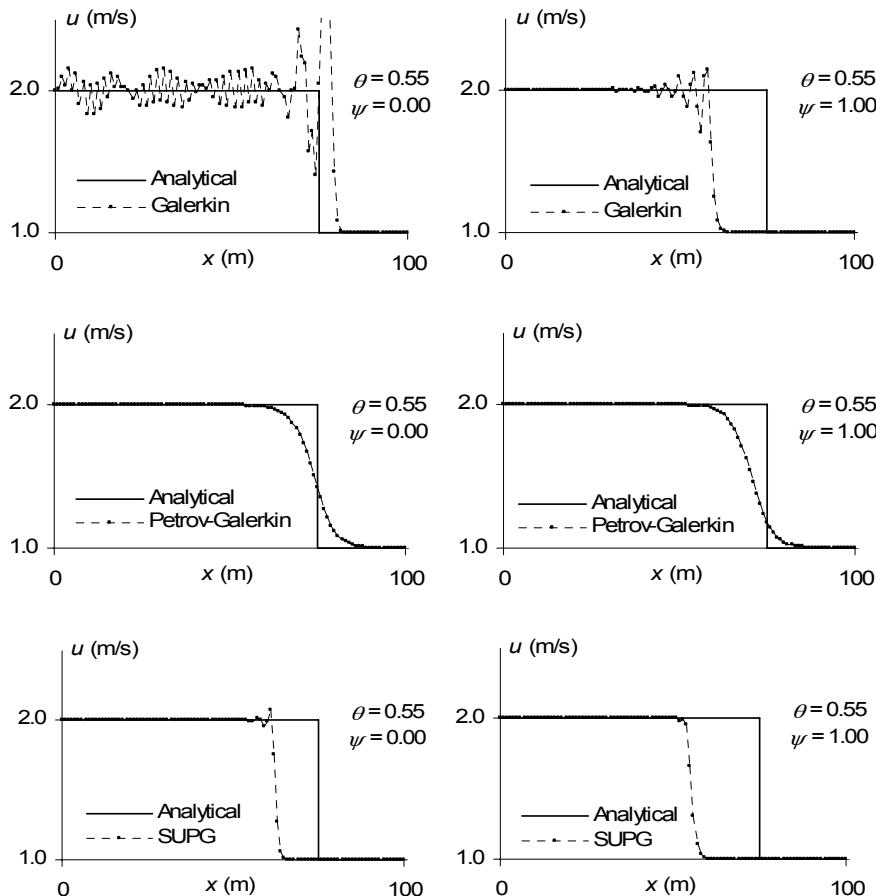


Figure 8.15b. The inviscid Burgers equation. Influence of the centering parameter ψ for $\theta = 0.55$

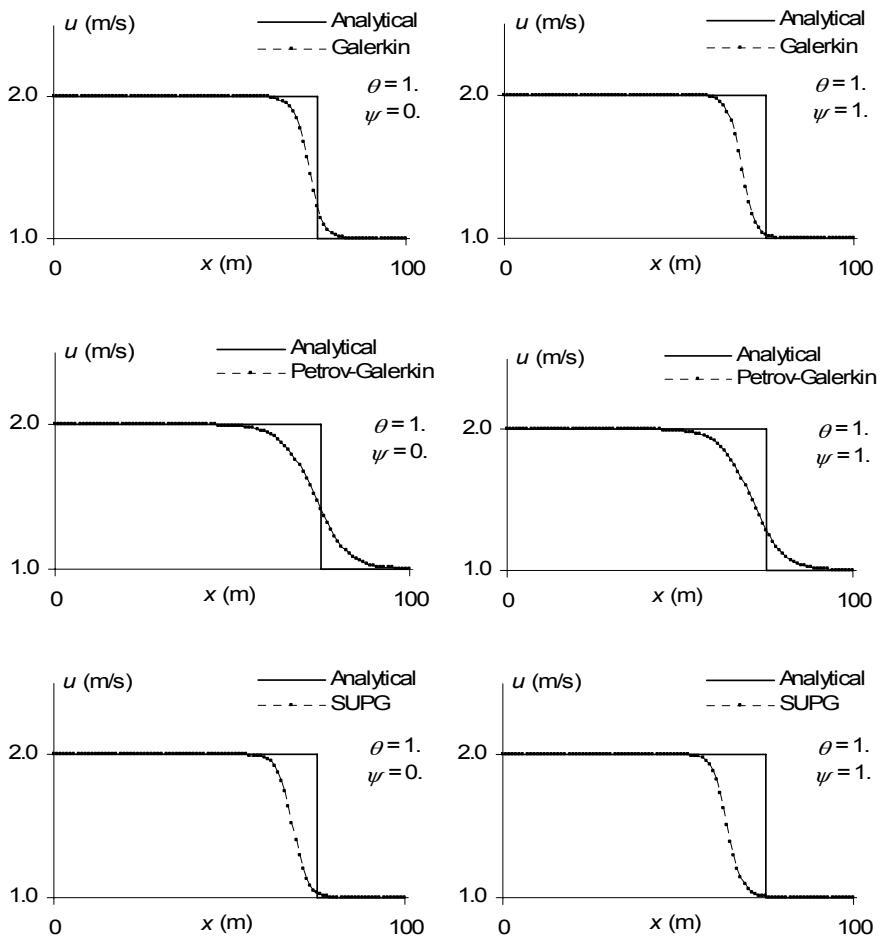


Figure 8.16. The inviscid Burgers equation. Influence of the centering parameter ψ for $\theta = 1$

The Petrov-Galerkin technique allows the shock to be located more accurately than the other two methods, at the expense of an increased numerical diffusion. No combination of θ and ψ , however, allows any of the methods to locate the shock correctly. This militates in favor of a more accurate estimate of the Courant number. In the next section, a semi-implicit estimate is proposed.

8.5.2.2. Solution by classical Galerkin techniques: semi-implicit estimate of Cr

The following semi-implicit estimate is proposed for the Courant number:

$$Cr = \left\{ (1 - \theta_{Cr}) \left[(1 - \psi) u_{i-1}^n + \psi u_i^n \right] + (\theta_{Cr} \left[(1 - \psi) u_{i-1}^{n+1} + \psi u_i^{n+1} \right]) \right\} \frac{\Delta t}{\Delta x} \quad [8.82]$$

where θ_{Cr} is an implicitation parameter that is specific to the estimate of the Courant number. It does not necessarily take the same value as the parameter θ in the scheme in [8.76–77]. Equation [8.82] is a generalization of equation [8.81], that is obtained for $\theta_{Cr} = 0$.

The semi-implicit estimate [8.82] is used within the following iterative procedure:

- 1) Assuming that the values at time level n are known, compute Cr for the first iteration using equation [8.81].
- 2) Use the thus estimated Cr to compute the coefficients in equations [8.76–77].
- 3) Solve the system of equations for the unknown values u_i^{n+1} . This provides a first estimate of the solution at the unknown time level $n + 1$.
- 4) Use the estimates u_i^{n+1} in equation [8.82] to update the estimate of the Courant number.

Steps 3) – 4) must be repeated until convergence is reached, that is, until two successive iterations yield the same (or almost the same) value for u_i^{n+1} . The appreciation of iteration convergence is left to the user of the method. In most practical cases, two or three iterations are seen to be sufficient.

The test case presented in section 8.4.2.1 is repeated. The parameters are given in Table 8.3.

As shown in Table 8.3, the iterative, semi-implicit estimate of the Courant number is a key factor in the accuracy of the method. For each of the techniques, it is possible to find a value for θ_{Cr} for which the shock is located correctly in the numerical solution. Indeed, in all three methods, $\theta_{Cr} = 0$ (that is, the explicit estimate [8.79]) gives a shock that propagates too slowly, while $\theta_{Cr} = 1.0$ yields a shock that propagates too fast.

The value $\theta_{\text{Cr}} = 1/2$ gives good results for the Petrov-Galerkin technique. In the Galerkin and SUPG schemes, θ_{Cr} must be taken as slightly smaller than the value to recover a correct shock location, thereby guaranteeing mass conservation.

Symbol	Meaning	Value
u_0	Initial condition	1 m/s
u_b	Prescribed velocity at the left-hand boundary	2 m/s
L	Domain length	100 m
T	Simulated time	50 s
Δt	Computational time step	1 s
Δx	Cell width	1 m
θ	Scheme implicitation parameter	0. and 0.7
θ_{Cr}	Implicitation parameter for the estimate of Cr	0.5 and 1.0
ρ	Value if $a\lambda/\Delta x$ (SUPG scheme)	1.
ψ	Centering parameter for the estimate of Cr	0.5

Table 8.3. Inviscid Burgers equation with semi-implicit estimate for Cr.
Parameters of the test case

These combinations of θ , θ_{Cr} and ψ must not be taken as general values that allow for conservation for all possible combinations of initial and boundary conditions. Consider the case of the Petrov-Galerkin technique with piecewise constant test functions [8.53].

As illustrated in Figure 8.17, the combination $\theta = 0.7$ and $\theta_{\text{Cr}} = \psi_{\text{Cr}} = 0.5$ guarantees conservation for $u_0 = 1$ m/s and $u_b = 2$ m/s. The test is repeated for the same combination of parameters, but the initial and boundary conditions are modified into $u_0 = 0$ m/s and $u_b = 5$ m/s. Figure 8.18 shows the computed profile at $T = 30$ s. It can be seen that the combination of numerical parameters is no longer optimal and does not guarantee conservation for this new set of initial and boundary conditions.

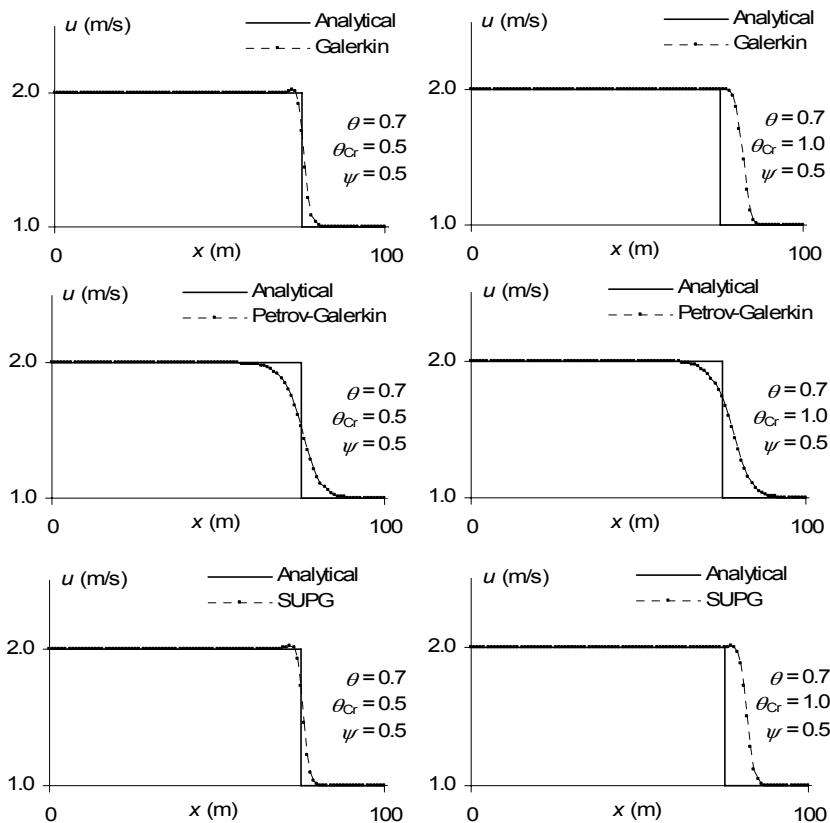


Figure 8.17. The inviscid Burgers equation. Influence of the implicitation parameter θ_{Cr} for the Galerkin, Petrov-Galerkin with piecewise constant test function and SUPG techniques

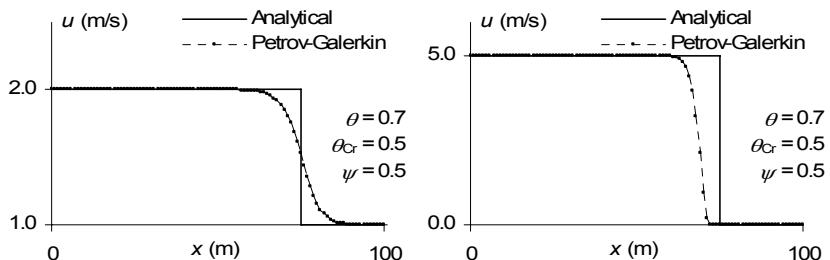


Figure 8.18. The inviscid Burgers equation. Solution using the Petrov-Galerkin technique with piecewise constant weighting functions for two different combinations of initial and boundary conditions

8.5.2.3. Solution by the RKDG2 technique

The RKDG2 technique is applied to the problem described in the previous sections. Figure 8.19 shows the numerical solution obtained at $T = 50$ seconds using the RKDG2 technique. The computational time step is $\Delta t = 0.1$ s. Obviously, conservation is ensured, at the expense however of computational rapidity, because the maximum permissible time step for solution stability is $\Delta t_{\max} = 0.15$ s.

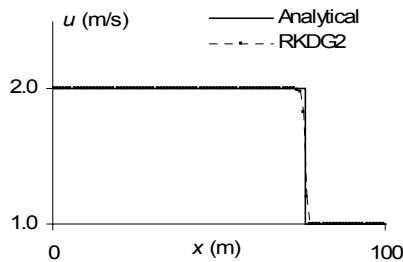


Figure 8.19. The inviscid Burgers equation. Solution using the RKDG2 technique

8.6. Summary

8.6.1. What you should remember

Finite element methods solve the weak form of conservation laws. The partial differential equation (PDE), or the system of PDEs, to be solved is multiplied by a pre-defined function, called a weighting or test function, and integrated over the solution domain. This means that the PDEs to be solved are solved in an average sense over the domain.

In classical Galerkin techniques (sections 8.1 to 8.3), space is discretized using computational points, also called nodes, that form elements. The solution is sought as the sum of pre-defined shape functions that take the value 1 at one node and 0 at all other nodes, multiplied by nodal values. The solution is known completely over the domain provided that each nodal value is known. In discontinuous Galerkin techniques (section 8.4), space is discretized using computational cells as in finite volume techniques. The solution is sought as a sum of elementary shape functions (for instance Legendre polynomials) defined over each cell, weighted by a cell coefficient. The solution is known over the computational domain provided that all the coefficients are known for all the elementary shape functions in each cell. In contrast with classical Galerkin techniques, discontinuous Galerkin techniques allow for discontinuous solutions at nodes.

Among the classical techniques, the Galerkin technique (see section 8.1.3.1) takes the shape and test functions from the same function space. In the Petrov-Galerkin technique (see section 8.1.3.2), the shape and test functions are taken from different function spaces. As a particular case, the SUPG approach (see section 8.1.3.3) uses test functions derived from the shape functions via the addition of a gradient-based term that allows for upwinding. This allows the oscillatory character of the numerical solution to be minimized to some extent.

When nonlinear PDEs (or systems) are to be solved in the framework of implicit or semi-implicit, classical Galerkin techniques, it is more convenient to solve the non-conservation form of the equations. However, this may lead to conservation problems, as shown by the computational examples of section 8.5.2. The example of the inviscid Burgers equation shows that a purely explicit estimate of the wave propagation speed gives incorrect shock speed estimates. The correct shock speed is recovered only if the computation of the wave propagation speed is made semi-implicit in the framework of an iterative procedure. In addition, a combination of implicitation and centering parameters that yield a correct shock speed for a given combination of initial and boundary conditions may give incorrect solutions for a different set of initial and boundary conditions.

Explicit, discontinuous Galerkin techniques introduced in section 8.4 are essentially conservative in that the computation of the average element value obeys a finite volume formalism. The higher-order components of the solution usually require limiting. Explicit discontinuous Galerkin techniques are stable when used in conjunction with Runge-Kutta algorithms, hence the initials RKDG for such methods. The stability constraints of RKDG techniques are more severe than those of classical finite volume schemes, thus constraining the permissible computational time steps to smaller values (see section 8.4.4).

8.6.2. Application exercises

Apply the Galerkin technique, the Petrov-Galerkin technique with piecewise constant test functions, the SUPG approach and the RKDG2 technique to the linear advection equation and the inviscid Burgers equation. Apply these discretizations to the test cases shown in section 8.5.

Indications and searching tips for the solution of these exercises can be found at the following URL: <http://vincentguinot.free.fr/waves/exercises.htm>.

Chapter 9

Treatment of Source Terms

9.1. Introduction

The numerical techniques presented in Chapters 6 to 8 mostly deal with hyperbolic conservation laws and hyperbolic systems without source terms. The equations or systems of equations found in practical engineering applications, however, contain source terms arising from variations in the geometry. Examples are the water hammer equations in pipes with variable cross-sections, the Saint Venant equations in non-prismatic channels and/or variable bottom slope, or the one-dimensional Euler equations in domains of variable cross-sectional areas. A careless discretization of the source terms may induce artificial perturbations in the computed profiles, if not solution instability. The need for source term discretization techniques that preserve equilibrium conditions without introducing spurious oscillations in the computed variables has led to the general notion of well-balanced schemes.

Giving a complete and exhaustive description of numerical techniques for source term discretization is beyond the scope of this book. The subject would actually deserve a book in itself. The purpose of this chapter is to present the broad lines of the main families of numerical techniques introduced over the past two decades to deal with source terms in hyperbolic systems of conservation laws. The water hammer equations (the simplest possible hyperbolic system of conservation laws) and the shallow water equations (a subject of intensive research over the past two decades, see e.g. [HER 07, TOR 07]) are used as illustrative examples.

Section 9.2 introduces the issue of geometric source terms discretization with two examples. The key notion of C -property is then introduced. Section 9.3 deals

with the upwind approach to source term discretization. Section 9.4 presents the quasi-steady wave algorithm, and section 9.5 is an introduction to well-balancing techniques.

9.2. Problem position

9.2.1. Example 1: the water hammer equations

The issue of source term discretization is illustrated using the water hammer equations introduced in section 2.3. The water hammer equations form the simplest possible 2×2 hyperbolic system of conservation laws, with constant, opposite wave speeds. This system is linear. The conservation form [2.2] of the equations, recalled here:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} = S$$

is obtained by defining the conserved variable U , the flux F and the source term S as in equation [2.68], recalled hereafter:

$$U = \begin{bmatrix} \rho A \\ \rho Q \end{bmatrix}, F = \begin{bmatrix} \rho Q \\ Ap \end{bmatrix}, S = \begin{bmatrix} 0 \\ \frac{\partial A}{\partial x} p - \rho g A \sin \theta - k|u|u \end{bmatrix}$$

where A is the cross-sectional area of the pipe, k is the friction coefficient, p is the pressure, Q is the liquid discharge, u is the flow velocity, θ is the angle between the axis of the pipe and the horizontal, and ρ is the density of the fluid. Recall that the variations in the mass per unit length ρA and the pressure force Ap obey relationship [2.44], that can be rewritten as:

$$d(Ap) = c^2 d(\rho A) \quad [9.1]$$

where c is the (assumed uniform) sound speed. Equation [9.1] is to be understood for a fixed value of the abscissa x along the pipe. It is also recalled that c is constant in time, regardless of the value of the pressure p .

Consider the simple configuration where the pipe is frictionless, horizontal, with a variable cross-sectional area A (see Figure 9.1). The source term S simplifies to

$$S = \begin{bmatrix} 0 \\ \frac{\partial A}{\partial x} p \end{bmatrix} \quad [9.2]$$

Assume that the initial and boundary conditions in the pipe are such that static equilibrium is verified, that is:

$$\left. \begin{array}{l} p(x,0) = p_0 \\ u(x,0) = 0 \\ p(0,t) = p(L,t) = p_0 \end{array} \right\} \quad \forall x \in [0, L] \quad \forall t > 0 \quad [9.3]$$

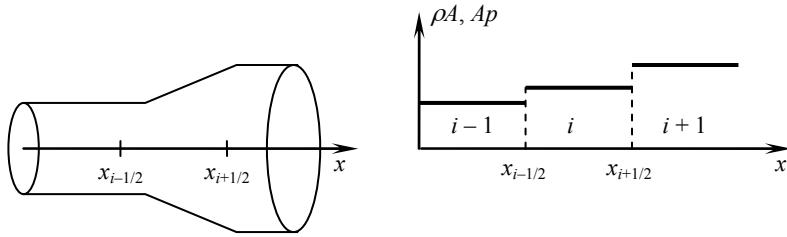


Figure 9.1. Definition sketch for the water hammer equations with variable pipe cross-section. Left: geometry of the pipe. Right: resulting initial, steady-state profile for the conserved variable ρA and the pressure force Ap

Substituting these conditions into equation [2.2] with definition [2.68] and simplification [9.2] leads to:

$$\left. \begin{array}{l} \frac{\partial}{\partial t}(\rho A) = -\frac{\partial}{\partial x}(\rho Q) = -\frac{\partial}{\partial x}(\rho u A) = 0 \\ \frac{\partial}{\partial t}(\rho Q) = -\frac{\partial}{\partial x}(Ap) + p \frac{\partial A}{\partial x} = -A \frac{\partial p}{\partial x} = 0 \end{array} \right\} \quad \forall \begin{cases} x \in [0, L] \\ t > 0 \end{cases} \quad [9.4]$$

In other words, static equilibrium is preserved at later times at all points in the pipe.

Assume now that the water hammer equations are to be solved using the finite volume discretization [7.3], recalled here:

$$U_i^{n+1} = U_i^n + \frac{\Delta t}{\Delta x} \left(F_{i-1/2}^{n+1/2} - F_{i+1/2}^{n+1/2} \right) + \Delta t S_i^{n+1/2}$$

Assume that a classical approximate Riemann solver, such as Roe's, Lax-Friedrichs' or the HLL solver (see Appendix C for more details), is used to compute the flux at the interfaces between the computational cells. The water hammer

equations being linear and c being assumed uniform, all three solvers lead to the same formula:

$$\begin{aligned} \left[\frac{\rho Q}{Ap} \right]_{i-1/2}^{n+1/2} &= \frac{1}{2} \left[\frac{(\rho Q)_{i-1}^n + (\rho Q)_i^n}{(Ap)_{i-1}^n + (Ap)_i^n} \right] + \frac{c}{2} \left[\frac{(\rho A)_{i-1}^n - (\rho A)_i^n}{(\rho Q)_{i-1}^n - (\rho Q)_i^n} \right] \\ &= \frac{1}{2} \left[\frac{(\rho Q)_{i-1}^n + (\rho Q)_i^n}{(Ap)_{i-1}^n + (Ap)_i^n} \right] + \frac{1}{2c} \left[\frac{(Ap)_{i-1}^n - (Ap)_i^n}{c^2 (\rho Q)_{i-1}^n - c^2 (\rho Q)_i^n} \right] \end{aligned} \quad [9.5]$$

If the initial state verifies static equilibrium at time level n , Q is zero and p is equal to p_0 within all the cells. Equation [9.5] becomes:

$$\left[\frac{\rho Q}{Ap} \right]_{i-1/2}^{n+1/2} = \begin{cases} \frac{A_{i-1} - A_i}{2c} p_0 \\ \frac{A_{i-1} + A_i}{2} p_0 \end{cases} \quad [9.6]$$

The first component of the vector equation [9.6] indicates that, if $A_{i-1} \neq A_i$, an artificial, non-zero mass discharge is computed at the interface $i - 1/2$. This non-zero discharge triggers pressure waves that propagate into the computational domain, thus destroying the steady-state character of the numerical solution.

This simple example shows that the source terms in hyperbolic systems should not be discretized independently of the conservation part, otherwise violating simple equilibrium requirements. Moreover, the presence of the source term in the momentum equation may lead us to revise the discretization of the continuity equation.

9.2.2. Example 2: the shallow water equations

The one-dimensional shallow water equations are obtained by restricting the two-dimensional shallow water equations (see section 5.4) to their one-dimensional projection. The one-dimensional shallow water equations can be written in the form [2.2] by defining U, F and S as in equation [7.83], recalled here:

$$U = \begin{bmatrix} h \\ q \end{bmatrix} = \begin{bmatrix} h \\ hu \end{bmatrix}, \quad F = \begin{bmatrix} q \\ M \end{bmatrix} = \begin{bmatrix} hu \\ hu^2 + gh^2 / 2 \end{bmatrix}, \quad S = \begin{bmatrix} 0 \\ (S_{0,x} - S_{f,x})gh \end{bmatrix}$$

where g is the gravitational acceleration, h is the water depth, u is the flow velocity, $S_{0,x}$ and $S_{f,x}$ are respectively the bottom and friction slope in the x -direction. q and M are respectively the unit discharge and specific force.

Assume as in the previous section that the shallow water equations are discretized using the finite volume technique [7.3], and that the HLL Riemann solver is used in the calculation of the fluxes (see Appendix C for more details). Consider the situation, shown in Figure 9.2, where the water is initially at rest ($\zeta = h + z_b = \text{Const}$, $u = 0$ everywhere).

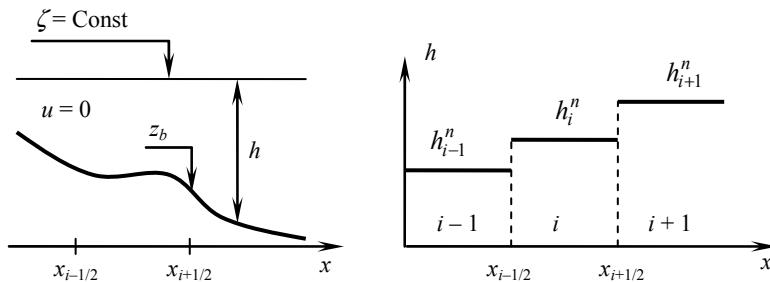


Figure 9.2. Definition sketch for the one-dimensional shallow water equations with variable bottom level: left: bottom geometry and steady-state equilibrium conditions; right: resulting discretized, initial state

The formula for the HLL Riemann solver is given by the second equation [C.3], recalled here:

$$F^* = \frac{\lambda^+ F_L - \lambda^- F_R}{\lambda^+ - \lambda^-} - \lambda^- \lambda^+ (U_L - U_R)$$

Applying this formula at the interface between the cells $i - 1$ and i with $u = 0$ yields the following expression for the unit discharge q :

$$q_{i-1/2} = -\frac{\lambda^- \lambda^+}{\lambda^+ - \lambda^-} (h_{i-1}^n - h_i^n) \quad [9.7]$$

where λ^- and λ^+ are respectively estimates of the minimum wave speed $u - c$ and the maximum wave speed $u + c$. Since the water is at rest, applying e.g. Davis' formula [C.5] for λ^- and λ^+ leads to:

$$\lambda^+ = -\lambda^- = \left[g \max(h_{i-1}^n, h_i^n) \right]^{1/2} \quad [9.8]$$

Substituting equation [9.8] into the flux formula [9.7] gives:

$$(hu)_{i-1/2} = \frac{[g \max(h_{i-1}^n, h_i^n)]^{1/2}}{2} (h_{i-1}^n - h_i^n) \quad [9.9]$$

If the bottom is horizontal, $h_{i-1}^n = h_i^n$ and the unit discharge remains equal to zero. But if the bottom is not horizontal, $h_{i-1}^n \neq h_i^n$. As a consequence, a non-zero discharge is computed even though the initial conditions correspond to water at rest under static equilibrium conditions. This generates artificial waves that propagate throughout the computational domain, eventually destroying the static character of the solution.

9.2.3. Stationary solution and C–property

In the above two examples, the artificial fluxes are diffusive fluxes stemming from second-order truncation errors (see Appendix B for detailed considerations on truncation errors and consistency aspects). As shown in [VAZ 99, CHA 03], these artificial fluxes can be eliminated if the discretization of the source terms can be made second-order accurate with respect to space.

In what follows, the source term is assumed to take the form:

$$S = f(U) \frac{\partial \varphi}{\partial x} \quad [9.10]$$

where f is a known vector function of U and φ is a parameter. For instance, in the case of the water hammer equations, $\varphi = A$ and $f = [0, p]^T$. In the case of the one-dimensional shallow water equations, $\varphi = z_b$ and $f = [0, gh]^T$.

First-order consistency condition imposes that if φ and U are identical in the cells i and $i + 1$, the discretized source term must be zero:

$$\left. \begin{array}{l} U_{i-1} = U_i \\ \varphi_{i-1} = \varphi_i \end{array} \right\} \Rightarrow S(\varphi_{i-1}, \varphi_i, U_{i-1}, U_i) = 0 \quad [9.11]$$

However, the first-order consistency condition is not sufficient to guarantee a satisfactory discretization of the source term. As an example, in sections 9.2.1 and 9.2.2, condition [9.11] is satisfied. Nevertheless, steady-state conditions are not preserved for arbitrary geometries. It is thus necessary to define so-called “enhanced

consistency” conditions [CHA 03], that allow the so-called *C*-property to be verified.

A flow field that verifies steady-state (or stationary) conditions at time level n should satisfy static equilibrium conditions at time level $n + 1$. This is true only if the difference between the fluxes is balanced exactly by the source term. For a finite volume scheme, this condition may be written as:

$$\frac{\partial U}{\partial t} = 0 \Rightarrow F_{i-1/2}^{n+1/2} - F_{i+1/2}^{n+1/2} + \Delta x (S_{i-1/2}^+ + S_{i+1/2}^-) = 0 \quad [9.12]$$

In [VAZ 99], the so-called *C*-property is defined as follows:

- the discretization is said to satisfy the exact *C*-property if the steady-state condition [9.12] is satisfied exactly;
- the discretization is said to satisfy the approximate *C*-property if the steady-state condition [9.12] is satisfied with at least second-order accuracy.

As in [VAZ 99], source terms involving sums of functions in the form [9.10] can be broken into several elementary source terms to be discretized independently from each other. The *C*-property was originally written for static conditions, that is, for a fluid at rest [VAZ 99]. However, more general steady-state preserving discretizations have been proposed by a number of authors, see for instance [HUB 00], [BUR 04].

9.3. Source term upwinding techniques

9.3.1. Principle

Source term upwinding has been applied to the calculation of the one-dimensional shallow water equations [BER 94] in channels of constant width, to the solution of the two-dimensional shallow water equations [BER 98, BRU 02] and to the solution of the Saint Venant equations in channels with variable width [VAZ 99, GAR 00]. Applications of the technique to higher-order TVD schemes can be found in [BUR 01]. Extensions of the method to various numerical techniques are provided in [CHA 03], where the source term is split into centered and upwind parts, which are discretized so as to preserve stationary solutions.

The technique is particularly adapted to discretizations where wave speeds and their propagation directions are well-identified and used explicitly in the discretization of the fluxes. This is the case in particular with finite volume methods (see Chapter 7) that use approximate Riemann solvers based on approximations of

the wave speeds (see Appendix C for examples). Consider a hyperbolic system discretized as in equation [7.3] (a slightly different writing is used hereafter):

$$U_i^{n+1} = U_i^n + \frac{\Delta t}{\Delta x} \left(F_{i-1/2}^{n+1/2} - F_{i+1/2}^{n+1/2} + \Delta x S_i^{n+1/2} \right) \quad [9.13]$$

where the source term can be written as in equation [9.10].

It is first recalled that in finite volume techniques using Riemann solvers, the solution at the interface between two adjacent cells is classically approximated as a succession of discontinuities separating regions of constant state (Figure 9.3).

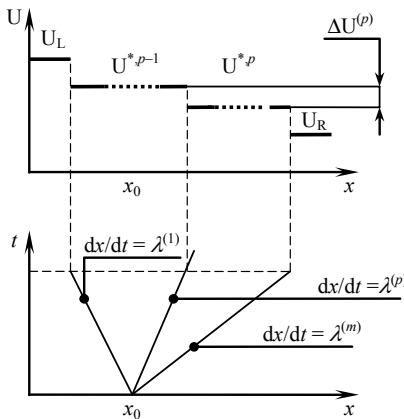


Figure 9.3. Source term upwinding across the waves in the solution of the Riemann problem.
Definition sketch in the physical space (top) and in the phase space (bottom)

Denoting by $U^{*,p-1}$ and $U^{*,p}$ the solution on the left- and right-hand sides of the p th wave, the jumps in U and F across the p th wave may be written as:

$$\left. \begin{aligned} U^{*,p} - U^{*,p-1} &= \alpha^{(p)} K^{(p)} \\ F^{*,p} - F^{*,p-1} &= \lambda^{(p)} (U^{*,p} - U^{*,p-1}) = \lambda^{(p)} \alpha^{(p)} K^{(p)} \end{aligned} \right\} \quad [9.14]$$

where the coefficients $\alpha^{(p)}$ are the wave strengths and the vectors $K^{(p)}$ are the eigenvectors of the Jacobian matrix A of F with respect to U . Note that summing the jumps across all the waves leads to:

$$\left. \begin{aligned} U_R - U_L &= K \alpha \\ F_R - F_L &= K \Lambda \alpha \end{aligned} \right\} \quad [9.15]$$

where K and a are respectively the matrix of eigenvectors of A and the vector formed by the wave strengths. Inverting the first equation [9.15] yields directly the expression of the wave strength vector α :

$$\alpha = K^{-1}(U_R - U_L) \quad [9.16]$$

Source term upwinding consists of discretizing the source term in the same way as the space derivative of the flux in upwind schemes:

$$\left. \begin{aligned} \Delta x S_i^{n+1/2} &= \sum_p \beta^{(p)} K^{(p)} = K\beta \\ \beta &= K^{-1} \Delta x S_i^{n+1/2} \end{aligned} \right\} \quad [9.17]$$

where β is a vector formed by the components $\beta^{(p)}$ of the wave strengths of the source term. The total source term [9.17] is split into two parts:

- the part that corresponds to negative wave speeds, $\lambda^{(p)} < 0$, is assigned to the cell on the left-hand side of the initial discontinuity,
- the part that corresponds to positive wave speeds, $\lambda^{(p)} > 0$, is assigned to the cell on the right-hand side of the initial discontinuity.

The discretization [9.13] is rewritten in the form:

$$U_i^{n+1} = U_i^n + \frac{\Delta t}{\Delta x} \left[F_{i-1/2}^{n+1/2} - F_{i+1/2}^{n+1/2} + \Delta x (S_{i-1/2}^+ + S_{i+1/2}^-) \right] \quad [9.18]$$

where $S_{i-1/2}^+$ is the contribution to the cell i of the source term between the cells $i-1$ and i , and $S_{i+1/2}^-$ is the contribution to the cell i of the source term between the cells i and $i+1$. The source term $S_{i-1/2}^+$ arises from the contribution of positive wave speeds, while the term $S_{i+1/2}^-$ arises from the contribution of the negative wave speeds:

$$\left. \begin{aligned} S_{i-1/2}^+ &= \sum_{\lambda^{(p)} \geq 0} \left[\beta^{(p)} K^{(p)} \right]_{i-1/2}^{n+1/2} \\ S_{i+1/2}^- &= \sum_{\lambda^{(p)} < 0} \left[\beta^{(p)} K^{(p)} \right]_{i+1/2}^{n+1/2} \end{aligned} \right\} \quad [9.19]$$

The contributions $S_{i-1/2}^+$ and $S_{i+1/2}^-$ are known if the wave strengths $\beta^{(p)}$ can be determined and if an estimate can be provided for the total source term. The wave strengths are obtained from the second equation [9.17]. The estimate of the source term is obtained by integrating equation [9.10] across each interface:

$$\left. \begin{aligned} \beta_{i-1/2}^{n+1/2} &= K^{-1} \Delta x S_{i-1/2}^{n+1/2} \approx (K^{-1} f)_{i-1/2}^{n+1/2} (\varphi_i - \varphi_{i-1}) \\ \beta_{i+1/2}^{n+1/2} &= K^{-1} \Delta x S_{i+1/2}^{n+1/2} \approx (K^{-1} f)_{i+1/2}^{n+1/2} (\varphi_{i+1} - \varphi_i) \end{aligned} \right\} \quad [9.20]$$

where $f_{i-1/2}^{n+1/2}$ is an “average” value of f between the cells $i-1$ and i . Its expression for f_{LR} must be devised in such a way that the exact or approximate C -property defined in section 9.2.3 is satisfied.

9.3.2. Application example 1: the water hammer equations

The source term upwinding technique is applied to the water hammer equations. The discretization is assumed to be given by equation [7.3], with formula [9.5] for the calculation of the fluxes. Equation [9.5] is recalled:

$$\left[\begin{array}{c} \rho Q \\ Ap \end{array} \right]_{i-1/2}^{n+1/2} = \frac{1}{2} \left[\begin{array}{c} (\rho Q)_{i-1}^n + (\rho Q)_i^n \\ (Ap)_{i-1}^n + (Ap)_i^n \end{array} \right] + \left[\begin{array}{c} \frac{(Ap)_{i-1}^n - (Ap)_i^n}{2c} \\ \frac{(\rho Q)_{i-1}^n - (\rho Q)_i^n}{2c} \end{array} \right]$$

The eigenvalues and eigenvectors of the water hammer equations (see section 2.3) are also recalled:

$$K = \begin{bmatrix} 1 & 1 \\ -c & c \end{bmatrix}, \quad K^{-1} = \frac{1}{2c} \begin{bmatrix} c & -1 \\ c & 1 \end{bmatrix} \quad [9.21]$$

The source term is discretized as (here at interface $i-1/2$):

$$\Delta x S_{i-1/2}^{n+1/2} \approx \begin{bmatrix} 0 \\ (A_i - A_{i-1}) p_{i-1/2} \end{bmatrix} \quad [9.22]$$

where the estimate $p_{i-1/2}$ must be determined so as to satisfy the C -property (this point is examined at the end of the section). The vector β of wave strengths is given by:

$$\beta = K^{-1} \Delta x S_{i-1/2}^{n+1/2} = \frac{1}{2c} \begin{bmatrix} (A_{i-1} - A_i) p_{i-1/2} \\ (A_i - A_{i-1}) p_{i-1/2} \end{bmatrix} \quad [9.23]$$

The waves with speeds $-c$ and $+c$ are respectively directed to the cells $i - 1$ and i . This leads to the following contributions for the source term at the interface $i - 1/2$:

$$\left. \begin{aligned} S_{i-1/2}^- &= \beta^{(1)} K^{(1)} = \left\{ \begin{aligned} &\frac{(A_{i-1} - A_i)p_{i-1/2}}{2c} \\ &(A_i - A_{i-1})p_{i-1/2} \end{aligned} \right\} \\ S_{i-1/2}^+ &= \beta^{(2)} K^{(2)} = \left\{ \begin{aligned} &\frac{(A_i - A_{i-1})p_{i-1/2}}{2c} \\ &(A_i - A_{i-1})p_{i-1/2} \end{aligned} \right\} \end{aligned} \right\} \quad [9.24]$$

The cell i receives on the left-hand interface $i - 1/2$ a total contribution formed by the sum of the flux $F_{i-1/2}^{n+1/2}$ and the contribution $S_{i-1/2}^+$ of the source term:

$$F_{i-1/2}^{n+1/2} + S_{i-1/2}^+ = \frac{1}{2} \left[\begin{aligned} &(\rho Q)_{i-1}^n + (\rho Q)_i^n + \frac{(Ap)_{i-1}^n - (Ap)_i^n}{c} \\ &\quad + \frac{(A_i - A_{i-1})p_{i-1/2}}{c} \\ &(Ap)_{i-1}^n + (Ap)_i^n + [(\rho Q)_{i-1}^n - (\rho Q)_i^n]c \\ &\quad + (A_i - A_{i-1})p_{i-1/2} \end{aligned} \right] \quad [9.25]$$

The expression for $p_{i-1/2}$ is now examined. For an initial, steady-state configuration, with a uniform pressure p_0 , discharge Q_0 and density ρ_0 , equation [9.25] simplifies to:

$$F_{i-1/2}^{n+1/2} + S_{i-1/2}^+ = \frac{1}{2} \left[\begin{aligned} &\rho_0 Q_0 + \frac{A_{i-1} - A_i}{2c} (p_0 - p_{i-1/2}) \\ &\frac{A_{i-1} + A_i}{2} p_0 + (A_i - A_{i-1})p_{i-1/2} \end{aligned} \right] \quad [9.26]$$

Any estimate for $p_{i-1/2}$ that satisfies the consistency condition:

$$\left. \begin{aligned} p_{i-1/2} &= f(p_{i-1}^n, p_i^n) \\ f(U, U) &= U \end{aligned} \right\} \quad [9.27]$$

satisfies the C -property exactly. This is the case with the following two estimates:

$$\left. \begin{aligned} p_{i-1/2} &= \frac{p_{i-1}^n + p_i^n}{2} \\ p_{i-1/2} &= \frac{A_{i-1} p_{i-1}^n + A_i p_i^n}{A_{i-1} + A_i} \end{aligned} \right\} [9.28]$$

In both cases, $p_{i-1/2} = p_0$ and equation [9.26] becomes:

$$F_{i-1/2}^{n+1/2} + S_{i-1/2}^+ = \frac{1}{2} \begin{bmatrix} \rho_0 Q_0 \\ A_i p_0 \end{bmatrix} [9.29]$$

It is easy to check that:

$$F_{i+1/2}^{n+1/2} + S_{i+1/2}^- = \frac{1}{2} \begin{bmatrix} \rho_0 Q_0 \\ A_i p_0 \end{bmatrix} [9.30]$$

The mass discharge is equal to the uniform-steady-state value $\rho_0 Q_0$ at all cell interfaces and the equilibrium condition [9.12] is satisfied.

9.3.3. Application example 2: the shallow water equations with HLL solver

Applications of the source term upwinding technique to the shallow water equations in conjunction with Roe's solver, Van Leer's Q-scheme and flux splitting techniques can be found in [BER 94, BER 98, GAR 00, VAZ 99]. In this section, the HLL solver is applied.

For the sake of simplicity, the frictionless restriction of the one-dimensional shallow water equations is considered:

$$\begin{aligned} U &= \begin{bmatrix} h \\ q \end{bmatrix} = \begin{bmatrix} h \\ hu \end{bmatrix}, & F &= \begin{bmatrix} q \\ M \end{bmatrix} = \begin{bmatrix} hu \\ hu^2 + gh^2 / 2 \end{bmatrix}, \\ S &= \begin{bmatrix} 0 \\ S_{0,x} gh \end{bmatrix} = \begin{bmatrix} 0 \\ -gh \partial z_b / \partial x \end{bmatrix} \end{aligned} [9.31]$$

where M and q are respectively the specific force and unit discharge. The eigenvalues and eigenvectors are obtained as particular cases of the Saint Venant equations in prismatic, rectangular channels:

$$\lambda^{(1)} = u - c, \quad \lambda^{(2)} = u + c,$$

$$K^{(1)} = \begin{bmatrix} 1 \\ \lambda^{(1)} \end{bmatrix}, \quad K^{(2)} = \begin{bmatrix} 1 \\ \lambda^{(2)} \end{bmatrix}, \quad K^{-1} = \frac{1}{\lambda^{(2)} - \lambda^{(1)}} \begin{bmatrix} \lambda^{(2)} & -1 \\ -\lambda^{(1)} & 1 \end{bmatrix} \quad [9.32]$$

The HLL solver is used for the estimate of the flux:

$$\left. \begin{aligned} q_{i-1/2} &= \frac{\lambda^+ q_{i-1}^n - \lambda^- q_i^n}{\lambda^+ - \lambda^-} - \frac{\lambda^- \lambda^+}{\lambda^+ - \lambda^-} (h_{i-1}^n - h_i^n) \\ M_{i-1/2} &= \frac{\lambda^+ M_{i-1}^n - \lambda^- M_i^n}{\lambda^+ - \lambda^-} - \frac{\lambda^- \lambda^+}{\lambda^+ - \lambda^-} (q_{i-1}^n - q_i^n) \end{aligned} \right\} \quad [9.33]$$

where λ^- and λ^+ are estimated as in equation [C.5]:

$$\left. \begin{aligned} \lambda^- &= \min[(u - c)_{i-1}^n, (u - c)_i^n, 0] \\ \lambda^+ &= \max[(u + c)_{i-1}^n, (u + c)_i^n, 0] \end{aligned} \right\} \quad [9.34]$$

The wave strengths of the source term are obtained from equation [9.20]:

$$\begin{aligned} \beta_{i-1/2}^{n+1/2} &= K^{-1} \Delta x S_{i-1/2}^{n+1/2} \\ &= \frac{1}{\lambda^{(2)} - \lambda^{(1)}} \begin{bmatrix} \lambda^{(2)} & -1 \\ -\lambda^{(1)} & 1 \end{bmatrix}_{i-1/2} \begin{bmatrix} 0 \\ g h_{i-1/2}^{n+1/2} (z_{b,i-1} - z_{b,i}) \end{bmatrix} \\ &= \frac{g h_{i-1/2}^{n+1/2}}{\lambda^{(2)} - \lambda^{(1)}} (z_{b,i-1} - z_{b,i}) \begin{bmatrix} -1 \\ 1 \end{bmatrix} \end{aligned} \quad [9.35]$$

The source term at the interface $i-1/2$ is split into two terms according to equation [9.19]:

$$\left. \begin{aligned} S_{i-1/2}^- &= \begin{cases} 0 & \text{if } \lambda^- \geq 0 \\ \beta^{(1)} K^{(1)} & \text{if } \lambda^- < 0 \text{ and } \lambda^+ \geq 0 \\ \beta^{(1)} K^{(1)} + \beta^{(2)} K^{(2)} & \text{if } \lambda^+ < 0 \end{cases} \\ S_{i-1/2}^+ &= \begin{cases} \beta^{(1)} K^{(1)} + \beta^{(2)} K^{(2)} & \text{if } \lambda^- \geq 0 \\ \beta^{(2)} K^{(2)} & \text{if } \lambda^- < 0 \text{ and } \lambda^+ \geq 0 \\ 0 & \text{if } \lambda^+ < 0 \end{cases} \end{aligned} \right\} \quad [9.36]$$

with:

$$\left. \begin{aligned} \beta^{(1)} K^{(1)} &= \frac{gh_{i-1/2}^{n+1/2}}{\lambda^{(2)} - \lambda^{(1)}} (z_{b,i-1} - z_{b,i}) \begin{bmatrix} -1 \\ -\lambda^{(1)} \end{bmatrix} \\ \beta^{(2)} K^{(2)} &= \frac{gh_{i-1/2}^{n+1/2}}{\lambda^{(2)} - \lambda^{(1)}} (z_{b,i-1} - z_{b,i}) \begin{bmatrix} 1 \\ \lambda^{(2)} \end{bmatrix} \end{aligned} \right\} [9.37]$$

The terms $S_{i-1/2}^-$ and $S_{i-1/2}^+$ are used in equation [9.18]. They contribute respectively to the cell $i-1$ and i . The second components of $\beta^{(1)} K^{(1)}$ and $\beta^{(2)} K^{(2)}$ in equation [9.37] indicate that the source term arising from the topography is split proportionally into the wave speeds $\lambda^{(1)}$ and $\lambda^{(2)}$.

A straightforward estimate for the wave speeds is:

$$\left. \begin{aligned} \lambda_{i-1/2}^{(1)n+1/2} &= \lambda^- \\ \lambda_{i-1/2}^{(1)n+1/2} &= \lambda^+ \end{aligned} \right\} [9.38]$$

where λ^- and λ^+ are given by equation [9.34]. The expression for $h_{i-1/2}^n$ is obtained by writing the requirements imposed by the C -property. Consider the case where the water is initially at rest. Then, $\lambda^+ = -\lambda^- = \max(c_{i-1}^n, c_i^n)$ and the following estimates are obtained:

$$\left. \begin{aligned} q_{i-1/2}^{n+1/2} &= -\frac{\lambda^- \lambda^+}{\lambda^+ - \lambda^-} (h_{i-1}^n - h_i^n) = \frac{-\lambda}{2} (h_{i-1}^n - h_i^n) = \frac{\lambda^+}{2} (h_{i-1}^n - h_i^n) \\ M_{i-1/2}^{n+1/2} &= \frac{\lambda^+ M_{i-1}^n - \lambda^- M_i^n}{\lambda^+ - \lambda^-} = \frac{M_{i-1}^n + M_i^n}{2} = \frac{g}{2} [(h_{i-1}^n)^2 - (h_i^n)^2] \\ \beta^{(1)} K^{(1)} &= \frac{gh_{i-1/2}^{n+1/2}}{2\lambda^+} (z_{b,i-1} - z_{b,i}) \begin{bmatrix} -1 \\ \lambda^+ \end{bmatrix} \\ \beta^{(2)} K^{(2)} &= \frac{gh_{i-1/2}^{n+1/2}}{2\lambda^+} (z_{b,i-1} - z_{b,i}) \begin{bmatrix} 1 \\ \lambda^+ \end{bmatrix} \end{aligned} \right\} [9.39]$$

Noting that $z_{b,i-1} - z_{b,i} = h_i^n - h_{i-1}^n$ under equilibrium conditions leads to:

$$\left. \begin{aligned} \beta^{(1)} K^{(1)} &= \frac{gh_{i-1}^{n+1/2}}{2\lambda^+} (h_i^n - h_{i-1}^n) \begin{bmatrix} -1 \\ \lambda^+ \end{bmatrix} \\ \beta^{(2)} K^{(2)} &= \frac{gh_{i-1}^{n+1/2}}{2\lambda^+} (h_i^n - h_{i-1}^n) \begin{bmatrix} 1 \\ \lambda^+ \end{bmatrix} \end{aligned} \right\} [9.40]$$

The contribution of the interface $i - 1/2$ to the cell i is thus given by (a similar reasoning may be made for the contribution of the interface to the cell $i - 1$):

$$F_{i-1/2}^{n+1/2} + \beta^{(2)} K^{(2)} = \begin{bmatrix} \frac{\lambda^+}{2} (h_{i-1}^n - h_i^n) + \frac{gh_{i-1/2}^C}{2\lambda^+} (h_i^n - h_{i-1}^n) \\ \frac{g}{2} [(h_{i-1}^n)^2 - (h_i^n)^2] + \lambda^+ gh_{i-1/2}^M \frac{h_i^n - h_{i-1}^n}{2\lambda^+} \end{bmatrix} [9.41]$$

The C -property is satisfied if the vector $F_{i-1/2}^{n+1/2} + \beta^{(2)} K^{(2)}$ is zero. It is worth noting that two different estimates are introduced for $h_{i-1/2}^{n+1/2}$ in equation [9.41]: $h_{i-1/2}^C$ is used in the continuity equation, while $h_{i-1/2}^M$ is used in the momentum equation. The reason is that it is not possible to find a single expression for $h_{i-1/2}^{n+1/2}$ that satisfies the C -property in both the continuity and momentum equations. It is easy to check that the following estimates allow the C -property to be verified:

$$\left. \begin{aligned} h_{i-1/2}^C &= \frac{\lambda^{+2} + \lambda^{-2}}{2g} \\ h_{i-1/2}^M &= \frac{h_{i-1}^n + h_i^n}{2} \end{aligned} \right\} [9.42]$$

The expression for $h_{i-1/2}^M$ coincides with the formula given in [BER 98] for Van Leer's Q -scheme. [BER 98] also give the C -preserving formula for Roe's solver (see Appendix C for a brief description of the solver):

$$h_{i-1/2}^{n+1/2} = (h_{i-1}^n h_i^n)^{1/2} [9.43]$$

9.4. The quasi-steady wave algorithm

9.4.1. Principle

The quasi-steady wave algorithm [LEV 98] was introduced for finite volume discretizations. In this method, the parameter φ in the source term [9.10] is assumed piecewise constant, with a discontinuity located in the middle of the computational cells (Figure 9.4).

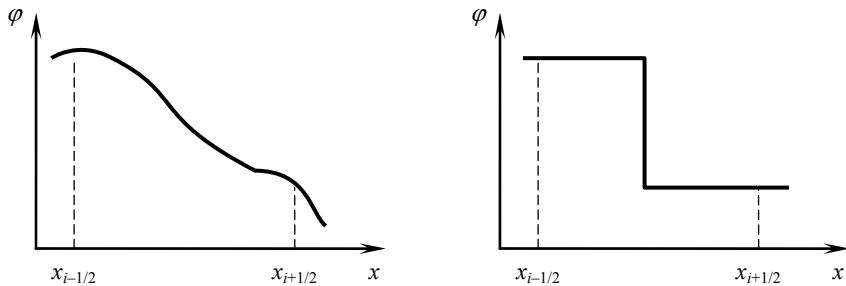


Figure 9.4. Quasi-steady wave algorithm. Left: variations in φ .
Right: discretized parameter in the cell i

The discontinuity in φ in the middle of the cell generates an additional Riemann problem, but the additional waves triggered by this Riemann problem do not reach the interfaces of the cell if the computational time step is kept small enough. Under these conditions, the additional Riemann problem does not influence the balance at the cell interfaces and the classical balance equation [7.3] can be used. The source term is assigned to the cell i :

$$\left. \begin{aligned} U_i^{n+1} &= U_i^n + \frac{\Delta t}{\Delta x} \left(F_{i-1/2}^{n+1/2} - F_{i+1/2}^{n+1/2} + \Delta x S_i^{n+1/2} \right) \\ \Delta x S_i^{n+1/2} &= S(U_i^n, \varphi_{i-1/2}, \varphi_{i+1/2}) \end{aligned} \right\} \quad [9.44]$$

In equation [9.44], an explicit estimate is proposed for the source term, but other estimates may be proposed. For definition [9.10] of the source term S , we have:

$$\Delta x S_i^{n+1/2} = \int_{x_{i-1/2}}^{x_{i+1/2}} f(x) \frac{\partial \varphi}{\partial x} dx \approx [\varphi(x_{i+1/2}) - \varphi(x_{i-1/2})] f_i^{n+1/2} \quad [9.45]$$

where $f_i^{n+1/2}$ is an average value of the vector function f over the cell i . If an explicit scheme is retained, f is defined from the known values at time level n . It must satisfy the consistency condition and the C -property presented in section 9.2.3:

(1) consistency condition:

$$\varphi_{i-1/2} = \varphi_{i+1/2} \Rightarrow S(\varphi_{i-1/2}, \varphi_{i+1/2}, U_i) = 0 \quad [9.46]$$

(2) steady-state condition: an initial-steady state must yield the following equality:

$$\frac{\partial U}{\partial t} = 0 \Rightarrow F_{i-1/2}^{n+1/2} - F_{i+1/2}^{n+1/2} + \Delta x S_i^{n+1} = 0 \quad [9.47]$$

9.4.2. Application to the water hammer equations

Consider the water hammer equations in a frictionless, horizontal pipe with a variable cross-sectional area A . The fluxes are computed as in equation [9.26], with the difference that A is constant across a given interface but may vary from one interface to the next. For interface $i - 1/2$, the formula for the flux becomes:

$$F_{i-1/2}^{n+1/2} = \left[\frac{\rho Q}{A p} \right]_{i-1/2}^{n+1/2} = \left[\frac{(\rho Q)_{i-1}^n + (\rho Q)_i^n}{2} + A_{i-1/2} \frac{p_{i-1}^n - p_i^n}{2c} \right. \\ \left. - A_{i-1/2} \frac{p_{i-1}^n + p_i^n}{2} + \frac{(\rho Q)_{i-1}^n - (\rho Q)_i^n}{2} c \right] \quad [9.48]$$

while the source term is computed as:

$$\Delta x S_i^{n+1/2} = (A_{i+1/2} - A_{i-1/2}) p_i^n \quad [9.49]$$

It is easy to check that equations [9.48–49] verify the consistency condition [9.46] and the C -property [9.47].

9.4.3. Application to the one-dimensional shallow water equations

The definition [9.33] of U , F and S for the one-dimensional shallow water equations is recalled:

$$\mathbf{U} = \begin{bmatrix} h \\ q \end{bmatrix} = \begin{bmatrix} h \\ hu \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} q \\ M \end{bmatrix} = \begin{bmatrix} hu \\ hu^2 + gh^2 / 2 \end{bmatrix},$$

$$\mathbf{S} = \begin{bmatrix} 0 \\ S_{0,x} \end{bmatrix} = \begin{bmatrix} 0 \\ -gh \partial z_b / \partial x \end{bmatrix}$$

In the present application example, the bottom level z_b is reconstructed using a piecewise constant function. The discontinuity in the bed level is assumed to be located at the center of the cell i . The average bottom level over the cell is thus:

$$z_{b,i} = \frac{z_{b,i-1/2} + z_{b,i+1/2}}{2} \quad [9.50]$$

The water depths $h_{i-1/2,L}$ and $h_{i-1/2,R}$ on the left- and right-hand sides of the interface $i - 1/2$ are given by:

$$\left. \begin{aligned} h_{i-1/2,L} &= h_{i-1}^n + z_{b,i-1} - z_{b,i-1/2} = h_{i-1}^n + \frac{z_{b,i-3/2} - z_{b,i-1/2}}{2} \\ h_{i-1/2,R} &= h_i^n + z_{b,i} - z_{b,i-1/2} = h_i^n + \frac{z_{b,i+1/2} - z_{b,i-1/2}}{2} \end{aligned} \right\} \quad [9.51]$$

Moreover, steady state is assumed over the cells. This means that:

$$\begin{aligned} q_{i-1/2,L} &= q_{i-1}^n \\ q_{i-1/2,R} &= q_i^n \end{aligned} \quad [9.52]$$

The flux $\mathbf{F}_{i+1/2}^{n+1/2}$ is computed by solving a Riemann problem with left and right states $[h_{i-1/2,L}, q_{i-1/2,L}]^T$ and $[h_{i-1/2,R}, q_{i-1/2,R}]^T$. For instance, if the HLL solver is used, we have:

$$\begin{aligned} \mathbf{F}_{i-1/2}^{n+1/2} &= \frac{\lambda^+}{\lambda^+ - \lambda^-} \left[\begin{bmatrix} q \\ M \end{bmatrix} \right]_{i-1/2,L}^n - \frac{\lambda^-}{\lambda^+ - \lambda^-} \left[\begin{bmatrix} q \\ M \end{bmatrix} \right]_{i-1/2,R}^n \\ &\quad - \frac{\lambda^- \lambda^+}{\lambda^+ - \lambda^-} \left(\left[\begin{bmatrix} h \\ q \end{bmatrix} \right]_{i-1/2,L}^n - \left[\begin{bmatrix} h \\ q \end{bmatrix} \right]_{i-1/2,R}^n \right) \end{aligned} \quad [9.53]$$

The source term for the continuity equation is zero. The expression of the source term in the momentum equation is derived from the force exerted by the bottom

discontinuity on the volume of water contained in the cell i . To do so, a control volume containing the discontinuity is defined (Figure 9.5). Remember that the pressure below the free surface obeys the following equation (see section 2.5.2.3):

$$\frac{p(z)}{\rho} = (\zeta - z)g \quad [9.54]$$

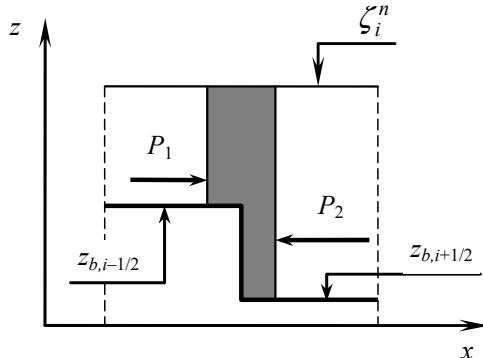


Figure 9.5. Quasi-steady wave algorithm for the shallow water equations. Balancing the forces exerted on a control volume that contains the bottom step

Integrating equation [9.54] between $z_{b,i-1/2}$ and ζ_i^n yields the expression for the pressure force per unit width P_1 on the left-hand side of the control volume:

$$\frac{P_1}{\rho} = \frac{g}{2} \left(h_i^n + \frac{z_{b,i+1/2} - z_{b,i-1/2}}{2} \right)^2 \quad [9.55]$$

Remember that in the Saint Venant and shallow water equations, the momentum equations are divided by the (constant) water density ρ . Conversely, the pressure force per unit width P_2 verifies:

$$\frac{P_2}{\rho} = - \frac{g}{2} \left(h_i^n + \frac{z_{b,i-1/2} - z_{b,i+1/2}}{2} \right)^2 \quad [9.56]$$

The reaction R exerted by the bottom step is obtained by integrating equation [9.53] between $z_{b,i+1/2}$ and $z_{b,i-1/2}$ (the reaction is in the direction of positive x if $z_{b,i-1/2} > z_{b,i+1/2}$):

$$\frac{R}{\rho} = gh_i^n(z_{b,i-1/2} - z_{b,i+1/2}) \quad [9.57]$$

By definition, R/ρ is the integral of the momentum source term over the cell i . Consequently, the source term is given by:

$$S = \begin{bmatrix} 0 \\ R/\rho \end{bmatrix} = \begin{bmatrix} 0 \\ gh_i^n(z_{b,i-1/2} - z_{b,i+1/2}) \end{bmatrix} \quad [9.58]$$

It is easy to check that expressions [9.53] and [9.58] verify the consistency condition and the C -property.

9.5. Balancing techniques

9.5.1. Well-balancing

9.5.1.1. Principle

In the traditional well-balanced approach, the parameter φ in the source term is discontinuous across the cell interfaces (Figure 9.6). This generates a Riemann problem with a discontinuity in the flux function. The solution of such a Riemann problem is not trivial. In [GRE 96], a solution is proposed for a scalar conservation law: the discontinuity in φ is taken as the limit case of a piecewise linear function.

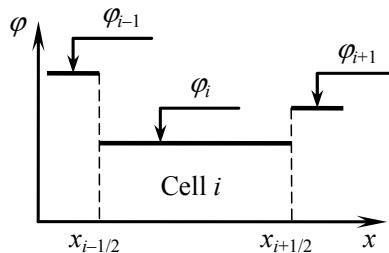


Figure 9.6. Discretization of the parameter φ

When hyperbolic systems of conservation laws are dealt with, the discontinuity in the flux function may induce extra waves in the solution of the Riemann problem compared to a Riemann problem with continuous fluxes. This means that extra, unknown intermediate regions of constant state may appear in the solution, thus requiring additional relationships to close the problem. As an example, exact

solutions to the Riemann problem for the shallow water equations with a bottom step are presented in [ALC 01]. At least 20 different possible wave patterns are identified, against 4 possible patterns for the shallow water equations on a flat bottom. Entropy considerations are used to connect the states on the left- and right-hand sides of the step.

The flux function being discontinuous, the classical finite volume formula [7.3] cannot be used. It must be modified into (here for an explicit formula):

$$U_i^{n+1} = U_i^n + \frac{\Delta t}{\Delta x_i} \left[F_{i-1/2,R}^{n+1/2} - F_{i+1/2,L}^{n+1/2} + \Delta x S(\varphi_{i+k}, U_{i+k}^n) \right] \quad [9.59]$$

where $F_{i-1/2,L}^{n+1/2}$ and $F_{i-1/2,R}^{n+1/2}$ are respectively the fluxes computed on the left- and right-hand side of the interface $i - 1/2$. These fluxes are computed by solving the Riemann problem with left and right states:

$$\left. \begin{array}{l} \frac{\partial U}{\partial t} + \frac{\partial F(U, \varphi)}{\partial x} = 0 \\ U(x, t^n) = U_L \\ \varphi(x) = \varphi_L \end{array} \right\} \text{ for } x < x_{i-1/2} \quad [9.60]$$

$$\left. \begin{array}{l} U(x, t^n) = U_R \\ \varphi(x) = \varphi_R \end{array} \right\} \text{ for } x > x_{i-1/2}$$

Solving this problem generates a flux F_L and F_R on the left- and right-hand sides of the interface.

In [CHA 03], the discretization [9.59] is applied to the shallow water equations for a number of flux splitting-based numerical techniques such as Roe's, Van Leer's, Lax-Friedrichs and Lax-Wendroff's schemes.

9.5.1.2. Application example: the water hammer equations

In the case of the water hammer equations, an analytical solution can be found to the Riemann problem [9.60]. The parameter φ of concern is the cross-sectional area A of the pipe, that is equal respectively to A_L and A_R on the left- and right-hand sides of the discontinuity:

$$\left. \begin{aligned} \frac{\partial U}{\partial t} + \frac{\partial F(U, A)}{\partial x} &= 0 \\ U(x, 0) &= \begin{cases} U_L & \text{for } x < x_0 \\ U_R & \text{for } x > x_0 \end{cases} \\ A(x) &= \begin{cases} A_L & \text{for } x < x_0 \\ A_R & \text{for } x > x_0 \end{cases} \end{aligned} \right\} \quad [9.61]$$

where U and F are given by [2.68]:

$$U = \begin{bmatrix} \rho A \\ \rho Q \end{bmatrix}, \quad F = \begin{bmatrix} \rho Q \\ Ap \end{bmatrix} \quad [9.62]$$

and $d(Ap) = c^2 d(\rho A)$.

When the cross-sectional area of the pipe is constant, the solution of the Riemann problem for the water hammer equations (see Chapter 4) is made of two contact discontinuities separating the left and right states from an intermediate region of constant state (Figure 9.7a). In the case of a piecewise constant cross-sectional area, there are two intermediate regions of constant state (Figure 9.7b).

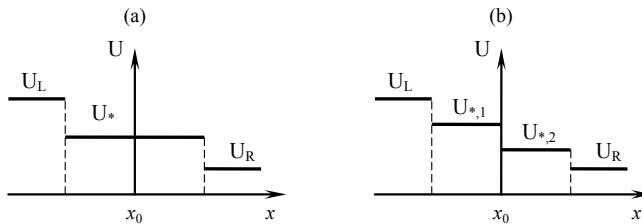


Figure 9.7. Solution of the Riemann problem for the water hammer equations: (a) constant cross-sectional area; (b) piecewise constant cross-sectional area

Applying the characteristic form [2.82] leads to:

$$\left. \begin{aligned} p_{*,1} + \rho c u_{*,1} &= p_L + \rho c u_L \\ p_{*,2} - \rho c u_{*,2} &= p_R - \rho c u_R \end{aligned} \right\} \quad [9.63]$$

where subscripts $*,1$ and $*,2$ indicate respectively the intermediate regions of constant state on the left- and right-hand sides of the discontinuity. Moreover, mass and momentum conservation impose the extra conditions:

$$\left. \begin{aligned} (\rho Q)_{*,1} &= (\rho Q)_{*,2} \\ p_{*,1} &= p_{*,2} \end{aligned} \right\} [9.64]$$

with $Q_{*,1} = A_L u_{*,1}$ and $Q_{*,2} = A_R u_{*,2}$. System [9.63–64] can be solved uniquely for p and Q :

$$\left. \begin{aligned} p_{*,1} = p_{*,2} &= \frac{A_L p_L + A_R p_R}{A_L + A_R} + \rho c \frac{Q_L - Q_R}{A_L + A_R} \\ (\rho Q)_{*,1} = (\rho Q)_{*,2} &= \frac{A_R Q_L + A_L Q_R}{A_L + A_R} + \frac{A_L A_R}{A_L + A_R} \frac{p_L - p_R}{c} \end{aligned} \right\} [9.65]$$

The flux and source terms at the interface are given by:

$$\begin{aligned} F_{i-1/2,L}^{n+1/2} &= \begin{bmatrix} (\rho Q)_{*,1} \\ A_L p_{*,1} \end{bmatrix}, & F_{i-1/2,R}^{n+1/2} &= \begin{bmatrix} (\rho Q)_{*,1} \\ A_R p_{*,1} \end{bmatrix}, \\ S_{i-1/2,L}^{n+1/2} &= \begin{bmatrix} 0 \\ \min(A_R - A_L, 0) p_{*,1} \end{bmatrix}, & S_{i-1/2,R}^{n+1/2} &= \begin{bmatrix} (\rho Q)_{*,1} \\ \max(A_R - A_L, 0) p_{*,1} \end{bmatrix} \end{aligned} [9.66]$$

With this discretization, the C -property is satisfied exactly.

9.5.2. Hydrostatic pressure reconstruction for free surface flow

The hydrostatic reconstruction method is presented in [AUD 05] for shallow water flows. The bottom is assumed constant over each cell, therefore bottom level discontinuities occur only at the interfaces between the cells (Figure 9.8). The method is described for a first-order finite volume scheme hereafter, but a second-order extension can be found in [AUD 05].

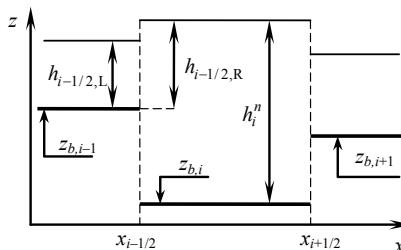


Figure 9.8. Definition sketch for the hydrostatic reconstruction method

In contrast with the approach presented in the previous section, the flux function is continuous across the interfaces, and the source term is exerted at each interface. The equations are discretized in the form:

$$U_i^{n+1} = U_i^n + \frac{\Delta t}{\Delta x_i} \left[F_{i-1/2}^{n+1/2} - F_{i+1/2}^{n+1/2} + \Delta x \left(S_{i-1/2,R}^{n+1/2} + S_{i+1/2,L}^{n+1/2} \right) \right] \quad [9.67]$$

where the source terms $S_{i-1/2,L}^{n+1/2}$ and $S_{i-1/2,R}^{n+1/2}$ are respectively the reaction of the bottom step onto the cell located on the left-hand side of the interface $i - 1/2$ (that is, cell $i - 1$) and the cell located on the right-hand side of interface $i - 1/2$ (that is, cell i). The reaction is assigned to the cell with the lower bottom level.

The flux at the interface $i - 1/2$ is computed by solving an equivalent Riemann problem with modified left and right states $U_{i-1/2,L}$ and $U_{i-1/2,R}$ defined as:

$$U_{i-1/2,L} = \begin{bmatrix} h_{i-1/2,L} \\ h_{i-1/2,L} u_{i-1}^n \end{bmatrix}, \quad U_{i-1/2,R} = \begin{bmatrix} h_{i-1/2,R} \\ h_{i-1/2,R} u_i^n \end{bmatrix} \quad [9.68]$$

where u is the flow velocity and the reconstructed water depths $h_{i-1/2,L}$ and $h_{i-1/2,R}$ on the left- and right-hand sides of the interface are given by:

$$\left. \begin{aligned} h_{i-1/2,L} &= \max(\zeta_{i-1}^n - z_{b,i-1/2}, 0) = \max(h_{i-1}^n + z_{b,i-1} - z_{b,i-1/2}, 0) \\ h_{i-1/2,R} &= \max(\zeta_i^n - z_{b,i-1/2}, 0) = \max(h_i^n + z_{b,i} - z_{b,i-1/2}, 0) \end{aligned} \right\} \quad [9.69]$$

where $z_{b,i-1/2}$ is the higher of the two bottom levels:

$$z_{b,i-1/2} = \max(z_{b,i-1}, z_{b,i}) \quad [9.70]$$

The source term at the interface is given by the integral of the pressure force on the bottom step:

$$\left. \begin{aligned} \Delta x S_{i-1/2,L}^{n+1/2} &= \left[\begin{array}{c} 0 \\ \frac{g}{2} \left(h_{i-1/2,L}^2 - (h_{i-1}^n)^2 \right) \end{array} \right] \\ \Delta x S_{i-1/2,R}^{n+1/2} &= \left[\begin{array}{c} 0 \\ \frac{g}{2} \left((h_i^n)^2 - h_{i-1/2,R}^2 \right) \end{array} \right] \end{aligned} \right\} \quad [9.71]$$

The C -property is satisfied exactly.

9.5.3. Auxiliary variable-based balancing

9.5.3.1. Introduction

As shown by the introductory examples in Section 9.2, flux and source term balancing problems arise from the fact that upwind numerical methods such as flux splitting and Riemann solver-based techniques use a flux function, part of which is proportional to the gradient in the conserved variable. As observed in [BUR 04], when geometry-induced source terms are present in the equations, the gradient in the conserved variable incorporates a geometric term that does not vanish under steady-state conditions. Source term upwinding allows the geometric term to be balanced with the flux gradient at least under static conditions (the C -property). The quasi-steady wave algorithm eliminates this drawback because the geometric parameter does not vary at the interface where the source term is computed. Hydrostatic reconstruction-based schemes also eliminate the problem by using the same value of the geometric parameter on both sides of the interface between the computational cells (the maximum of the two bottom levels in the case of the shallow water equations). Alternative balancing techniques have been explored in the literature:

(1) A first possible option proposed in the literature consists of modifying the system of equations to be solved. In [GAL 03], the shallow water equations on irregular topography are augmented by a third equation involving the bottom level and various alternative variables, such as the hydraulic head, are considered in the discretization of the equations.

(2) A second option consists of reconstructing the variations in the geometric parameter in such a way that steady-state conditions are preserved in the balancing of flux gradients and source terms. This is the case in [KES 10] for example, where the topographic source terms are discretized in Discontinuous Galerkin (DG) techniques (see Chapter 8) by projecting the topographic data onto the space of reconstruction functions.

(3) A third approach consists of estimating the fluxes and source terms in a coupled way. The influence of the source terms is accounted for in the calculation of the fluxes, thereby ensuring the balance between the gradient of the fluxes and the source terms. This is the approach followed by [LHO 07] and [FIN 10] in their derivation of approximate-state solvers for the water hammer equations, the shallow water equations and the shallow water equations with porosity.

(4) A fourth path, that is well-suited to finite volume techniques, consists of defining “auxiliary” variables to be used in the estimates of the gradients for the calculation of the fluxes. Such auxiliary variables do not need to be conserved variables. They must be defined so as to allow the C -property and/or steady-state conditions to be preserved. As an example, [NUJ 95] proposes that the free surface elevation ζ be used instead of the water depth h in the calculation of the mass flux

for the shallow water equations. The authors of [ZHO 01] note that reconstructing the free surface elevation instead of the water depth enhances the stable character of higher-order schemes and allows spurious oscillations to be eliminated from the numerical solutions of the shallow water equations. [LIA 09] proposes a free surface elevation-based discretization of the pressure terms. Although not justified theoretically in Nujic's original publication in 1995, the use of the free surface elevation instead of the water depth for the shallow water equations can be justified as follows: since z_b is constant, the free surface elevation $\zeta = h + z_b$ verifies $\partial\zeta/\partial t = \partial h/\partial t$. Therefore, ζ may be used as the first component of the conserved variable U instead of h . The free surface elevation is also used instead of the conserved variable ϕh in the solution of the two-dimensional shallow water equations with a variable porosity ϕ [GUI 06]. In [BUR 04], the open channel equations are solved by expressing the variation in the cross-sectional area as a function of momentum balance. A similar formula is proposed in [LEE 10] for the HLL and Roe's Riemann solver (see Appendix C). This formula coincides with those of an approximate-state Riemann solver for the shallow water equation [LHO 07, FIN 10]. This leads us to wonder whether a general methodology can be used to define auxiliary flow variables that preserve the C -property and steady-state flow conditions more efficiently than the conserved variables do. This option is explored in the following sections. The approach is called "auxiliary variable-based balancing" for the sake of convenience.

9.5.3.2. Principle of Auxiliary Variable-based (AV) balancing

As mentioned in Appendix C and underlined by a number of authors, flux estimates based on classical Riemann solvers such as the HLL/HLLC, Roe's, the Lax-Friedrichs solver, etc. can be expressed as the sum of a centered flux (that is unconditionally unstable when used with first- or second-order time stepping [VIC 82]) and a diffusive flux (that contributes to stabilize the numerical solution):

$$F(U_L, U_R) = aF_L + (1 - a)F_R + D(U_L - U_R) \quad [9.72]$$

where a is a coefficient between 0 and 1, and D is a diffusion matrix that depends on the flux formula used. For instance, in the discretization [9.5] provided for the water hammer equations (section 9.2.1), $a = 1/2$ and D is a diagonal matrix equal to $c/2$ times the identity matrix:

$$D = \frac{c}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \frac{c}{2} I \quad [9.73]$$

In the HLL Riemann solver used for the shallow water equations example of section 9.2.2:

$$\left. \begin{aligned} a &= \frac{\lambda^+}{\lambda^+ - \lambda^-} \\ D &= -\frac{\lambda^- \lambda^+}{\lambda^+ - \lambda^-} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = -\frac{\lambda^- \lambda^+}{\lambda^+ - \lambda^-} I \end{aligned} \right\} [9.74]$$

Although playing a crucial role to solution stability in the absence of source terms, the diffusive flux fails to preserve the C -property, as shown in the examples of section 9.2. The question then arises whether the diffusive flux may be formulated in a slightly different way, such that the C -property is preserved:

$$F(U_L, U_R) = a F_L + (1-a) F_R + D_V (V_L - V_R) \quad [9.75]$$

where D_V is a diffusion matrix, not necessarily equal to D , and V_L and V_R are appropriately defined auxiliary variables, functions of U and φ :

$$\left. \begin{aligned} V_L &= V(U_L, \varphi_L) \\ V_R &= V(U_R, \varphi_R) \end{aligned} \right\} [9.76]$$

V may also be defined in differential form as:

$$dV = dV(dU, \varphi) = \left(\frac{\partial V}{\partial U} \right)_{x=\text{Const}} dU + \left(\frac{\partial V}{\partial \varphi} \right)_{U=\text{Const}} \frac{\partial \varphi}{\partial x} dx \quad [9.77]$$

The term $(\partial V / \partial U)_{x=\text{Const}} dU$ accounts for the variations in the conserved variable U due to the gradients in the flow conditions, while the term $(\partial V / \partial x)_{U=\text{Const}} \partial \varphi / \partial x d\varphi$ accounts for the influence of the geometric source terms. The differential definition [9.77] is more convenient than the definition [9.76] when the definition of V contains terms that cannot be recast in the form of the differential of a variable (as it is the case with source terms). Assuming that the variable V is defined appropriately, the diffusion matrix D_V must be determined. A desirable property for this matrix is that it does not change the stability constraints of the original discretization. This means that the strength of the diffusion term $D_V (V_L - V_R)$ should be identical to that of the original diffusion term $D (U_L - U_R)$. The diffusive term in equation [9.72] is rewritten as (subscripts $x = \text{Const}$ and $U = \text{Const}$ are omitted for the sake of clarity):

$$\left. \begin{aligned} D(U_L - U_R) &\approx -\Delta x D \frac{\partial U}{\partial x} + O(\Delta x^2) \\ D_V(V_L - V_R) &= -\Delta x D_V \frac{\partial V}{\partial U} \frac{\partial U}{\partial x} - \Delta x D_V \frac{\partial V}{\partial \varphi} \frac{\partial \varphi}{\partial x} + O(\Delta x^2) \end{aligned} \right\} [9.78]$$

The stability properties of the scheme are related to the first-order derivative. In the absence of source terms, equation [9.78] reduces to:

$$\left. \begin{aligned} D(U_L - U_R) &\approx -\Delta x D \frac{\partial U}{\partial x} + O(\Delta x^2) \\ D_V(V_L - V_R) &= -\Delta x D_V \frac{\partial V}{\partial U} \frac{\partial U}{\partial x} + O(\Delta x^2) \end{aligned} \right\} [9.79]$$

The diffusion terms in equation [9.79] are equivalent provided that:

$$D_V \frac{\partial V}{\partial U} = D \quad \Leftrightarrow \quad D_V = D \left(\frac{\partial V}{\partial U} \right)^{-1} [9.80]$$

9.5.3.3. Discretization of the momentum source term

Estimate [9.75] satisfies steady-state conditions in the continuity equation. Indeed, under steady-state conditions, the first components $F_{1,L}$ and $F_{1,R}$ of F_L and F_R are identical, $F_{1,L} = F_{1,R} = q$. If V is chosen appropriately, $V_L = V_R$ for steady flow configurations, and the first component of equation [9.75] becomes:

$$F_1 = aF_{1,L} + (1-a)F_{1,R} = q [9.81]$$

In contrast with the source term upwinding technique, no extra source term needs to be added to the continuity equation. The issue remains, however, of the discretization of the momentum source term. Numerical experiments (see section 9.6) show that using the source term upwinding technique only for the momentum source term provides good results.

9.5.3.4. Application to the water hammer equations

Consider the water hammer equations in a non-horizontal pipe with variable cross-sectional area. The angle of the pipe with the horizontal is denoted by θ and the friction coefficient is denoted by k . The discretization [9.5] is used, with D given by equation [9.75]. The conserved variable is $U = [\rho A, \rho Q]^T$. A possible choice for the auxiliary variable V is:

$$dV = \begin{bmatrix} dp + (\rho g \sin \theta + k|u|u/A) dx \\ d(\rho Q) \end{bmatrix} \quad [9.82]$$

This choice is motivated by the fact that under steady-state conditions, the vector dV is zero. Indeed, the first component expresses steady-state energy conservation (the slope of the energy line is equal to the head loss per unit length) while the second component accounts for mass conservation. Therefore, definition [9.82] not only verifies the C -property exactly, but it also verifies steady-state conditions. The following estimate is obtained at the interface $i - 1/2$:

$$\begin{aligned} \Delta x_{i-1/2} \left(\frac{\partial V}{\partial x} \right)_{i-1/2} &= \Delta x_{i-1/2} \begin{bmatrix} \frac{\partial p}{\partial x} + \rho g \sin \theta + \frac{k|u|u}{A} \\ \frac{\partial}{\partial x}(\rho Q) \end{bmatrix}_{i-1/2} \\ &\approx \begin{bmatrix} p_i^n - p_{i-1}^n + \rho g(z_i - z_{i-1}) + \Delta x_{i-1/2} \left(\frac{k|u|u}{A} \right)_{i-1/2}^n \\ (\rho Q)_i^n - (\rho Q)_{i-1}^n \end{bmatrix} \end{aligned} \quad [9.83]$$

where z_i is the elevation of the cell i and $\Delta x_{i-1/2}$ is the distance between the centers of the cells $i - 1$ and i . The Jacobian matrix of V with respect to U is given by (the second equation [2.74] is used):

$$\frac{\partial V}{\partial U} = \begin{bmatrix} \partial p / \partial(\rho A) & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} c^2 \partial p / \partial(Ap) & 0 \\ 0 & 1 \end{bmatrix} \approx \begin{bmatrix} c^2 / A & 0 \\ 0 & 1 \end{bmatrix} \quad [9.84]$$

Substituting equations [9.73] and [9.84] into equation [9.80] leads to the following expression for the matrix D_V :

$$D_V = \frac{1}{2} \begin{bmatrix} A/c & 0 \\ 0 & c \end{bmatrix} \quad [9.85]$$

and the final discretization becomes:

$$F_{i-1/2}^{n+1/2} = \frac{1}{2} \begin{bmatrix} (\rho Q)_{i-1}^n + (\rho Q)_i^n \\ (Ap)_{i-1}^n + (Ap)_i^n \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \frac{A_{i-1/2}}{c} \Delta p_{i-1/2}^n \\ c[(\rho Q)_{i-1}^n - (\rho Q)_i^n] \end{bmatrix} \quad [9.86]$$

where $A_{i-1/2}$ is an average value of A between the cells $i-1$ and i , and the variation Δp is estimated as:

$$\Delta p_{i-1/2}^n = p_{i-1}^n - p_i^n + \rho g(z_{i-1} - z_i) - \frac{\Delta x_{i-1/2}}{2} \left[\left(\frac{k|u|u}{A} \right)_{i-1}^n + \left(\frac{k|u|u}{A} \right)_i^n \right] \quad [9.87]$$

Examples of possible estimates for $A_{i-1/2}$ are:

$$\left. \begin{aligned} A_{i-1/2} &= \frac{A_{i-1/2} + A_{i-1/2}}{2} \\ A_{i-1/2} &= \min(A_{i-1/2}, A_{i-1/2}) \\ A_{i-1/2} &= (A_{i-1/2} A_{i-1/2})^{1/2} \end{aligned} \right\} \quad [9.88]$$

These are not the only possible estimates. Any estimate satisfying the consistency condition [9.46] may be used.

9.5.3.5. Application to the shallow water equations

Consider discretization [9.35] presented in section 9.3.3 for the shallow water equations discretized with the HLL Riemann solver. The components of the flux are given by:

$$\left. \begin{aligned} q_{i-1/2}^n &= \frac{\lambda^+ q_{i-1}^n - \lambda^- q_i^n}{\lambda^+ - \lambda^-} - \frac{\lambda^- \lambda^+}{\lambda^+ - \lambda^-} \Delta h_{i-1/2} \\ M_{i-1/2}^n &= \frac{\lambda^+ M_{i-1}^n - \lambda^- M_i^n}{\lambda^+ - \lambda^-} - \frac{\lambda^- \lambda^+}{\lambda^+ - \lambda^-} (q_{i-1}^n - q_i^n) \end{aligned} \right\} \quad [9.89]$$

where $\Delta h_{i-1/2}$ is an approximation of the difference in h that satisfies the C -property. The source term contributions are changed to:

$$\left. \begin{aligned} \beta^{(1)} \mathbf{K}^{(1)} &= \frac{gh_{i-1}^{n+1/2}}{2\lambda^+} (h_i^n - h_{i-1}^n) \begin{bmatrix} 0 \\ \lambda^+ \end{bmatrix} \\ \beta^{(2)} \mathbf{K}^{(2)} &= \frac{gh_i^{n+1/2}}{2\lambda^+} (h_i^n - h_{i-1}^n) \begin{bmatrix} 0 \\ \lambda^+ \end{bmatrix} \end{aligned} \right\} \quad [9.90]$$

The only remaining problem is to define a suitable estimate for $\Delta h_{i-1/2}$ in the first equation [9.90]. Note that the classical estimate:

$$\Delta h_{i-1/2}^{(0)} = h_{i-1}^n - h_i^n \quad [9.91]$$

does not satisfy the C -property.

(1) A first option is Nujic's variable transformation [NUJ 95], $\zeta = h + z_b$:

$$\mathbf{V} = \begin{bmatrix} h + z_b \\ q \end{bmatrix}, \quad \frac{\partial \mathbf{V}}{\partial \mathbf{U}} = \mathbf{I}, \quad D_V = D \quad [9.92]$$

This leads to the following estimate for Δh in equation [9.89]:

$$\Delta h_{i-1/2}^{(1)} = \zeta_{i-1}^n - \zeta_i^n \quad [9.93]$$

The C -property is verified. However, arbitrary steady-state configurations are not preserved. Consider for instance uniform flow conditions over a non-zero bottom slope. Then both the water depth h and the unit discharge q are constant. Since the slope is non-zero, ζ is not constant and the first equation [9.88] yields a flux that is different from the average cell value.

(2) A second possible choice for \mathbf{V} is proposed in [BUR 04] and [LEE 10]:

$$d\mathbf{V} = \begin{bmatrix} dM - (S_0 - S_f)gh dx \\ dq \end{bmatrix}, \quad \frac{\partial \mathbf{V}}{\partial \mathbf{U}} = \begin{bmatrix} c^2 - u^2 & 2u \\ 0 & 1 \end{bmatrix} \quad [9.94]$$

Note that the first row of the Jacobian matrix $\partial \mathbf{V} / \partial \mathbf{U}$ is obtained using the relationship $dM = (c^2 - u^2) dh + 2u dq$ and by excluding the integral in the differentiation of \mathbf{V} . This can be justified by the fact that the integral is a function of x and that its derivative with respect to x is non-zero even if \mathbf{V} is constant. Equation [9.94] leads to the following estimate for Δh :

$$\begin{aligned} \Delta h_{i-1/2}^{(2)} &= \frac{M_{i-1}^n - M_i^n - 2u_{i-1/2}(q_{i-1}^n - q_i^n) + (z_{b,i-1} - z_{b,i} - \Delta x S_f)}{(c^2 - u^2)_{i-1/2}} \\ &= -\frac{M_{i-1}^n - M_i^n - (\lambda^- + \lambda^+)(q_{i-1}^n - q_i^n) + (z_{b,i-1} - z_{b,i} - \Delta x S_f)}{\lambda^- \lambda^+} \end{aligned} \quad [9.95]$$

This formula coincides with that of the approximate-state Riemann solver presented in [LHO 07]. [BUR 04] and [LHO 07] acknowledge the need for a

specific treatment of critical points, that correspond to $\lambda^- = 0$ or $\lambda^+ = 0$. This yields a division by zero in formula [9.95]. [BUR 04] proposed that the final estimate for Δh be the minmod of the estimates given by equations [9.90] and [9.95]. [LHO 07] and [FIN 10] proposed that the estimate of the flux be based on a characteristics-based estimate of the flow variables under critical conditions.

When combined with the HLL or Roe's solver, formula [9.95] is seen to introduce a downwinding effect on the unit discharge in the neighborhood of critical points [FIN 10]. This can be avoided by defining dV as:

$$dV = \begin{bmatrix} dM - 2u dq - (S_0 - S_f)gh dx \\ dq \end{bmatrix}, \quad \frac{\partial V}{\partial U} = \begin{bmatrix} c^2 - u^2 & 0 \\ 0 & 1 \end{bmatrix} \quad [9.96]$$

thus eliminating the difference $q_{i-1}^n - q_i^n$ from the numerator in equation [9.95]. The following estimate is obtained:

$$\Delta h_{i-1/2}^{(3)} = -\frac{M_{i-1}^n - M_i^n + (z_{b,i-1} - z_{b,i} - \Delta x S_f)}{\lambda^- \lambda^+} \quad [9.97]$$

(3) A third possibility is to use the hydraulic head $H = \zeta + u^2/(2g)$ instead of M in the first component of V :

$$dV = \begin{bmatrix} dH + S_f dx \\ dq \end{bmatrix}, \quad \frac{\partial V}{\partial U} = \begin{bmatrix} c^2 - u^2 & u \\ 0 & 1 \end{bmatrix} \quad [9.98]$$

The first row in the Jacobian matrix is obtained using the relationship $dH = (1 - u^2/c^2) dh + u dq + dz_b$. The following estimate is obtained for Δh :

$$\Delta h_{i-1/2}^{(4)} = -c_{i-1/2}^2 \frac{H_{i-1}^n - H_i^n - z_{b,i-1} + z_{b,i} + \frac{\lambda^- + \lambda^+}{2}(q_i^n - q_{i-1}^n) + \Delta x S_f}{\lambda^- \lambda^+} \quad [9.99]$$

For the same reason as in option (2), the following simplification is proposed for transcritical flow:

$$dV = \begin{bmatrix} dH - u dq + S_f dx \\ dq \end{bmatrix}, \quad \frac{\partial V}{\partial U} = \begin{bmatrix} c^2 - u^2 & 0 \\ 0 & 1 \end{bmatrix} \quad [9.100]$$

thus leading to:

$$\Delta h_{i-1/2}^{(4)} = -c_{i-1/2}^2 \frac{H_{i-1}^n - H_i^n - z_{b,i-1} + z_{b,i} + \Delta x S_f}{\lambda^- \lambda^+} \quad [9.101]$$

9.6. Computational example

The performance of the various source term discretization techniques presented in the previous sections is illustrated by an application to the one-dimensional shallow water equations. The following steady-state configuration is considered: water flows with a constant unit discharge in a channel with a non-horizontal bottom (Figure 9.9). The bottom level forms a bump. It is obtained from the following equation:

$$z_b(x) = Z \exp \left[- \left(\frac{x - x_0}{X} \right)^2 \right] \quad [9.102]$$

The water is initially at rest, with a constant free surface elevation and a zero discharge at all points of the computational domain. A constant unit discharge and water level are prescribed at the upstream and downstream end of the channel respectively.

The numerical solution is examined after a sufficiently long simulated time, so that the transient regime vanishes and steady state is reached. The constant unit discharge and downstream water level are chosen such that the flow regime is subcritical upstream of the bump, supercritical on the top part of the bump and a hydraulic jump appears on the downstream side of the bump.

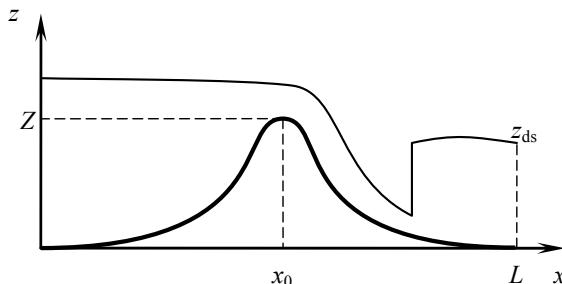


Figure 9.9. Steady-state flow configuration. Bold line: bottom level.
Thin line: free surface flow elevation

The parameters of the test case are given in Table 9.1. The numerical solution is computed using the first-order Godunov scheme, with left and right states of the Riemann problems at the cell interfaces taken equal to the average values of the variables over the computational cells. The HLL solver presented in Appendix C is used in the computation of the fluxes at the interfaces between the computational cells.

Figure 9.10 shows the longitudinal profiles for the water level and unit discharge computed at $t = 10^4$ seconds using the source term upwinding approach. Under steady-state conditions, the unit discharge should be expected to be uniform over the entire domain. This is not the case with the numerical solution: the transition from subcritical to supercritical conditions is observed to induce strong variations in the cell average of the unit discharge.

Note that this should not be attributed to continuity issues, because the finite volume technique is intrinsically conservative. The unit discharge profile plotted in Figure 9.10 is the cell average of the quantity $q = hu$, that is, the momentum in the computational cells, and not the value of the mass flux at the interfaces between the cells. Under steady-state conditions, the unit discharge at the interfaces between the computational cells is uniform all over the computational domain.

Symbol	Meaning	Value
g	Gravitational acceleration	9.81 m/s ²
L	Length of the computational domain	400 m
n_M	Manning's friction coefficient	0.01
q_{us}	Upstream discharge	1 m ² /s
X	Characteristic size of the bump (equation [9.102])	50 m
x_0	Abscissa of the center of the bump	200 m
Z	Height of the bump	1 m
z_{ds}	Water level at the downstream boundary	1 m
Δx	Computational cell size	2 m
ζ	Initial water level	1.01 m

Table 9.1. Parameters of the test case

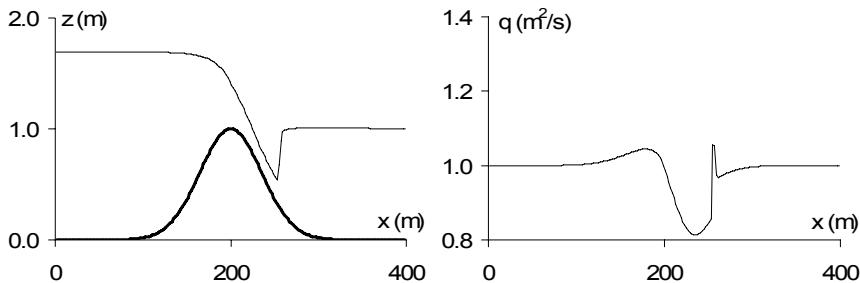


Figure 9.10. Steady-state water level and unit discharge profiles computed by the source term upwinding method. Bold line (left): bottom level

Also note that the computed discharge profile exhibits a peak at the location of the jump. This is because the water depth varies abruptly across the jump. Since the unit discharge at the cell interfaces is given by an equation in the form [9.72], q_L and q_R cannot be identical to the unit discharge at the interface if $h_L \neq h_R$. A higher-order reconstruction technique would be needed to reconstruct the left and right states at each interface more accurately and obtain a constant discharge.

Figure 9.11 shows the water level and unit discharge profiles computed using the quasi steady wave propagation algorithm. Both the extent and amplitude of the zone of non-uniform unit discharge are reduced compared to the source term upwinding method, except for the peak that corresponds to the hydraulic jump.

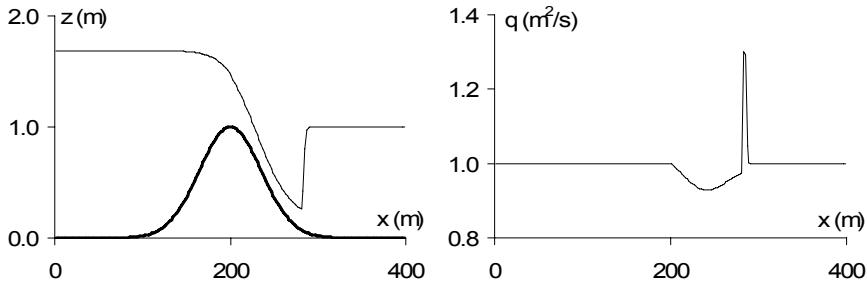


Figure 9.11. Steady-state water level and unit discharge profiles computed by the quasi-steady wave propagation method. Bold line (left): bottom level

Figure 9.12 shows the free surface and unit discharge profiles obtained using the hydrostatic reconstruction method. The amplitude of the discharge peak at the

location of the hydraulic jump is similar to that in Figure 9.11, but the amplitude of the non-uniform discharge zone upstream of the jump is smaller in Figure 9.12.

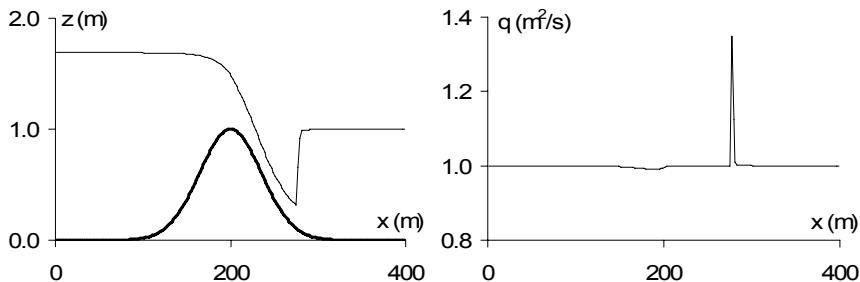


Figure 9.12. Steady-state water level and unit discharge profiles computed by the hydrostatic reconstruction method

Figures 9.13 and 9.14 show the profiles obtained using the auxiliary variable method.

Figure 9.13a shows the result given by Nujic's method [NUJ 95], equation [9.93]. The quality of the numerical solution is similar to that of the hydrostatic reconstruction method, but the discharge peak is smaller.

Figure 9.13b shows the results obtained from the momentum-based method proposed in [BUR 04]. As in [BUR 04], $\Delta h_{i-1/2}$ is computed as the minmod of the estimates [9.91] and [9.95]. The amplitude of the peak is reduced compared to that in Figure 9.13a, but a small variation in the unit discharge is still visible upstream of the bump.

This variation is eliminated when $\Delta h_{i-1/2}$ is computed as the minmod of [9.91] and [9.97] (see Figure 9.13c). This, however, is achieved at the expense of a sharper oscillation at the location of the hydraulic jump. An optimally accurate method should therefore use estimate [9.95] at the shock and [9.97] at other points.

Figure 9.14a shows the profiles obtained using the minmod of [9.91] and the hydraulic head-based estimate [9.99]. Figure 9.14b shows the profiles obtained from the minmod of [9.91] and the hydraulic head-based estimate [9.101]. Both methods give similar results in the neighborhood of the shock. However, estimate [9.101] is better upstream of the bump.

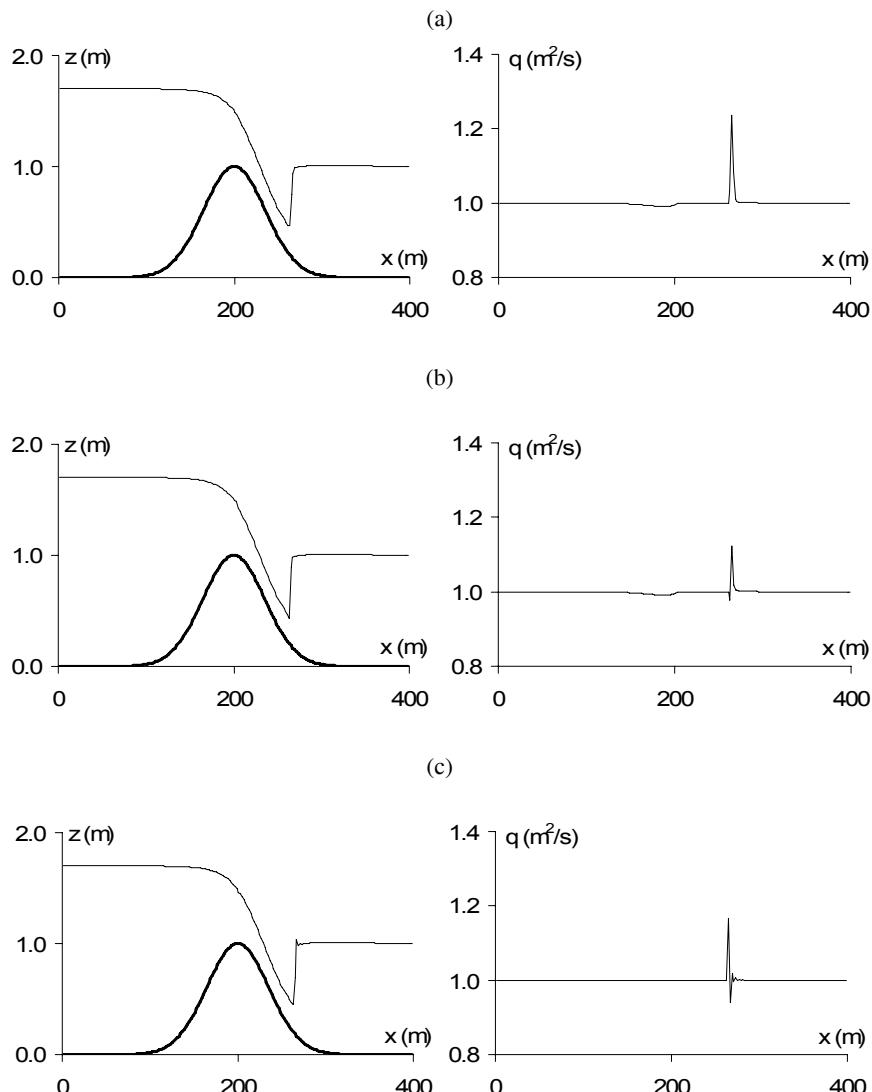


Figure 9.13. Steady-state water level and unit discharge profiles computed by the auxiliary variable-based method. (a) $\Delta h_{i-1/2} = \Delta h_{i-1/2}^{(1)}$, (b) $\Delta h_{i-1/2} = \min \text{mod}[\Delta h_{i-1/2}^{(0)}, \Delta h_{i-1/2}^{(2)}]$, (c) $\Delta h_{i-1/2} = \min \text{mod}[\Delta h_{i-1/2}^{(0)}, \Delta h_{i-1/2}^{(3)}]$

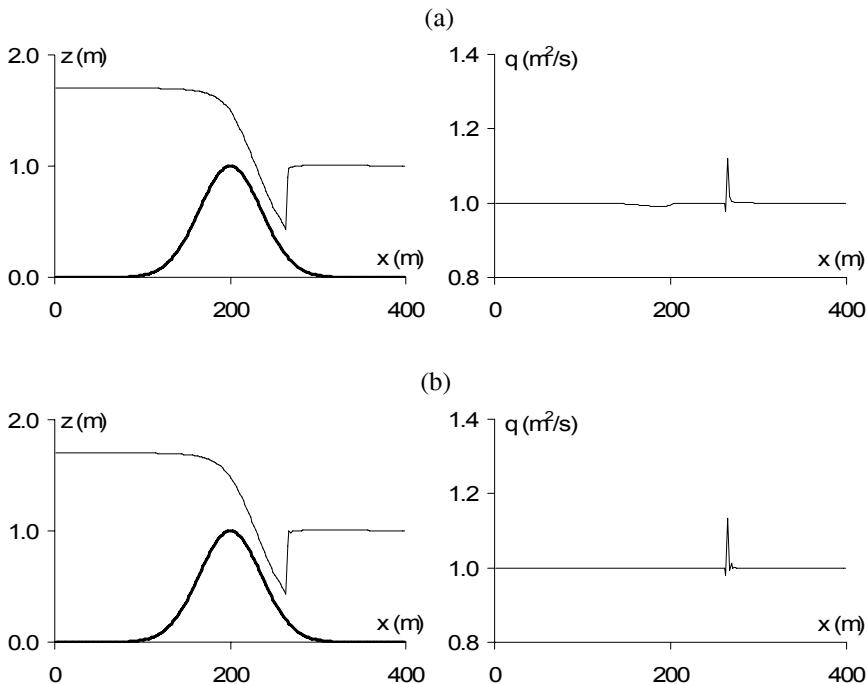


Figure 9.14. Steady-state water level and unit discharge profiles computed by the auxiliary variable-based method. (a) $\Delta h_{i-1/2} = \min \text{mod}[\Delta h_{i-1/2}^{(0)}, \Delta h_{i-1/2}^{(4)}]$,
(b) $\Delta h_{i-1/2} = \min \text{mod}[\Delta h_{i-1/2}^{(0)}, \Delta h_{i-1/2}^{(5)}]$

9.7. Summary

Real-world applications of computational hydraulics or computational fluid dynamics involve the discretization of geometry-induced source terms. Discretizing such source terms independently from the fluxes may induce stability problems. This is due to a lack of balance between the second-order (diffusion-like) terms in the discretization of both fluxes and source terms.

The notion of C -property plays an essential role in the definition of well-balanced schemes. The C -property states that any initial condition verifying static equilibrium conditions should yield a static solution at later times for arbitrary geometries. A discretization of the flow equations may satisfy the C -property exactly or approximately. The C -property may be extended to steady-state flow conditions (of which static equilibrium is only a particular case).

A number of methods are presented in this chapter for geometric source term discretization in finite volume techniques:

– Source term upwinding (section 9.3) uses a decomposition of the source term in the base of eigenvectors of the Jacobian matrix A . The source term is thus broken into as many components as there are waves. Each component is assigned to the computational cell into which the wave originating from the cell interface travels. Some of the flow variables must be estimated at the interfaces between the computational cells. The estimates must be derived in such a way that the C -property be verified.

– The quasi-steady wave algorithm (section 9.4) uses a particular reconstruction of the geometry. The variations in the geometric parameter are lumped in the middle of the computational cell. The parameter is continuous at the cell interfaces, which allows a standard Riemann problem (that is, with a continuous flux function and no source term) to be defined and solved.

– In well-balancing techniques (section 9.5.1), a solution is sought for the Riemann problem with a discontinuous flux function and source term. In the original publication of the method, the discontinuity is achieved as the limit case of a continuous, ramp-shaped function.

– Balancing techniques such as the hydrostatic reconstruction (section 9.5.2) consist of modifying the states of the Riemann problem at the interface between two cells. The cross-sectional areas (or water depth) on both sides of the interface are modified using the water level and the higher of the bottom levels on both sides of the interface. The unit discharge is re-computed as the product of the flow velocity in the cell and the modified water depth.

– The auxiliary variable-based balancing technique (section 9.5.3) consists of redefining the flow variables used in the diffusive part of the flux estimate. These so-called auxiliary variables are defined so as to achieve a zero gradient under steady-state conditions, which allows the C - and extended C -property to be satisfied.

Chapter 10

Sensitivity Equations for Hyperbolic Systems

10.1. Introduction

Most conservation laws presented in the previous chapters involve flux functions or source terms that are functions of the conserved variable U and a number of parameters that are known *a priori*. For instance, the flux F in linear advection equations (section 1.3) is equal to the product of the conserved variable AC by the flow velocity u . The knowledge of u is required to solve the equation. In the kinematic wave equations (section 1.5), the friction coefficient, channel slope and channel geometry are parameters on which the discharge Q depends. In the Buckley-Leverett equation (section 1.6), the flux F is a function of the conserved variable s , the shape parameter b_{BL} and the Darcy velocity V_d . In the advection-adsorption equation (section 1.7), the parameters are the flow velocity and the adsorption constants (k_{lin} for a linear law, k_F and b in a Freundlich model, k_L and C_L in a Langmuir-based model). The initial conditions may also be considered as a parameter in the problem that consists of solving the equations over time.

In practical engineering applications, the parameters are known with a given imprecision or uncertainty. It is sometimes impossible to measure these parameters directly, hence the need for a “calibration” of the model. Calibration consists of adjusting or constraining the parameters so that the model output reproduces measurements or observations as faithfully as possible. In some other cases, such as scenario analysis, the modeler is interested in knowing the consequences of imprecision or uncertainty in the knowledge of the parameters on the model output.

In optimization and control problems, we are interested in the parameter values that allow a certain objective (or cost) function to be minimized.

In all these problems, the sensitivity of the solution to the model parameters (that is, how the output variables change with the parameters) plays an essential role. For instance, trying to “calibrate” a model parameter to which the model output is not sensitive would be meaningless. In contrast, oversensitivity of a model to a given parameter may indicate the hidden influence of additional parameters that have not been identified. Model sensitivity is used in optimization problems for optimum search. In uncertainty analysis, it can be used in the framework of efficient, first-order second moment techniques. In the field of flow control, it is often used in the framework of adjoint modeling.

Sensitivity is classically defined as the partial derivative of a model variable with respect to a parameter. Cacuci [CAC 03] defines the sensitivity as a Gateaux (also called directional) derivative. Several techniques may be used to compute the sensitivity of a model variable to a parameter:

- The empirical, or finite difference approach, consists of carrying out a simulation using two slightly different values of the parameter of interest. The sensitivity is defined as the limit of the ratio of the variation in the variable to the variation in the parameter. This approach is very efficient in estimating the sensitivity of a model, the governing equations of which are unknown. It may also be used in the local sensitivity analysis of models with known governing equations, such as hyperbolic systems of conservation laws.

- The complex differentiation technique [LYN 67] consists of introducing a pure imaginary perturbation in the parameter of interest (see [LU 07] for an application). The solution of the governing equations is a complex number, the imaginary part of which is used to compute the sensitivity. This technique is second-order accurate with respect to the perturbation in the parameter. It has the advantage that only one run of the model is needed. The drawback is that all the mathematical operations in the model must be carried out on complex numbers, which is computationally expensive.

- Code differentiation consists of differentiating the programmed instructions instead of the governing equations in the informatic implementation of the model. Code differentiation may be manual or automatic [ELI 07].

- Analytical differentiation consists of differentiating the analytical solution of the governing equations with respect to the parameter. This implies that (i) an analytical solution is available, which is not often the case in real-world problems, and (ii) the parameter is assumed to be perturbed uniformly over the entire solution domain. This is another constraining assumption.

– Direct sensitivity calculation consists of differentiating the governing equations of the model with respect to the parameter of interest, thus yielding a set of governing sensitivity equations. The sensitivity equations are solved in a coupled way with the original model equations. This approach is examined in the present chapter. Two types of sensitivity equations are considered hereafter: forward (or direct) sensitivity equations, and adjoint (or backward) sensitivity equations.

A particular aspect of sensitivity equations for hyperbolic systems is that the solutions of the hyperbolic system may become locally discontinuous. In this case, the derivatives of the solution with respect to time, space and the parameter of interest become locally undefined and the sensitivity equations must be modified because sensitivity becomes a locally non-conserved variable (see section 10.2.2). Classical numerical techniques yield unstable sensitivity solutions [GUN 99]. The governing sensitivity equations must be modified to account for this (see [BAR 02], where the governing sensitivity equations are reformulated in the framework of the theory of distributions).

10.2. Forward sensitivity equations for scalar laws

10.2.1. Derivation for continuous solutions

Consider a one-dimensional scalar hyperbolic equation written in the conservation form [1.1], recalled here:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} = S$$

where U , F and S are respectively the conserved variable, the flux and the source term. In what follows, F and S are functions in the form:

$$\left. \begin{aligned} F &= F(U, \varphi) \\ S &= S(U, \varphi) \end{aligned} \right\} \quad [10.1]$$

where φ is a parameter with respect to which the sensitivity analysis is to be carried out. φ may be a parameter in the conservation law, in the initial or boundary conditions.

The purpose is to study the influence of variations in φ on the solution U over the solution domain. This is done via a perturbation analysis. The parameter φ is expressed in the form:

$$\varphi(x, t) = \varphi_0(x, t) + \varepsilon(x, t)\varphi' \quad [10.2]$$

where φ_0 and φ' are respectively the “nominal” value of the parameter and φ' is an infinitesimal perturbation. $\epsilon(x, t)$ is called the support, or characteristic function of the perturbation. It expresses the fact that the parameter may not be perturbed with the same magnitude at all points and all times of the solution domain. The solution of equation [1.1] with F , S and φ as defined in equations [10.1–2] is written in the form:

$$U(x, t) = U_0(x, t) + U'(x, t) \quad [10.3]$$

where U_0 is the solution of equation [1.1] for $\varphi = \varphi_0$ and U' is the perturbation in U caused by the perturbation in φ . The sensitivity of U with respect to φ is defined as the limit:

$$s(x, t) = \lim_{\varphi' \rightarrow 0} \frac{U'(x, t)}{\varphi'} \quad [10.4]$$

The governing equation for s is obtained by writing two equations [1.1] with two different values of φ' :

$$\left. \begin{aligned} \frac{\partial U_0}{\partial t} + \frac{\partial}{\partial x} F(U_0, \varphi_0) &= S(U_0, \varphi_0) \\ \frac{\partial}{\partial t}(U_0 + U') + \frac{\partial}{\partial x} F(U_0 + U', \varphi_0 + \epsilon\varphi') &= S(U_0 + U', \varphi_0 + \epsilon\varphi') \end{aligned} \right\} \quad [10.5]$$

A first-order Taylor series expansion yields:

$$\left. \begin{aligned} F(U_0 + U', \varphi_0 + \epsilon\varphi') &\approx F(U_0, \varphi_0) + \frac{\partial F}{\partial U} U' + \frac{\partial F}{\partial \varphi} \epsilon\varphi' \\ S(U_0 + U', \varphi_0 + \epsilon\varphi') &\approx S(U_0, \varphi_0) + \frac{\partial S}{\partial U} U' + \frac{\partial S}{\partial \varphi} \epsilon\varphi' \end{aligned} \right\} \quad [10.6]$$

Substituting equations [10.6] into equations [10.5] and subtracting yields:

$$\frac{\partial U'}{\partial t} + \frac{\partial}{\partial x} \left[\frac{\partial F}{\partial U} U' + \frac{\partial F}{\partial \varphi} \epsilon\varphi' \right] = \frac{\partial S}{\partial U} U' + \frac{\partial S}{\partial \varphi} \epsilon\varphi' \quad [10.7]$$

Dividing equation [10.7] by the perturbation φ' and introducing definition [10.4] gives:

$$\frac{\partial s}{\partial t} + \frac{\partial}{\partial x} \left[\frac{\partial F}{\partial U} s + \frac{\partial F}{\partial \varphi} \varepsilon \right] = \frac{\partial S}{\partial U} s + \varepsilon \frac{\partial S}{\partial \varphi} \quad [10.8]$$

Noting from Chapter 1 that $\partial F / \partial U$ is defined as the wave speed λ , equation [10.8] becomes:

$$\frac{\partial s}{\partial t} + \frac{\partial G}{\partial x} = \frac{\partial S}{\partial U} s + \varepsilon \frac{\partial S}{\partial \varphi} - \frac{\partial}{\partial x} \left(\varepsilon \frac{\partial F}{\partial \varphi} \right), \quad G = \lambda s \quad [10.9]$$

10.2.2. Conservation, non-conservation and characteristic forms

Equations [1.1] and [10.9] can be written in vector conservation form as:

$$\frac{\partial V}{\partial t} + \frac{\partial H}{\partial x} = T \quad [10.10]$$

where V , H and T are defined as:

$$V = \begin{bmatrix} U \\ s \end{bmatrix}, \quad H = \begin{bmatrix} F \\ G \end{bmatrix} = \begin{bmatrix} F \\ \lambda s \end{bmatrix}, \quad T = \begin{bmatrix} S \\ Q \end{bmatrix} = \left[\frac{\partial S}{\partial U} s + \varepsilon \frac{\partial S}{\partial \varphi} - \frac{\partial}{\partial x} \left(\varepsilon \frac{\partial F}{\partial \varphi} \right) \right] \quad [10.11]$$

The non-conservation form of [10.10] is obtained by differentiating H with respect to V :

$$\frac{\partial V}{\partial t} + B \frac{\partial V}{\partial x} = T' \quad [10.12]$$

with:

$$B = \begin{bmatrix} \lambda & 0 \\ \lambda_U s & \lambda \end{bmatrix}, \quad T' = T - \left(\frac{\partial H}{\partial x} \right)_{V=\text{Const}} \quad [10.13]$$

where λ_U is the derivative of λ with respect to U . The matrix B has a double eigenvalue, $\lambda^{(1)} = \lambda^{(2)} = \lambda$. System [10.12] is not strictly hyperbolic because a necessary condition for hyperbolicity is that all the eigenvalues of the Jacobian matrix be distinct. System [10.10–11] is said to be linearly degenerate.

The characteristic form of sensitivity equations is derived directly from equations [1.17] and [10.12]:

$$\left. \begin{aligned} \frac{\partial U}{\partial t} + \lambda \frac{\partial U}{\partial x} &= S - \left(\frac{\partial F}{\partial x} \right)_{U=\text{Const}} \\ \frac{\partial s}{\partial t} + \lambda \frac{\partial s}{\partial x} &= \frac{\partial S}{\partial U} s + \varepsilon \frac{\partial S}{\partial \varphi} - \frac{\partial}{\partial x} \left(\varepsilon \frac{\partial F}{\partial \varphi} \right) - \frac{\partial \lambda}{\partial x} s - \left(\frac{\partial G}{\partial x} \right)_{s=\text{Const}} \end{aligned} \right\} \quad [10.14]$$

This system can be written in vector form as:

$$\frac{dV}{dt} = T'' \text{ for } \frac{dx}{dt} = \lambda \quad [10.15]$$

Both the conserved variable U and its sensitivity s are Riemann invariants for system [10.10]. In the absence of source term Q'' , U and s are constant along the characteristics.

10.2.3. Extension to discontinuous solutions

Chapter 3 deals with the properties of discontinuous solutions U for the original hyperbolic conservation law [1.1] (see section 3.4). Discontinuous solutions verify the jump relationship [3.28], recalled here:

$$(U_1 - U_2)c_s = F_1 - F_2$$

where subscripts 1 and 2 denote the values of U on the left- and right-hand sides of the discontinuity, and c_s is the speed of the discontinuity. This relationship cannot be transposed as such to the sensitivity. As shown in [BAR 02], extra terms appear in the jump relationship for the sensitivity equations. Several derivation methods are available for the sensitivity jump relationships:

- A first approach consists of carrying out two balances over an infinitesimal control volume $[x_1, x_2]$ containing the discontinuity. The first is written for the “nominal” value φ_0 of the parameter, the second is carried out for the perturbed value $\varphi_0 + \varepsilon\varphi'$. This approach is presented in [GUI 09c] and will not be detailed here.

- Another possible approach consists of representing the propagation of the discontinuity in the (x, φ) plane for a given time t and linking both sides of the discontinuity using two different paths (Figure 10.1). This approach is used hereafter.

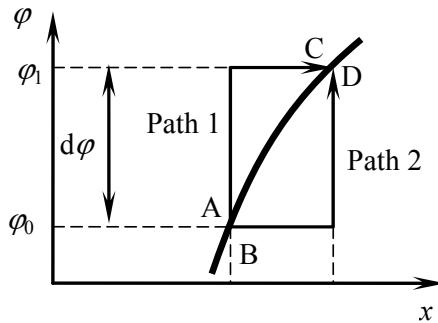


Figure 10.1. Representation of a discontinuity in the (x, φ) plane for a given time t

The jump relationships are written for two different values of the parameter: the nominal value φ_0 and a slightly different value $\varphi_1 = \varphi_0 + d\varphi$. The points on the left- and right-hand sides of the discontinuity for $\varphi = \varphi_0$ are denoted by A and B respectively. The points on the left- and right-hand sides of the discontinuity for $\varphi = \varphi_1$ are denoted by C and D respectively. The jump relationships hold:

$$\begin{aligned} F_A - F_B &= (U_A - U_B) c_s(\varphi_0) \\ F_C - F_D &= (U_C - U_D) c_s(\varphi_1) \end{aligned} \quad [10.16]$$

Connecting the points A and C along Path 1 and points B and D along Path 2 yields:

$$\left. \begin{aligned} F_C - F_A &= (x_C - x_A) \frac{\partial F}{\partial x} + (\varphi_1 - \varphi_0) \frac{\partial F}{\partial \varphi} = \left[\frac{\partial x_s}{\partial \varphi} \left(\frac{\partial F}{\partial x} \right)_L + G_L \right] d\varphi \\ F_D - F_B &= \left[\frac{\partial x_s}{\partial \varphi} \left(\frac{\partial F}{\partial x} \right)_R + G_R \right] d\varphi \\ U_C - U_A &= \left[\frac{\partial x_s}{\partial \varphi} \left(\frac{\partial U}{\partial x} \right)_L + \left(\frac{\partial U}{\partial \varphi} \right)_L \right] d\varphi = \left[\frac{\partial x_s}{\partial \varphi} \left(\frac{\partial U}{\partial x} \right)_L + s_L \right] d\varphi \\ U_D - U_B &= \left[\frac{\partial x_s}{\partial \varphi} \left(\frac{\partial U}{\partial x} \right)_R + \left(\frac{\partial U}{\partial \varphi} \right)_R \right] d\varphi = \left[\frac{\partial x_s}{\partial \varphi} \left(\frac{\partial U}{\partial x} \right)_R + s_R \right] d\varphi \end{aligned} \right\} [10.17]$$

where x_s is the abscissa of the shock. Subtracting equations [10.16] from one another leads to:

$$\begin{aligned} F_C - F_A - (F_D - F_B) &= (U_C - U_D) c_s (\varphi_1) - (U_A - U_B) c_s (\varphi_0) \\ &= [(U_C - U_A) - (U_D - U_B)] c_s (\varphi_0) \\ &\quad + (U_C - U_D) \frac{\partial c_s}{\partial \varphi} d\varphi \end{aligned} \quad [10.18]$$

Substituting relationships [10.17] into [10.18] and simplifying by $d\varphi$ leads to:

$$\left. \begin{aligned} (s_L - s_R) c_s &= G_L - G_R + R \\ R &= \left[\left(\frac{\partial F}{\partial x} - c_s \frac{\partial U}{\partial x} \right)_L - \left(\frac{\partial F}{\partial x} - c_s \frac{\partial U}{\partial x} \right)_R \right] \frac{\partial x_s}{\partial \varphi} - (U_L - U_R) \frac{\partial c_s}{\partial \varphi} \end{aligned} \right\} \quad [10.19]$$

where subscripts L and R denote the values on the left- and right-hand sides of the discontinuity. Compared to the original Rankin-Hugoniot condition [3.28], the first equation [10.19] contains an extra source term R that takes effect only at discontinuities. The expression of the source term Q in equation [10.11] is modified into:

$$T = \left[\frac{\partial S}{\partial U} s + \varepsilon \frac{\partial S}{\partial \varphi} - \frac{\partial}{\partial x} \left(\varepsilon \frac{\partial F}{\partial \varphi} \right) + R \delta_{x_s} \right] \quad [10.20]$$

where δ_{x_s} is Dirac's function centered at $x = x_s$ and R is the extra source term as defined in [10.19].

10.2.4. Solution of the Riemann problem

10.2.4.1. Definition

The Riemann problem is defined for the hyperbolic part (that is, without source term) of equation [10.10]:

$$\left. \begin{aligned} \frac{\partial V}{\partial t} + \frac{\partial H}{\partial x} &= 0 \\ V(x, 0) &= \begin{cases} V_L & \text{for } x < x_0 \\ V_R & \text{for } x > x_0 \end{cases} \end{aligned} \right\} \quad [10.21]$$

The properties of the solution of the Riemann problem are detailed in Chapter 4. The properties of the solutions of the Riemann problem for the sensitivity equations of scalar hyperbolic conservation laws are given in [GUI 07]. The reader is referred to the original publication for the details of the proofs, only the broad lines being recalled hereafter.

Recall from Chapter 4 that the solution of the Riemann problem for a convex or concave scalar hyperbolic law is made of a simple wave connecting the left and right states of the Riemann problem. If the flux function is non-convex, the wave may be a compound wave, that is, the combination of a rarefaction wave and a shock. The solution of the Riemann problem is self-similar regardless of the nature of the wave. In other words, the solution U is a function of the ratio $(x - x_0)/t$ alone.

10.2.4.2. Solution for a convex or concave law

The following configurations are considered for a convex or concave law.

(1) Rarefaction wave (Figure 10.2a). This is the case if $\lambda_L = \lambda(U_L) < \lambda_R = \lambda(U_R)$. Then the solution U is given by:

$$U(x, t) = \begin{cases} U_L & \text{for } x \leq x_0 + \lambda_L t \\ \lambda^{-1}[(x - x_0)/t] & \text{for } x_0 + \lambda_L t \leq x \leq x_0 + \lambda_R t \\ U_R & \text{for } x \geq x_0 + \lambda_R t \end{cases} \quad [10.22]$$

where the function $\lambda^{-1}(\xi)$ is the inverse function of $\lambda(U)$. The rarefaction wave can be shown to be a void sensitivity region [GUI 07], that is, the sensitivity is zero within a rarefaction wave:

$$s(x, t) = \begin{cases} s_L & \text{for } x < x_0 + \lambda_L t \\ 0 & \text{for } x_0 + \lambda_L t < x < x_0 + \lambda_R t \\ s_R & \text{for } x > x_0 + \lambda_R t \end{cases} \quad [10.23]$$

(2) Contact discontinuity (Figure 10.2b). This type of wave verifies $\lambda_L = \lambda_R = \lambda_{LR}$. In such a case, the solution U of the Riemann problem is:

$$U(x, t) = \begin{cases} U_L & \text{for } x < x_0 + \lambda_{LR} t \\ U_R & \text{for } x > x_0 + \lambda_{LR} t \end{cases} \quad [10.24]$$

The sensitivity s is given by:

$$s(x, t) = \begin{cases} s_L & \text{for } x < x_0 + \lambda_{LR} t \\ s_R & \text{for } x > x_0 + \lambda_{LR} t \end{cases} \quad [10.25]$$

Note that $\lambda_L = \lambda_R$ by definition of the contact discontinuity. This is not necessarily an indication that $R = 0$ in equation [10.19], because $\partial\lambda/\partial\varphi$ (to mention but one of the terms in [10.19]) may be non-zero.

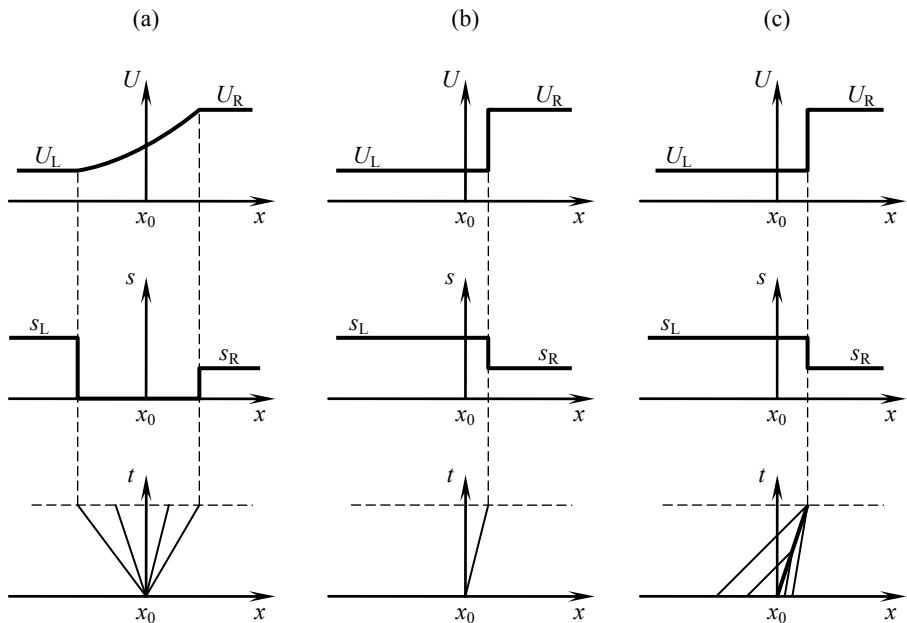


Figure 10.2. Solution of the Riemann problem for a convex/concave law.
(a) rarefaction wave; (b) contact discontinuity; (c) shock wave

(3) Shock wave (Figure 10.2c). For a convex law, this is the case if $\lambda_L > \lambda_R$. U is given by:

$$U(x, t) = \begin{cases} U_L & \text{for } x < x_0 + c_s t \\ U_R & \text{for } x > x_0 + c_s t \end{cases} \quad [10.26]$$

The sensitivity is discontinuous across the shock. Since s is a Riemann invariant, it verifies:

$$s(x, t) = \begin{cases} s_L & \text{for } x < x_0 + c_s t \\ s_R & \text{for } x > x_0 + c_s t \end{cases} \quad [10.27]$$

10.2.4.3. Non-convex conservation laws

In addition to the three configurations presented in section 10.2.4.2 non-convex flux functions give rise to a fourth possible configuration.

(4) Compound (mixed) wave: a rarefaction wave bounded by a shock. This configuration appears when the left and right state of the Riemann problem located on both sides of the value U_{\max} for which the wave speed λ is maximum (see Chapter 4):

$$\left. \begin{array}{l} U_L < U_{\max} < U_R \\ U_R < U_{\max} < U_L \\ \frac{d\lambda}{dU}(U_{\max}) = 0 \end{array} \right\} [10.28]$$

Then U is given by:

$$U(x, t) = \begin{cases} U_L & \text{for } x \leq x_0 + \lambda_L t \\ \lambda^{-1}[(x - x_0)/t] & \text{for } x_0 + \lambda_L t \leq x < x_0 + \lambda_R t \\ U_R & \text{for } x > x_0 + \lambda_R t \end{cases} [10.29]$$

for a mixed wave with a shock facing to the right (Figure 10.3a), and by:

$$U(x, t) = \begin{cases} U_L & \text{for } x < x_0 + \lambda_L t \\ \lambda^{-1}[(x - x_0)/t] & \text{for } x_0 + \lambda_L t < x \leq x_0 + \lambda_R t \\ U_R & \text{for } x \geq x_0 + \lambda_R t \end{cases} [10.30]$$

for a mixed wave with a shock facing to the left (Figure 10.3b).

In both cases the sensitivity is zero within the rarefaction wave. It verifies equation [10.22].

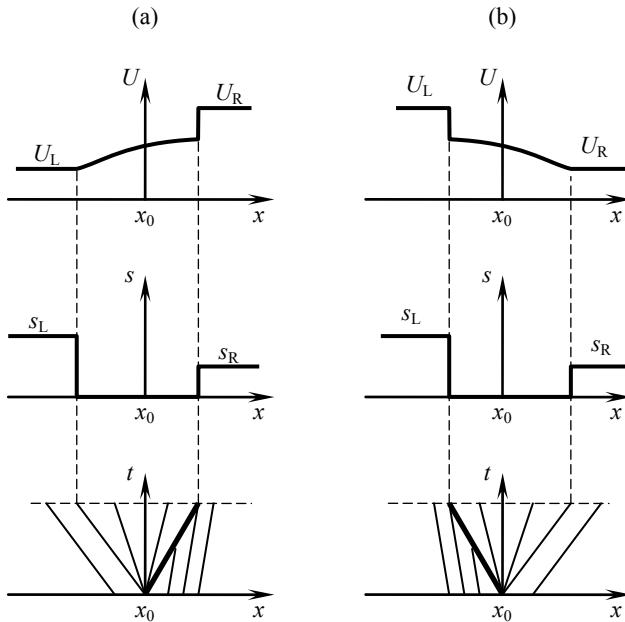


Figure 10.3. Solution of the Riemann problem for a non-convex law: compound (mixed) wave: (a) shock on the right-hand side of the wave; (b) shock on the left-hand side of the wave

10.3. Forward sensitivity equations for hyperbolic systems

10.3.1. Governing equations

The governing sensitivity equations for hyperbolic systems are derived as explained in section 10.2. Consider the conservation form [2.2] of a hyperbolic system:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} = \mathbf{S}$$

where \mathbf{F} and \mathbf{S} are functions of the conserved variable \mathbf{U} and the parameter φ .

$$\begin{cases} \mathbf{F} = \mathbf{F}(\mathbf{U}, \varphi) \\ \mathbf{S} = \mathbf{S}(\mathbf{U}, \varphi) \end{cases} \quad [10.31]$$

As in section 10.2, the parameter φ is subjected to an infinitesimal perturbation (equation [10.2]). Since both the “nominal” solution U and the perturbed solution $U + U'$ verify equation [2.2], we have:

$$\left. \begin{aligned} \frac{\partial U}{\partial t} + \frac{\partial}{\partial x} F(U, \varphi) &= S(U, \varphi) \\ \frac{\partial}{\partial t}(U + U') + \frac{\partial}{\partial x} F(U + U', \varphi + \varepsilon\varphi') &= S(U + U', \varphi + \varepsilon\varphi') \end{aligned} \right\} [10.32]$$

Subtracting equations [10.32] from each other, dividing by φ' and using the definition of the sensitivity vector $s = \partial F / \partial \varphi = U' / \varphi'$ leads to the following sensitivity equations (see [GUI 09c] for the details of the derivation):

$$\left. \begin{aligned} \frac{\partial s}{\partial t} + \frac{\partial G}{\partial x} &= Q \\ G &= As \\ Q &= \frac{\partial S}{\partial U} s + \varepsilon \frac{\partial S}{\partial \varphi} - \frac{\partial}{\partial x} \left(\varepsilon \frac{\partial F}{\partial \varphi} \right) + R \delta_s \\ R &= \left[\left(\frac{\partial F}{\partial x} - c_s \frac{\partial U}{\partial x} \right)_L - \left(\frac{\partial F}{\partial x} - c_s \frac{\partial U}{\partial x} \right)_R \right] \frac{\partial x_s}{\partial \varphi} - (U_L - U_R) \frac{\partial c_s}{\partial \varphi} \end{aligned} \right\} [10.33]$$

where $A = \partial F / \partial U$ and δ_s is Dirac's function that takes effect only at shocks. The subscripts L and R in the definition of the point source term R indicate the values of the variables on the left- and right-hand side of the shock respectively. Equation [10.33] is valid for both continuous and discontinuous solutions. Note that the jump relationships for the sensitivity are obtained as the vector version of equations [10.19]:

$$\left. \begin{aligned} (s_L - s_R) c_s &= G_L - G_R + R \\ R &= \left[\left(\frac{\partial F}{\partial x} - c_s \frac{\partial U}{\partial x} \right)_L - \left(\frac{\partial F}{\partial x} - c_s \frac{\partial U}{\partial x} \right)_R \right] \frac{\partial x_s}{\partial \varphi} - (U_L - U_R) \frac{\partial c_s}{\partial \varphi} \end{aligned} \right\} [10.34]$$

Equations [2.2] and [10.33] are rewritten in the form of a single system defining a variable V as the union of U and s :

$$\frac{\partial V}{\partial t} + \frac{\partial H}{\partial x} = T [10.35]$$

where V , H and T are defined as:

$$\begin{aligned} V &= [U_1, \dots, U_m, s_1, \dots, U_m]^T \\ H &= [F_1, \dots, F_m, G_1, \dots, G_m]^T \\ T &= [S_1, \dots, S_m, Q_1, \dots, Q_m]^T \end{aligned} \quad [10.36]$$

10.3.2. Non-conservation and characteristic forms

The non-conservation form [10.12] is recalled:

$$\frac{\partial V}{\partial t} + B \frac{\partial V}{\partial x} = T'$$

where T' is defined as in equation [10.13] by the union of two vectors:

$$T' = T - \left(\frac{\partial H}{\partial x} \right)_{V=\text{Const}} = \begin{bmatrix} S' \\ Q' \end{bmatrix} = \begin{bmatrix} S - (\partial F / \partial x)_{U=\text{Const}} \\ Q - (\partial G / \partial x)_{\substack{U=\text{Const} \\ s=\text{Const}}} \end{bmatrix} \quad [10.37]$$

and B is the square matrix defined as the union of four square matrices:

$$B = \begin{bmatrix} A & 0 \\ C & A \end{bmatrix} \quad [10.38]$$

where A and C are the square Jacobian matrices:

$$\left. \begin{array}{l} A = \frac{\partial F}{\partial U} = \frac{\partial G}{\partial s} \\ C = \frac{\partial G}{\partial U} \end{array} \right\} \quad [10.39]$$

The first row in matrix B is related to the non-conservation form for the variable U , the second row expresses the non-conservation form for the sensitivity s . Note that $\partial G / \partial s = A$ because $G = As$. A particular consequence is that all the eigenvalues of system [10.35] with B defined as in equation [10.38] are double because an eigenvalue for the conserved variable U is also an eigenvalue for the sensitivity s . The system is said to be linearly degenerate.

The characteristic form for U is derived in Chapter 2 and will not be recalled here. The characteristic form for s is obtained as follows. The vector equation [10.12] is written for the sensitivity only:

$$\frac{\partial s}{\partial t} + A \frac{\partial s}{\partial x} = -C \frac{\partial U}{\partial x} + Q' \quad [10.40]$$

and the system is diagonalized by left-multiplying equation [10.40] with the inverse of the matrix K of eigenvectors of A :

$$K^{-1} \frac{\partial s}{\partial t} + K^{-1} AK K^{-1} \frac{\partial s}{\partial x} = K^{-1} \left(Q' - C \frac{\partial U}{\partial x} \right) \quad [10.41]$$

Introducing the diagonal matrix Λ formed by the eigenvalues of A , equation [10.41] is rewritten as:

$$K^{-1} \frac{\partial s}{\partial t} + \Lambda K^{-1} \frac{\partial s}{\partial x} = K^{-1} \left(Q' - C \frac{\partial U}{\partial x} \right) \quad [10.42]$$

The vectors Y and Q'' are introduced as:

$$\begin{aligned} dY &= K^{-1} ds \\ Q'' &= K^{-1} \left(Q' - C \frac{\partial U}{\partial x} \right) \end{aligned} \quad [10.43]$$

Substituting definitions [10.43] into equation [10.42] leads to:

$$\frac{\partial Y}{\partial t} + \Lambda \frac{\partial Y}{\partial x} = Q'' \quad [10.44]$$

which is equivalent to:

$$\frac{dY_p}{dt} = Q''_p \quad \text{for } \frac{dx}{dt} = \lambda^{(p)} \quad [10.45]$$

The vector Y is the vector of sensitivity Riemann invariants. Note that Y can also be defined as:

$$Y = \frac{\partial W}{\partial \varphi} \quad [10.46]$$

10.3.3. The Riemann problem

10.3.3.1. Structure of the solution

Consider the Riemann problem [10.21] defined for the hyperbolic part the governing equations. Recall that the solution of the Riemann problem is self-similar and that it verifies the following property (see Chapter 4):

$$\left(\mathbf{B} - \frac{x - x_0}{t} \mathbf{I} \right) \frac{\partial \mathbf{V}}{\partial x} = 0 \quad [10.47]$$

As shown in Chapter 4, equation [10.47] leads to Property (P4.3): the solution is made of m simple waves (rarefaction waves, shocks or contact discontinuities) separating regions of constant state. These waves originate from the location x_0 of the initial discontinuity.

The solution of the Riemann problem is determined uniquely from the left and right states provided that m independent relationships in s can be written across each wave. The next two sections focus on the derivation of such relationships for rarefaction waves and discontinuities.

10.3.3.2. Rarefaction waves

Assume that the p th wave is a rarefaction wave. Within this wave, $(x - x_0)/t$ is an eigenvalue for \mathbf{B} , as indicated by equation [10.47]. Consequently, it is also an eigenvalue for \mathbf{A} ($p = 1, \dots, m$):

$$\lambda^{(p)} = \frac{x - x_0}{t} \quad [10.48]$$

Differentiating equation [10.48] with respect to the parameter φ leads to:

$$\frac{\partial \lambda^{(p)}}{\partial \varphi} = 0 \quad [10.49]$$

Moreover, the Riemann invariants Y_q ($q \neq p$) provide $m - 1$ relationships:

$$Y_q = \frac{\partial W_q}{\partial \varphi} = \text{Const} \quad \text{for } \frac{dx}{dt} \neq \lambda^{(p)}, q \neq p \quad [10.50]$$

Consequently, m independent relationships are available across a rarefaction wave.

10.3.3.3. Discontinuities

If the p th wave is a shock or a contact discontinuity, the jump relationships [10.34] are applicable. However, both U and s being constant on each side of the wave in the solution of the Riemann problem, the space derivatives vanish in the expression of the source term R and the jump relationships are simplified into:

$$[s^{(p,-)} - s^{(p,+)}]c_s = G^{(p,-)} - G^{(p,+)} - [U^{(p,-)} - U^{(p,+)}] \frac{\partial c_s}{\partial \varphi} \quad [10.51]$$

where c_s is the speed of the discontinuity and superscripts $(p, -)$ and $(p, +)$ indicate respectively the values of the variables on the left- and right-hand side of the discontinuity. The expression for the derivative of c_s with respect to φ can be derived from the jump relationships for the variable U . Indeed, the Rankin-Hugoniot conditions yield:

$$c_s = \frac{F_q^{(p,-)} - F_q^{(p,+)}}{U_q^{(p,-)} - U_q^{(p,+)}} \quad \forall q = 1, \dots, m \quad [10.52]$$

Consequently:

$$\frac{\partial c_s}{\partial \varphi} = \frac{G_q^{(p,-)} - G_q^{(p,+)}}{U_q^{(p,-)} - U_q^{(p,+)}} - \frac{\left[\frac{F_q^{(p,-)} - F_q^{(p,+)}}{U_q^{(p,-)} - U_q^{(p,+)}} \right] [s_q^{(p,-)} - s_q^{(p,+)}]}{\left[U_q^{(p,-)} - U_q^{(p,+)} \right]^2} \quad \forall q = 1, \dots, m \quad [10.53]$$

Since $\partial c_s / \partial \varphi$ is entirely determined from the values of s and U on both sides of the discontinuity, m independent equations [10.51] can be written across a discontinuity.

10.3.4. Application example: the one-dimensional shallow water sensitivity equations

10.3.4.1. Governing equations

The purpose is to derive the analytical sensitivity solution for the dambreak problem presented in section 4.3.3. Recall that the dambreak problem is a Riemann problem under the assumption of zero bottom slope and frictionless motion.

If the channel is rectangular, the Saint Venant equations (see section 2.5) are equivalent to the one-dimensional version of the shallow water equations (see sections 5.4 and 7.4.2). U and F are defined as:

$$U = \begin{bmatrix} h \\ q \end{bmatrix}, F = \begin{bmatrix} q \\ M \end{bmatrix} = \begin{bmatrix} uh \\ q^2/h + gh^2/2 \end{bmatrix} \quad [10.54]$$

where g is the gravitational acceleration, h is the water depth, M is the specific force, q is the unit discharge and u is the flow velocity.

The complete description of the solution of the dambreak problem is given in Chapter 4. Recall that the solution is made of an intermediate region of constant state separated from the left state by a rarefaction wave and from the right state by a shock. The Riemann invariant $W_2 = u + c$ is applicable across the rarefaction wave, while the jump relationships apply across the shock.

The Jacobian matrix A is recalled:

$$A = \begin{bmatrix} 0 & 1 \\ c^2 - u^2 & 2u \end{bmatrix} \quad [10.55]$$

where $c = (gh)^{1/2}$ is the propagation speed of the waves in still water and $u = q/h$. The sensitivity vector s and the sensitivity flux G are defined as:

$$s = \begin{bmatrix} \eta \\ \theta \end{bmatrix}, \quad G = As = \begin{bmatrix} \theta \\ (c^2 - u^2)\eta + 2u\theta \end{bmatrix} \quad [10.56]$$

where η and θ are respectively the sensitivity of h and q to the parameter φ (that may be any parameter). The matrix B is given by:

$$B = \begin{bmatrix} 0 & 1 & 0 & 0 \\ c^2 - u^2 & 2u & 0 & 0 \\ 0 & 0 & 0 & 1 \\ (g - 2u^2/h)\eta - 2u\theta/h & (\theta - u\eta)/h & c^2 - u^2 & 2u \end{bmatrix} \quad [10.57]$$

It is easy to check that B has the following double eigenvalues:

$$\left. \begin{aligned} \lambda^{(1)} &= \lambda^{(2)} = u - c \\ \lambda^{(3)} &= \lambda^{(4)} = u + c \end{aligned} \right\} \quad [10.58]$$

10.3.4.2. Riemann problem definition

The purpose is to solve the Riemann problem with left and right states in U:

$$U_L = \begin{bmatrix} h_L \\ 0 \end{bmatrix}, \quad U_R = \begin{bmatrix} h_R \\ 0 \end{bmatrix} \quad [10.59]$$

with $h_L > h_R$.

The left and right states for the sensitivity depend on the parameter φ examined in the sensitivity analysis. In what follows, it is chosen to examine the sensitivity of the solution of the dambreak problem to the initial water level h_L on the left-hand side of the dam, while h_R is assumed fixed. Then $\varphi = h_L$ and we have:

$$s_L = \begin{bmatrix} \eta_L \\ \theta_L \end{bmatrix} = \begin{bmatrix} \partial h_L / \partial h_L \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad s_R = \begin{bmatrix} \eta_R \\ \theta_R \end{bmatrix} = \begin{bmatrix} \partial h_R / \partial h_L \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad [10.60]$$

Also note that if the sensitivity analysis focuses on the influence of initial or boundary conditions, the value of the parameter (the initial or boundary conditions) does not influence the expression of the flux F and source term S at times $t > 0$ within the domain. Therefore, $\partial F / \partial \varphi = \partial S / \partial \varphi = 0$ in the expression of the source term Q.

10.3.4.3. Sensitivity solution

In the rarefaction wave $\lambda^{(1)} = u - c$, equation [10.49] is applicable:

$$\nu - \chi = 0 \quad [10.61]$$

where ν and χ are respectively the sensitivity of u and c to the left state h_L . The speed c of the waves in still water is defined as $c = (gh)^{1/2}$.

Moreover, the Riemann invariant $W_2 = u + 2c$ may be used across the wave:

$$u + 2c = u_L + 2c_L \quad [10.62]$$

Differentiating equation [10.62] with respect to φ yields:

$$\nu + 2\chi = \nu_L + 2\chi_L \quad [10.63]$$

System [10.61–62] can be solved uniquely for v and χ in the rarefaction wave:

$$v = \chi = \frac{v_L + 2\chi_L}{3} = \frac{2}{3}\chi_L \text{ for } -c_L < \frac{x - x_0}{t} < u^* - c^* \quad [10.64]$$

where the $*$ superscript indicates the variables in the intermediate region of constant state. The sensitivities η and θ are derived from equation [10.64] by noting that $h = c^2/h$ and $q = hu$. Consequently:

$$\left. \begin{aligned} \eta &= \frac{2c}{g} \chi \\ \theta &= \eta u + h v \end{aligned} \right\} \quad [10.65]$$

Applying the first equation [10.65] yields the expression for the left state c_L in equation [10.64]:

$$\chi_L = \frac{g}{2c_L} \eta_L \quad [10.66]$$

Moreover, the profile for c is given by equation [4.48], recalled here:

$$c(x, t) = \frac{1}{3} \left(2c_L - \frac{x - x_0}{t} \right)$$

Substituting equations [10.64] and [4.48] into equation [10.65] leads to:

$$\eta(x, t) = \frac{2}{9c_L} \left(2c_L - \frac{x - x_0}{t} \right) \eta_L \quad \text{for } -c_L < \frac{x - x_0}{t} < u^* - c^* \quad [10.67]$$

A similar reasoning yields the following expression for the sensitivity of the unit discharge:

$$\theta(x, t) = \frac{2\eta_L}{9c_L} \left(c_L - \frac{x - x_0}{t} \right) \left(2c_L + \frac{x - x_0}{t} \right) \text{ for } -c_L < \frac{x - x_0}{t} < u^* - c^* \quad [10.68]$$

It is visible from equation [10.67] that the sensitivity of the water depth varies linearly with x in the rarefaction wave. In addition, using $(x - x_0)/t = -c_L$ in equation [10.67] yields a limit value $\eta = 2\eta_L/3$. Consequently, the sensitivity h is discontinuous at the left-hand boundary of the rarefaction wave. Equation [10.67] gives $\eta = 0$ for $(x - x_0)/t = 2c_L$ (Figure 10.4).

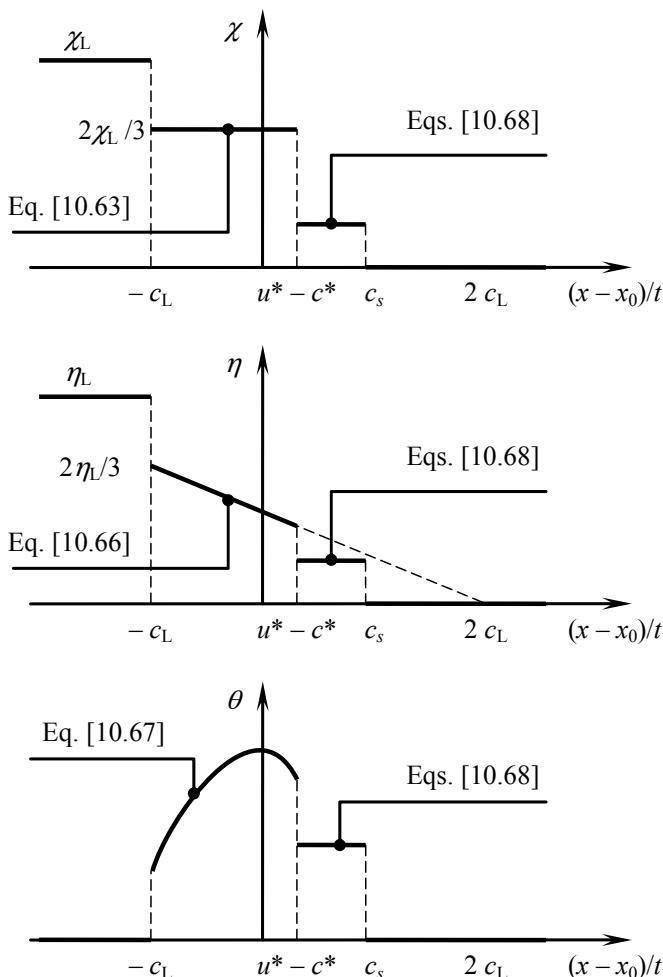


Figure 10.4. Dambreak problem. Definition sketch for the sensitivity solution

The sensitivities η^* and θ^* in the intermediate region of constant state are obtained by applying the jump relationships [10.51] and [10.53] and noting that $u_R = \eta_R = \theta_R = 0$:

$$\left. \begin{aligned} \eta^* c_s &= \theta^* - \eta^* \frac{\partial c_s}{\partial h_L} \\ \theta^* c_s &= (c^2 - u^2)^* \eta^* + 2u^* \theta^* - \theta^* \frac{\partial c_s}{\partial h_L} \end{aligned} \right\} \quad [10.69]$$

The shock speed c_s is known from the solution of the Riemann problem for the flow equations:

$$c_s = \frac{q^* - q_R}{h^* - h_R} = \frac{q^*}{h^* - h_R} \quad [10.70]$$

Differentiating with respect to $\varphi = h_L$ leads to:

$$\frac{\partial c_s}{\partial h_L} = \frac{\theta^*}{h^* - h_R} - \frac{q^* \eta^*}{(h^* - h_R)^2} \quad [10.71]$$

System [10.69–71] is nonlinear. It can be solved using iterative techniques such as Newton-Raphson's method. Note that in the general case, equations [10.69] and [10.67–68] do not yield a continuous sensitivity profile at the right-hand boundary of the rarefaction wave (Figure 10.4).

Figure 10.5 illustrates the solution obtained for the parameters in Table 10.1.

Symbol	Meaning	Value
g	Gravitational acceleration	9.81 m s^{-2}
h_L	Initial upstream water depth	10 m
h_R	Initial downstream water depth	1 m
η_L	Sensitivity to the initial water depth on the upstream side	1
η_R	Sensitivity to the initial water depth on the downstream side	0

Table 10.1. Dambreak problem parameters

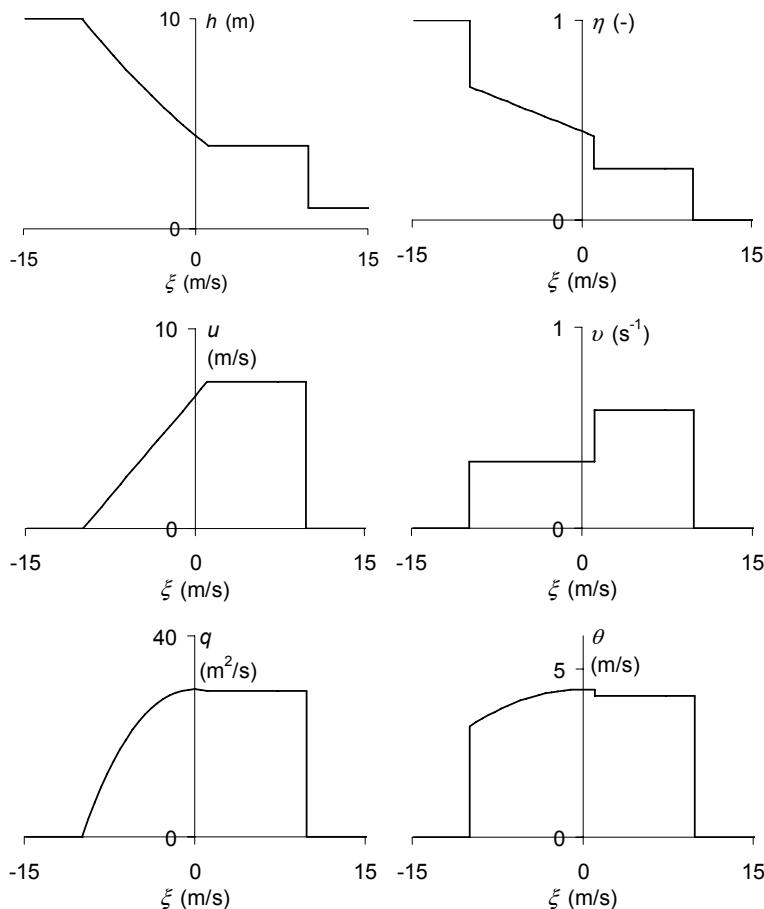


Figure 10.5. Dambreak problem. Analytical solution for the parameter set in Table 10.1

Figure 10.6 shows the empirical sensitivity profiles obtained from the numerical solution of the dam-break problem using a finite volume technique. The first-order Godunov scheme described in Chapter 7 is used with the HLL approximate Riemann solver described in Appendix C. The empirical sensitivity solution is computed by solving the shallow water equations twice. In the first simulation, the initial water depth on the left-hand side of the dam is set to $h_L + \varepsilon/2$. In the second simulation, it is set to $h_L - \varepsilon/2$. The sensitivities η , χ and θ are obtained by dividing the difference in h , c and q between the two simulations by ε .

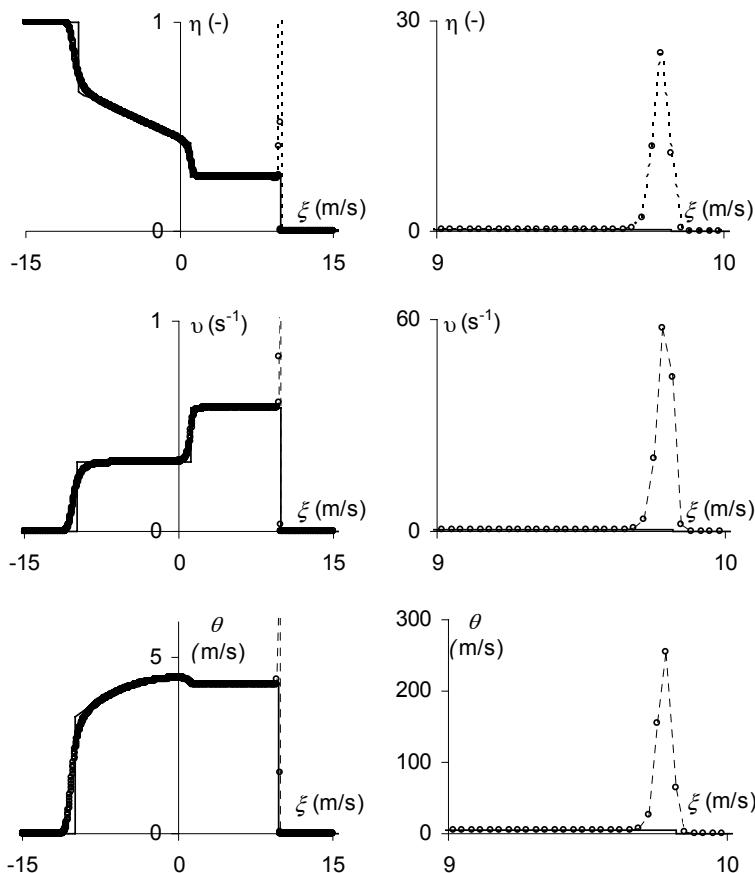


Figure 10.6. Dambreak problem. Analytical solution (solid line) and empirical solution (dashed, dotted line) for the sensitivity

The profiles are plotted as functions of the ratio $\xi = (x - x_0)/t$. The empirical sensitivity profiles clearly exhibit artificial peaks in the neighborhood of the shock. Rescaling the graph on the right-hand side of Figure 10.6 indicates that the amplitude of the artificial peak is up to 60 times that of the theoretical value of the sensitivity next to the shock. This example illustrates the need for numerical methods that do not exhibit such undesirable behaviors. Examples of such techniques are presented in section 10.5.

10.4. Adjoint sensitivity equations

10.4.1. Introduction

The forward sensitivity equations presented in the previous sections are most useful when the purpose of the sensitivity analysis is to study the influence of a single parameter on many model outputs (or variables). Examples of such situations are the sensitivity analysis of the dambreak problem presented in section 10.3.4. The influence of the upstream water level (a single simulation parameter) is studied for all the flow variables at all points and at all times.

In a number of inverse problems, however, the objective of the sensitivity analysis is to investigate the influence of numerous model parameters on a reduced set of flow variables. This is the case for instance when time-varying boundary conditions or spatially-varying initial conditions in a flow model are to be adjusted so as to reproduce flow measurements at a limited number of locations as accurately as possible. In this case, the number of points and time steps at which the initial and boundary conditions are to be adjusted may be much larger than the number of measurement points. The forward sensitivity approach is extremely time-consuming because it requires that one forward sensitivity calculation be carried out for each point (and each time) where the initial and/or boundary conditions are to be adjusted. In such situations, that are typical of inverse problems, the adjoint sensitivity analysis approach is more appropriate [CAC 03].

Adjoint sensitivity analysis is used in many fields of engineering such as model inversion (data assimilation [LED 86], model calibration, [PAN 89]), optimization problems, flow control and uncertainty analysis. Its earliest applications can be found in the field of meteorology and atmospheric sciences [HAL 82, HAL 83]. The reader interested in a formal, general definition of adjoint operators and an introduction to the underlying theory may refer to [CAC 03]. The theory of adjoint models can be found in [MAR 95]. The purpose of this section is to introduce the broad principles of adjoint sensitivity equations for one-dimensional hyperbolic conservation laws. The same derivation principle may be applied to systems of conservation laws. The reader interested in application examples of the adjoint sensitivity analysis technique to systems of conservation laws may refer to [SAN 00, SAN 99], which describe applications to free surface flow.

10.4.2. Adjoint models for scalar laws

10.4.2.1. Derivation

Consider a one-dimensional flow model obeying a scalar hyperbolic conservation law in the form [1.1], with a forward sensitivity equation in the

form [10.9]. For the sake of conciseness, the support function ε of the perturbation is considered to be uniformly zero and the solution is assumed to be continuous, therefore $R = 0$. It is also assumed that the wave speed λ is positive. The purpose is to optimize the value of the parameter φ so as to minimize a so-called objective function $J(\varphi)$ defined in general form as:

$$J(\varphi) = \int_0^T \int_0^L f(U, \varphi) dx dt \quad [10.72]$$

where L and T are respectively the length of the domain and the time interval over which the objective function is defined. The parameter φ may be an initial or boundary condition, or a model parameter such as a friction coefficient, bottom slope (for an open channel model), sound speed or pipe diameter (for a water hammer model), etc. The function f depends on the nature of the optimization problem to be solved. Assume for instance that a measurement device provides an experimental value for U for a given time τ at a given abscissa X . No other measurement is available. If the purpose is to minimize the difference between the computed and measured value at $[X, \tau]$, the function f may be defined as a least square-based distance:

$$f(U, \varphi) = [U(X, \tau) - U_{\text{meas}}]^2 = \int_0^T \int_0^L [U(x, t) - U_{\text{meas}}]^2 \delta_{(X, \tau)} dx dt \quad [10.73]$$

where $\delta_{(X, \tau)}$ is the Dirac function that takes effect at the abscissa X and time τ . The objective function J is zero (that is, minimum) when the distance between the simulated and measured variable U is zero.

Classically, gradient-based methods (sometimes called quasi-Newton methods) are used to find the minimum or zero of the objective function $J(\varphi)$. The gradient of J in the parameter space is given by:

$$\frac{\partial J}{\partial \varphi} = \int_0^T \int_0^L \frac{\partial f}{\partial \varphi} + \frac{\partial f}{\partial U} \frac{\partial U}{\partial \varphi} dx dt = \int_0^T \int_0^L \frac{\partial f}{\partial \varphi} + \frac{\partial f}{\partial U} s dx dt \quad [10.74]$$

The expression of f being known, $\partial f / \partial \varphi$ and $\partial f / \partial U$ can be computed and the gradient of the objective function can be computed provided that the sensitivity s is known. However, as mentioned in the previous section, computing s for a large set of parameters φ is computationally expensive. The adjoint formulation is obtained

by modifying equation [10.74] as follows. The first equation [10.9] is rewritten as (note the simplification $\varepsilon = 0, R = 0$):

$$\frac{\partial s}{\partial t} + \frac{\partial G}{\partial x} - \frac{\partial S}{\partial U} s = 0 \quad [10.75]$$

Equation [10.75] is multiplied by an arbitrary function $\mu(x, t)$ called the Lagrange multiplier. The resulting product, which is zero, is added to the integrand in equation [10.74], thus leading us to redefine the gradient of the objective function as:

$$\frac{\partial J}{\partial \varphi} = \int_0^T \int_0^L \left[\frac{\partial f}{\partial \varphi} + \frac{\partial f}{\partial U} s + \left(\frac{\partial s}{\partial t} + \frac{\partial G}{\partial x} - \frac{\partial S}{\partial U} s \right) \mu \right] dx dt \quad [10.76]$$

The adjoint equation is obtained by deriving a governing equation for the Lagrange multiplier μ . This is done via integration by parts:

$$\left. \begin{aligned} \int_0^T \int_0^L \frac{\partial s}{\partial t} \mu dx dt &= \int_0^L [s\mu]_0^T dx - \int_0^T \int_0^L \frac{\partial \mu}{\partial t} s dx dt \\ \int_0^T \int_0^L \frac{\partial G}{\partial x} \mu dx dt &= \int_0^T [\mu G]_0^L dt - \int_0^T \int_0^L \frac{\partial \mu}{\partial x} G dx dt \end{aligned} \right\} \quad [10.77]$$

Substituting equations [10.77] into equation [10.76] and noting that $G = \lambda s$ leads to:

$$\begin{aligned} \frac{\partial J}{\partial \varphi} = &+ \int_0^L [\mu s]_0^T dx + \int_0^T [\mu G]_0^L dt + \int_0^T \int_0^L \frac{\partial f}{\partial \varphi} dx dt \\ &- \int_0^T \int_0^L \left(\frac{\partial \mu}{\partial t} + \lambda \frac{\partial \mu}{\partial x} + \frac{\partial S}{\partial U} \mu - \frac{\partial f}{\partial U} \right) s dx dt \end{aligned} \quad [10.78]$$

Assume that μ verifies the following equation:

$$\frac{\partial \mu}{\partial t} + \lambda \frac{\partial \mu}{\partial x} + \frac{\partial S}{\partial U} \mu - \frac{\partial f}{\partial U} = 0 \quad [10.79]$$

Equation [10.79] is similar to equation [10.75], with the difference that (i) an additional term $\partial f / \partial U$ is introduced and (ii) the sign of the source term $\partial S / \partial U$ is

changed. This has consequences on the stability of μ . The solution s of equation [10.75] is stable when computed for increasing times. Changing the sign of the term $\partial S / \partial U$ in equation [10.79] may result in instability if μ is also computed for increasing times. The initial sign for the source term can be recovered by introducing the reverse time and space coordinates $t' = T - t$ and $x' = L - x$. Equation [10.79] becomes:

$$\frac{\partial \mu}{\partial t'} + \lambda \frac{\partial \mu}{\partial x'} = \frac{\partial S}{\partial U} \mu + \frac{\partial f}{\partial U} \quad [10.80]$$

This is the final form of the adjoint sensitivity equation. The solution μ is stable if computed in the direction of positive t' , that is, for negative t . Assuming that λ is positive, a boundary condition is needed at $x' = 0$ (that is, at $x = L$). The initial condition is needed for $t' = 0$ (which corresponds to $t = T$). The simplest possible conditions are:

$$\left. \begin{array}{ll} \mu(x, T) = 0 & \forall x \in [0, L] \\ \mu(L, t) = 0 & \forall t \in [0, T] \end{array} \right\} \quad [10.81]$$

Assuming that the adjoint equation [10.80] with initial and boundary conditions [10.81] is satisfied, substituting equations [10.79] and [10.81] into equation [10.78] leads to the following expression for the gradient of the objective function:

$$\begin{aligned} \frac{\partial J}{\partial \varphi} &= - \int_0^L (\mu s)(x, 0) dx - \int_0^T (\mu G)(0, t) dt + \int_0^T \int_0^L \frac{\partial f}{\partial \varphi} dx dt \\ &= - \int_0^L (s\mu)(x, 0) dx - \int_0^T (\lambda s\mu)(0, t) dt + \int_0^T \int_0^L \frac{\partial f}{\partial \varphi} dx dt \end{aligned} \quad [10.82]$$

10.4.2.2. Physical interpretation – algorithmic aspects

The adjoint equation [10.80] may be interpreted as follows: the origin of a perturbation in the variable U at a given location (x, t) in the solution domain is to be sought at earlier times, at the points the perturbation is likely to come from. Solving equation [10.80] is equivalent to traveling backward along the characteristic lines in the (x, t) plane (Figure 10.7). The smaller the value of μ at $t = 0$ and/or $x = 0$, the smaller the value of the integrals in equation [10.82], thus the smaller the magnitude of $\partial J / \partial \varphi$. The Lagrange multiplier may be seen as an indicator of the influence of the flow solution at a given time and abscissa on the final value of the objective function.

From an algorithmic point of view, the adjoint formalism associated with a quasi-Newton procedure implies the following steps:

(1) Forward step: solve the flow and forward sensitivity equation in the direction of positive time over the interval $[0, T]$ for a given value of the parameter φ . This provides the value of U and s . The objective function J is computed from equation [10.72]. In the general case, J is not minimum for the selected value of φ .

(2) Backward step: solve the adjoint sensitivity equation [10.80] with initial and boundary conditions [10.81] in the direction of negative times. This yields the Lagrange multiplier μ at all points (x, t) of the solution domain $[0, L] \times [0, T]$.

(3) Compute the gradient $\partial J / \partial \varphi$ using equation [10.82]. Update the parameter φ using a classical Newton procedure:

$$\varphi \mapsto \varphi - \frac{J(\varphi)}{\partial J / \partial \varphi} \quad [10.83]$$

Steps (1) – (3) are repeated sequentially.

In practical computer implementations, the sequence (1) – (2) involves that the results of the forward flow and sensitivity calculation being stored and available for the backward solution of the adjoint problem. In contrast with the classical forward computation procedure, the previously computed time levels cannot be erased from the memory of the computer because the flow variables must be available over the entire time interval $[0, T]$. This may imply considerable memory and storage requirements.

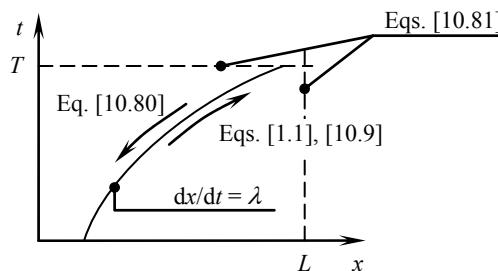


Figure 10.7. Forward and adjoint sensitivity equations. Definition sketch in the (x, t) plane

10.4.2.3. Extension to hyperbolic systems

The adjoint sensitivity system is derived as follows. Each component of the vector sensitivity equation is multiplied by a Lagrange multiplier μ_p :

$$\left(\frac{\partial s_p}{\partial t} + \frac{\partial G_p}{\partial x} - Q_p \right) \mu_p = 0, \quad p = 1, \dots, m \quad [10.84]$$

where subscript p denotes the component of the vector equation. Adding this system to the objective function as in section 10.4.2.1 leads to:

$$\frac{\partial s_\Delta}{\partial t} M + \frac{\partial G_\Delta}{\partial x} M - Q_\Delta M - \frac{\partial f}{\partial U} M = 0 \quad [10.85]$$

where M is the vector of Lagrange multipliers and s_Δ , G_Δ and Q_Δ are diagonal matrices constructed from the components of the vectors s , G and Q :

$$s_\Delta = \begin{bmatrix} \ddots & & 0 \\ & s_p & \\ 0 & & \ddots \end{bmatrix}, \quad G_\Delta = \begin{bmatrix} \ddots & & 0 \\ & G_p & \\ 0 & & \ddots \end{bmatrix}, \quad Q_\Delta = \begin{bmatrix} \ddots & & 0 \\ & Q_p & \\ 0 & & \ddots \end{bmatrix} \quad [10.86]$$

As in section 10.4.2.1, the differential operators are swapped via integration by parts, leading to the following equation:

$$s_\Delta \frac{\partial M}{\partial t'} + G_\Delta \frac{\partial M}{\partial x'} = Q_\Delta M + \frac{\partial f}{\partial U} M \quad [10.87]$$

Multiplying equation [10.87] by the inverse of s_Δ leads to the adjoint system:

$$\frac{\partial M}{\partial t'} + s_\Delta^{-1} G_\Delta \frac{\partial M}{\partial x'} = s_\Delta^{-1} Q_\Delta M + s_\Delta^{-1} \frac{\partial f}{\partial U} M \quad [10.88]$$

Riemann invariants can be derived for the Lagrange multipliers. Left-multiplying equation [10.88] by K_Δ^{-1} leads to:

$$\frac{\partial \psi}{\partial t'} + \Lambda \frac{\partial \psi}{\partial x'} = K_\Delta^{-1} \left(s_\Delta^{-1} Q_\Delta M + s_\Delta^{-1} \frac{\partial f}{\partial U} M \right) \quad [10.89]$$

where K_Δ is the matrix of eigenvectors of $s_\Delta^{-1} G_\Delta$ and ψ is the vector of adjoint Riemann invariants:

$$d\psi = K_\Delta^{-1} dM \quad [10.90]$$

A more convenient way of deriving the characteristic form of the adjoint equation consists of starting from the characteristic form [10.44] of the sensitivity equation:

$$\frac{\partial Y_\Delta}{\partial t} M + \Lambda \frac{\partial Y_\Delta}{\partial x} M = Q''_\Delta M + \frac{\partial f}{\partial U} M \quad [10.91]$$

Applying integration by parts yields the adjoint equation:

$$Y_\Delta \frac{\partial M}{\partial t'} + \Lambda Y_\Delta \frac{\partial M}{\partial x'} = Q^{(3)} + \frac{\partial f}{\partial U} M \quad [10.92]$$

where $Q^{(3)} = Q''_\Delta M$. Multiplying by the inverse of Y_Δ and noting that $Y_\Delta^{-1} \Lambda Y_\Delta = \Lambda$ leads to:

$$\frac{\partial M}{\partial t'} + \Lambda \frac{\partial M}{\partial x'} = Y_\Delta^{-1} \left(Q^{(3)} + \frac{\partial f}{\partial U} M \right) \quad [10.93]$$

The vector equation [10.93] forms a system of m characteristic equations. The Lagrange multipliers μ_p ($p = 1, \dots, m$) are the adjoint Riemann invariants.

10.5. Finite volume solution of the forward sensitivity equations

10.5.1. Introduction

As shown in section 10.3.4, the empirical solution of sensitivity equations yields numerical artifacts such as abnormally large (if not locally infinite) values of the computed sensitivity. Other artifacts have been observed in two-dimensional free surface flow simulations, such as artificial sensitivity swirls in regions where they should not be present [GUI 09c]. Such artifacts can be eliminated to a large extent if the sensitivity equations are solved directly.

Most methods for direct sensitivity calculation presented in the literature deal with continuous solutions (see e.g. [GUN 99, LU 07]). The purpose of this section is to provide the broad lines for a finite volume-based solution technique of sensitivity equations with discontinuous flow solutions. The principle of the method is presented in [GUI 07] for scalar laws, with an application to the kinematic wave model seen in Chapter 1.

The technique is extended to the shallow water equations in [DEL 08, GUI 09c]. A more accurate version of the Riemann solver with an extension to 3×3 systems including a passive scalar transport equation is given in [GUI 09a, GUI 09b].

Practical implementation aspects, including boundary conditions, can be found in [GUI 09b].

10.5.2. Discretization

The purpose is to solve the $2m \times 2m$ hyperbolic system [10.35] formed by the governing equations for the flow variable U and the sensitivity s . For the sake of clarity, only the case $m = 2$ is considered and the source terms S and Q are assumed to be zero. Consequently, the vector source term T as defined in equation [10.36] only incorporates the sensitivity Dirac source term at discontinuities. A finite volume discretization is proposed in [DEL 08, GUI 09a, GUI 09b, GUI 09c] (Figure 10.8):

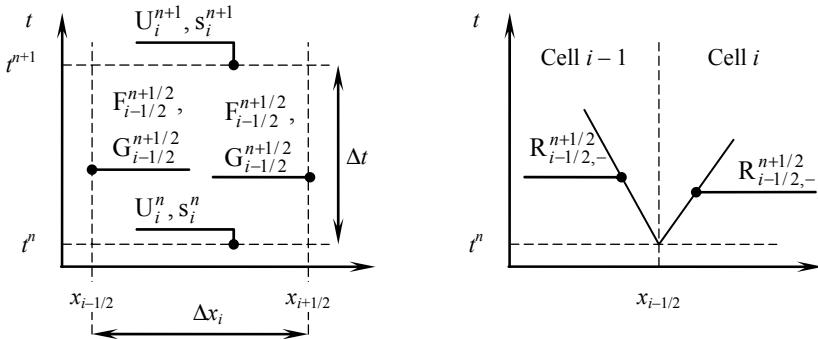


Figure 10.8. Direct solution of the sensitivity equations. Definition sketch for the finite volume solution technique. Left: principle of the finite volume discretization. Right: splitting the sensitivity source term into two parts at each interface

$$\left. \begin{aligned} U_i^{n+1} &= U_i^n + \frac{\Delta t}{\Delta x_i} (F_{i-1/2}^{n+1/2} - F_{i+1/2}^{n+1/2}) \\ s_i^{n+1} &= s_i^n + \frac{\Delta t}{\Delta x_i} (G_{i-1/2}^{n+1/2} - G_{i+1/2}^{n+1/2} + R_{i-1/2,+}^{n+1/2} + R_{i+1/2,-}^{n+1/2}) \end{aligned} \right\} \quad [10.94]$$

where subscripts i , $i - 1/2$ and $i + 1/2$ indicate respectively the average value in the cell i , the value at the interfaces $i - 1/2$ and $i + 1/2$, and superscripts n and $n + 1/2$ denote respectively the value at the time level n and the average value between the time levels n and $n + 1$. Subscript $(i - 1/2, +)$ denotes the contribution of the source term R arising from interface $i - 1/2$ to the cell located in the direction of positive x (that is, cell i). Conversely, subscript $(i + 1/2, -)$ denotes the contribution of the source term arising from interface $i + 1/2$ to the cell located on the left-hand side (that is, cell i).

F , G and R are estimated from the solution of Riemann problems at the interfaces between the computational cells. The contributions $R_{i-1/2,-}^{n+1/2}$ and $R_{i-1/2,+}^{n+1/2}$ shown in Figure 10.8 are computed from the solution of the Riemann problem. $R_{i-1/2,-}^{n+1/2}$ arises from the waves with negative propagation speeds, while $R_{i-1/2,+}^{n+1/2}$ is associated with waves that have a positive propagation speeds. The proposed approach being explicit, the estimates of F , G and R in equations [10.94] are based on the values of U , s , F and G at the known time level n .

10.5.3. A modified HLL Riemann solver for sensitivity solutions

10.5.3.1. Principle of the solver

The fluxes $F_{i-1/2}^{n+1/2}$ and $G_{i-1/2}^{n+1/2}$ at the interface $i - 1/2$ are computed by solving a Riemann problem with left and right states V_L and V_R . If the first-order Godunov scheme is used, V_L and V_R are respectively equal to the average values in the cells $i - 1$ and i . If higher-order schemes are used, V_L and V_R are inferred from the reconstructed profiles in the cells $i - 1$ and i . The states V_L and V_R are assumed known hereafter.

The Riemann solver proposed in [GUI 09a, GUI 09b] is an extension of the approximate HLL Riemann solver described in Appendix C (see section C.1). Recall that the HLL Riemann solver uses the assumption of an intermediate region of constant state (U^*, s^*) separated from the left and right states of the Riemann problem by two discontinuities. The jump relationships hold across these discontinuities:

$$\left. \begin{aligned} F_L - F^* &= (U_L - U^*) \lambda^{(1)} \\ F^* - F_R &= (U^* - U_R) \lambda^{(2)} \\ G_L - G^* &= (s_L - s^*) \lambda^{(1)} + (U_L - U^*) \frac{\partial \lambda^{(1)}}{\partial \varphi} \\ G^* - G_R &= (s^* - s_R) \lambda^{(2)} + (U^* - U_R) \frac{\partial \lambda^{(2)}}{\partial \varphi} \end{aligned} \right\} \quad [10.95]$$

where $\lambda^{(1)}$ and $\lambda^{(2)}$ are the speeds of the left- and right-hand discontinuities. These speeds are assumed known *a priori* from the left and right states. Various estimates for them are provided in Appendix C. Note that the jump relationships for the sensitivity in [10.95] are obtained as particular cases of the more general jump relationships [10.34] because the x -derivatives of the HLL-solution are zero.

10.5.3.2. Flux formulae

System [10.94] can be solved uniquely for U^* , s^* , F^* and G^* :

$$\left. \begin{aligned} U^* &= \frac{-\lambda^{(1)}U_L + \lambda^{(2)}U_R}{\lambda^{(2)} - \lambda^{(1)}} + \frac{F_L - F_R}{\lambda^{(2)} - \lambda^{(1)}} \\ F^* &= \frac{\lambda^{(2)}F_L - \lambda^{(1)}F_R}{\lambda^{(2)} - \lambda^{(1)}} - \frac{\lambda^{(1)}\lambda^{(2)}}{\lambda^{(2)} - \lambda^{(1)}}(U_L - U_R) \\ s^* &= \frac{-\lambda^{(1)}s_L + \lambda^{(2)}s_R}{\lambda^{(2)} - \lambda^{(1)}} + \frac{G_L - G_R}{\lambda^{(2)} - \lambda^{(1)}} \\ &\quad + \frac{1}{\lambda^{(2)} - \lambda^{(1)}} \frac{\partial \lambda^{(2)}}{\partial \varphi} (U^* - U_R) + \frac{1}{\lambda^{(2)} - \lambda^{(1)}} \frac{\partial \lambda^{(1)}}{\partial \varphi} (U_L - U^*) \\ G^* &= \frac{\lambda^{(2)}G_L - \lambda^{(1)}G_R}{\lambda^{(2)} - \lambda^{(1)}} - \frac{\lambda^{(1)}\lambda^{(2)}}{\lambda^{(2)} - \lambda^{(1)}}(s_L - s_R) \\ &\quad + \frac{\lambda^{(1)}}{\lambda^{(2)} - \lambda^{(1)}} \frac{\partial \lambda^{(2)}}{\partial \varphi} (U^* - U_R) + \frac{\lambda^{(2)}}{\lambda^{(2)} - \lambda^{(1)}} \frac{\partial \lambda^{(1)}}{\partial \varphi} (U_L - U^*) \end{aligned} \right\} [10.96]$$

The second equation [10.96] provides the expression for the flux F^* in the intermediate region of constant state. The flux at the interface $i - 1/2$ is equal to F_L if $\lambda^{(1)} > 0$, to F^* if $\lambda^{(1)} \leq 0 \leq \lambda^{(2)}$, and to F_R if $\lambda^{(2)} < 0$. These three formulae can be gathered into a single expression as:

$$F_{i-1/2}^{n+1/2} = \frac{\lambda^+ F_L - \lambda^- F_R}{\lambda^+ - \lambda^-} - \frac{\lambda^- \lambda^+}{\lambda^+ - \lambda^-} (U_L - U_R) \quad [10.97]$$

where λ^- and λ^+ are bounded expressions for $\lambda^{(1)}$ and $\lambda^{(2)}$:

$$\left. \begin{aligned} \lambda^- &= \min(\lambda^{(1)}, 0) \\ \lambda^+ &= \max(\lambda^{(2)}, 0) \end{aligned} \right\} [10.98]$$

A similar expression can be proposed for $G_{i-1/2}^{n+1/2}$:

$$\left. \begin{aligned} G^* &= \frac{\lambda^+ G_L - \lambda^- G_R}{\lambda^+ - \lambda^-} - \frac{\lambda^- \lambda^+}{\lambda^+ - \lambda^-} (s_L - s_R) \\ &\quad + \frac{\lambda^-}{\lambda^+ - \lambda^-} \frac{\partial \lambda^+}{\partial \varphi} (U^* - U_R) + \frac{\lambda^+}{\lambda^+ - \lambda^-} \frac{\partial \lambda^-}{\partial \varphi} (U_L - U^*) \end{aligned} \right\} [10.99]$$

In contrast with equation [10.97], equation [10.99] requires that the intermediate state U^* in the intermediate region of constant state be computed. This is done using the first equation [10.96]. Note that in this equation, $\lambda^{(1)}$ and $\lambda^{(2)}$ must not be replaced with λ^- and λ^+ , the role of which is only to provide a unified formula to handle the subcritical/supercritical transition at the interface.

A pending question is the expression of the derivatives $\partial\lambda^\pm/\partial\varphi$ in equation [10.99]. As proposed in [GUI 09a, GUI 09b], these derivatives are computed by noting that λ^- and λ^+ are known functions of the left and right states U_L and U_R :

$$\frac{\partial\lambda^\pm}{\partial\varphi} = \frac{\partial\lambda^\pm}{\partial U_L} \frac{\partial U_L}{\partial\varphi} + \frac{\partial\lambda^\pm}{\partial U_R} \frac{\partial U_R}{\partial\varphi} = \frac{\partial\lambda^\pm}{\partial U_L} s_L + \frac{\partial\lambda^\pm}{\partial U_R} s_R \quad [10.100]$$

The derivatives $\partial\lambda^\pm/\partial U_L$ and $\partial\lambda^\pm/\partial U_R$ are row vectors. Their product with the column vectors s_L and s_R gives a scalar quantity.

10.5.3.3. Dirac source term

The discretization of the Dirac source term R is examined. The source term $R_{i-1/2,-}^{n+1/2}$ is given by:

$$R_{i-1/2,-}^{n+1/2} = \begin{cases} \beta_1 \frac{\partial\lambda^{(1)}}{\partial\varphi} (s_L - s^*) + \beta_2 \frac{\partial\lambda^{(2)}}{\partial\varphi} (s^* - s_R) & \text{if } \lambda^{(2)} < 0 \\ \beta_1 \frac{\partial\lambda^{(1)}}{\partial\varphi} (s_L - s^*) & \text{if } \lambda^{(1)} < 0 < \lambda^{(2)} \\ 0 & \text{if } \lambda^{(1)} > 0 \end{cases} \quad [10.101]$$

where β_1 and β_2 are indicators, $\beta_p = 1$ if the wave p is a shock or contact discontinuity, $\beta_p = 0$ otherwise. Conversely, the source term $R_{i-1/2,+}^{n+1/2}$ is given by:

$$R_{i-1/2,+}^{n+1/2} = \begin{cases} 0 & \text{if } \lambda^{(2)} < 0 \\ \beta_2 \frac{\partial\lambda^{(2)}}{\partial\varphi} (s^* - s_R) & \text{if } \lambda^{(1)} < 0 < \lambda^{(2)} \\ \beta_1 \frac{\partial\lambda^{(1)}}{\partial\varphi} (s_L - s^*) + \beta_2 \frac{\partial\lambda^{(2)}}{\partial\varphi} (s^* - s_R) & \text{if } \lambda^{(1)} > 0 \end{cases} \quad [10.102]$$

The following criterion is used for shock detection [GUI 09a, GUI 09b]:

$$\begin{aligned}\beta_1 &= 1 \quad \text{if } \left\{ \begin{array}{l} \lambda^{(1)}(U_L) > \lambda^{(1)}(U^*) \\ \lambda^{(2)}(U_L) > \lambda^{(2)}(U^*) \end{array} \right\} \\ \beta_2 &= 1 \quad \text{if } \left\{ \begin{array}{l} \lambda^{(1)}(U^*) > \lambda^{(1)}(U_R) \\ \lambda^{(2)}(U^*) > \lambda^{(2)}(U_R) \end{array} \right\}\end{aligned}\quad [10.103]$$

10.5.4. Application example: the one-dimensional shallow water equations

The sensitivity Riemann solver is applied to the dambreak problem presented in section 10.3.4, with the parameters given in Table 10.1. More details can be found on the analytical solution of this problem and extensions to constant bottom slopes in [GUI 09a-c].

The numerical solution is computed over a domain discretized into 1,000 cells of width 1 m. The computational time step is set to the maximum permissible value given by the stability constraint $(u + c) \Delta t_{\max} = \Delta x$. Figure 10.9 shows the analytical and numerical solution for the sensitivity variables η , χ , v and θ . In contrast with the empirical solution shown in Figure 10.6, there is no artificial peak in the solution computed by the sensitivity solver.

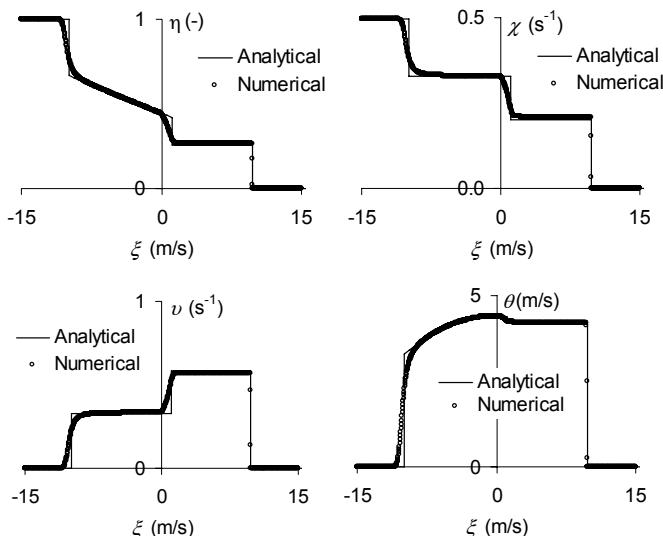


Figure 10.9. Dambreak problem. Analytical solution and numerical solution obtained using the modified HLL solver

10.6. Summary

Sensitivity equations for hyperbolic conservation laws or hyperbolic systems are obtained by differentiating the flow governing equations with respect to the parameter of interest. This parameter may be an initial condition, a boundary condition, or a parameter in the flux and/or in the source term. The sensitivity equations can be formulated in forward form (see sections 10.2 and 10.3) and in adjoint form (see section 10.4).

The wave speed of the sensitivity system are identical to the wave speeds of the flow system (sections 10.2 and 10.3). Consequently, the system formed by the flow equations and their sensitivity equations is not strictly hyperbolic but linearly degenerate because all the eigenvalues in the system are double eigenvalues. In the adjoint formulation, invariants can be defined for the Lagrange multipliers. The wave propagation speeds of these adjoint Riemann invariants are identical to those of the flow system. The adjoint invariants are calculated by following the characteristics in the backward time direction.

Discontinuous flow solutions generate Dirac sensitivity source terms that take effect at the flow discontinuities (sections 10.2.3 and 10.3.3). The jump relationships for the sensitivity are given by equation [10.19]. In the solution of the Riemann problem (sections 10.2.4 and 10.3.3), the sensitivity is discontinuous at the edges of rarefaction waves. In the solution of the Riemann problem for scalar conservation laws, rarefaction waves are zero sensitivity regions. The analytical sensitivity solution of the dambreak problem for the Saint Venant equations is derived in section 10.3.4.

Numerical methods are available for the discretization of the forward sensitivity equations in the presence of discontinuous solutions. A modified HLL Riemann solver is presented in section 10.5, with an application to the dambreak problem derived in section 10.3.4.

Chapter 11

Modeling in Practice

11.1. Modeling software

11.1.1. *Introduction*

The purpose of this chapter is to advise on a number of modeling precautions and to provide a few guidelines to users of modeling software packages.

In an industrial context, the choice of a specific modeling package is most often not made by the engineer or technician who operates it. Whether in a research team or in a consulting company, the selection (or development policy) of the modeling tool proceeds from numerous considerations. The quality and accuracy of the modeling results is only one of the many criteria used in the selection process. This is because numerical modeling is increasingly integrated into multi-criteria decision-making processes. Sophisticated algorithms and numerical techniques are nowadays “encapsulated” in user-friendly modules, served by efficient graphical interfaces with the purpose of facilitating modeling result interpretation and decision-making. To give but one example, commercial river flow modeling packages still use numerical schemes (for example, Preissmann’s scheme presented in Chapter 6) developed in the 1960s. For most of them, the development effort has concentrated on graphical interfaces or exchange modules with graphical or decision support tools such as database management systems, geographical information systems, etc.

As indicated in Chapters 6 to 9, numerical techniques provide only approximations to the solutions of the governing equations. Some techniques are more accurate than others, some are more computationally efficient. Sometimes, accuracy and/or computational efficiency are optimal only for a specific type of

situation (e.g. rapid or slow transients, in the presence or absence of solution discontinuities, etc.). There is often little room for the modeling engineer to question the choice of a given software or numerical technique used in his team or company. The engineer's task is rather to be as much aware as possible of the limitations (in terms of robustness, accuracy, validity of model assumptions) of the modeling package used, so as to minimize the risk of misuse and optimize the quality of the modeling results.

The main two questions that must be answered by a modeler are (i) what are the key phenomena involved in the configuration to be modeled, and (ii) what are the basic requirements to be fulfilled by the numerical method so as to guarantee the quality of the numerical solution? Two basic examples of such modeling issues are given in the next sections.

11.1.2. Conservation

In many engineering applications related to environmental fluid mechanics, conservation is a key issue. As an example, mass conservation (both in terms of water and solute transport) may seem a natural requirement for a river flow modeling package. Mass conservation implies that the variation in the amount of water (or contaminant) stored within a reach is equal to the difference between the discharge across the upstream and downstream sections. If pollution or water resource allocation studies are to be carried out, conservation appears as an indispensable prerequisite in the model selection process. This simple condition, however, may not be satisfied if the modeling package solves the wrong form of the equations, as shown in the example hereafter.

The continuity equation for transient flow in a channel has been presented in section 1.5.1. The conservation form is given by equation [1.84], recalled here:

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0$$

where A is the channel cross-sectional area and Q is the volume discharge. In some software packages, however, the dependent variables are not A and Q , but A and $u = Q/A$, where u is the flow velocity. Since $Q = Au$, equation [1.84] can be rewritten as:

$$\frac{\partial A}{\partial t} + A \frac{\partial u}{\partial x} + u \frac{\partial A}{\partial x} = 0 \quad [11.1]$$

In the case of smooth channel geometry, equations [11.1] and [1.84] can be approximated with a similar order of accuracy by a given numerical technique. If the geometry of the channel is locally discontinuous (see Figure 11.1), A and u are locally discontinuous (while Q is not) and thus locally non-differentiable. Approximating the terms $u \partial A / \partial x$ and $A \partial u / \partial x$ may lead to mass conservation problems.

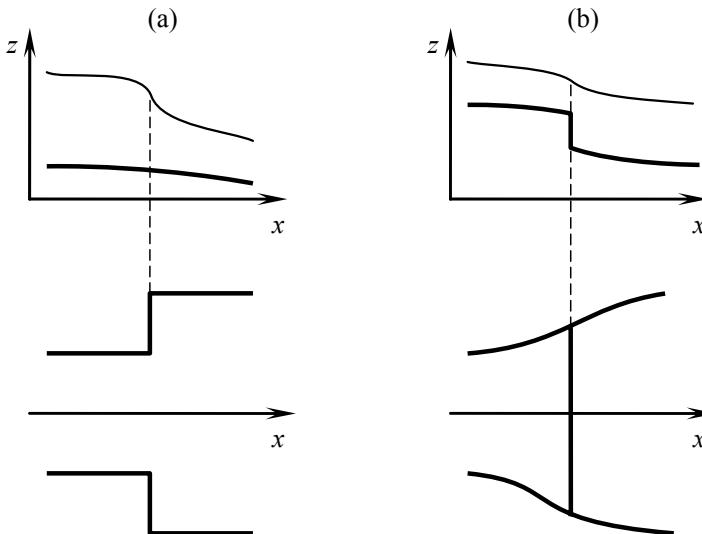


Figure 11.1. Typical examples of discontinuous geometries in river modeling: sudden widening (a), bottom step (b). Side view (top), plan view (bottom)

Similar conservation problems may appear if the term $\partial A / \partial t$ is rewritten as:

$$\frac{\partial A}{\partial t} = b \frac{\partial h}{\partial t} = b \frac{\partial \zeta}{\partial t} \quad [11.2]$$

where b is the top width, h is the water depth and ζ is the free surface elevation (see section 2.5.2 and Figure 2.12 for the notation). If the channel width is a discontinuous function of z (Figure 11.2), b may become discontinuous and the estimate of $b \partial h / \partial t$ (or $b \partial \zeta / \partial t$) may become incorrect.

In the example of Figure 11.2, the section is piecewise rectangular. The free surface width switches discontinuously from b_1 to b_2 at $z=z_1$. The derivative $b = \partial A / \partial h$ is thus undefined for $z=z_1$. Assume that the free surface elevation ζ is lower than z_1 (thus $b=b_1$) at the beginning of the computational time step and

higher than z_1 (thus $b = b_2$) at the end of the computational time step. Any purely explicit or purely implicit estimate of b in the term $b \partial h / \partial t$ yields an incorrect estimate for $\partial A / \partial t$.

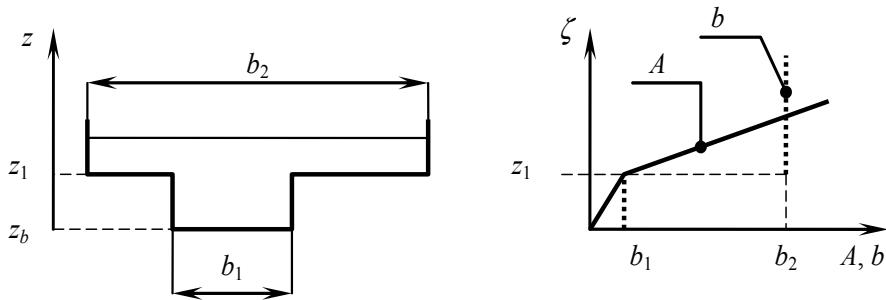


Figure 11.2. Example of a channel geometry with a discontinuous top width. Left: cross-sectional view of the geometry. Right: variations in A and b with the free surface elevation ζ

Even if the geometry is smooth, not solving the conservation form of the equations may yield erroneous solutions in the presence of discontinuous solutions (shock waves in the field of gas dynamics; hydraulic jumps or moving bores in free surface hydraulics). This is due to the non-uniqueness of weak solutions (see Chapter 3 for more details). Indeed, the conservation, non-conservation and characteristic forms of the governing equations are not equivalent in the presence of discontinuous solutions because the partial derivatives are locally undefined.

This is illustrated by the dambreak simulation shown in Figure 11.3.

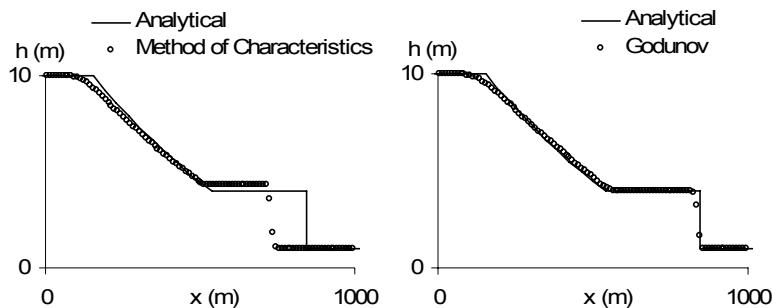


Figure 11.3. Dambreak problem. Solution at $t = 35$ s

The problem and its analytical solution are presented in detail in section 4.3.3. Recall that the solution consists of a region of constant state separated from the upstream and downstream sides of the dam by a rarefaction wave and a shock respectively. The initial water levels on the left- and right-hand sides of the dam are respectively 10 m and 1 m. Two numerical techniques are used to solve this problem. Figure 11.3 (left) shows the results obtained at $t = 35$ seconds with the finite difference-based, first-order method of characteristics (see section 6.2). Figure 11.3 (right) shows the results obtained using the finite volume-based, Godunov scheme (section 7.2), with fluxes calculated by the HLL approximate Riemann solver (see section C.1 in Appendix C for a description of this solver).

The first-order method of characteristics yields an overestimated water depth in the intermediate region of constant state, while the speed of the shock is underestimated. It is clearly visible from the figure that the total volume of water is not conserved with this method. A numerical integration indicates that the volume of water per unit width computed by the method of characteristics is $5,168 \text{ m}^2$, for an initial volume per unit width of $5,520 \text{ m}^2$. Approximately 6% of the total volume is lost artificially over the 35 simulated seconds. In contrast, the volume is preserved exactly in the Godunov simulation.

11.1.3. Solution monotony

The solutions of hyperbolic systems of conservation laws are essentially TVD. The practical consequence is that oscillations cannot develop spontaneously in the solutions of hyperbolic systems (unless specific combinations of boundary conditions and/or source terms are used). This is why TVD and upwind schemes are particularly praised by environmental modelers.

As an example, centered schemes are not popular in the field of contaminant transport modeling because they induce artificial oscillations in the computed profiles. For instance, artificial oscillations around a small or zero background concentration may yield locally negative concentration values in the numerical solution of a contaminant transport model. Although such oscillations may be easily justified from a mathematical point of view on the basis of truncation error analyses or from scheme phase and amplitude portraits (see Appendix B), their physically unrealistic character makes them hardly acceptable to decision-makers. The modeler is left with two options:

(1) Using a TVD scheme allows the monotone character of the solution to be preserved, thus eliminating physically unrealistic solutions. Moreover, the numerical diffusion applied by TVD schemes in the neighborhood of steep gradients contributes to make the solution more “realistic” than non-TVD schemes, because they imitate the diffusion and dissipation mechanisms that are present in natural

processes. It must be remembered however that numerical diffusion or dispersion proceed from truncation errors and that they should be seen as a sign that the numerical solution is not optimally accurate.

(2) In contrast, non-TVD schemes such as centered schemes (see section 6.5) minimize numerical dissipation. For this reason, they are most appreciated in modeling fields where numerical diffusion and dissipation are seen as “parasitic” phenomena that jeopardize the quality of the solution. This is the case for instance in the field of turbulence modeling, where upwind schemes are considered too dissipative.

In solving real-world modeling problems, the choice of a numerical technique most often results from a trade-off between the accuracy of the numerical solution and the monotony or positivity properties of the analytical solution that are deemed essential by the modeler.

11.2. Mesh quality

Meshing is an important step in the modeling process. Despite the availability of efficient mesh generation packages, meshing complex geometries still requires human supervision. Optimal result quality is achieved if the mesh is regular and isotropic. These issues are illustrated with the two-dimensional shallow water equations.

Mesh regularity. Numerical methods usually need 3 to 5 grid points (or cells, or elements) to represent steep gradients or discontinuities, as shown by the numerical results presented in Chapters 6 to 8. Strongly diffusive numerical methods may induce front smearing over 10 to 20 cells (see the tests presented in Chapter 8). A strong size contrast between neighboring cells in a grid may amplify the artificial smearing. The consequence may be an artificial damping of transients in some parts of the model.

An example of mesh generation is given in Figure 11.4. The purpose is to generate a mesh for a river main channel and floodplain simulations. The width of the river banks imposes the minimum size for the elements. The banks being narrow compared to the floodplain, the modeler may be tempted to generate a mesh with elements rapidly increasing in size in the direction of the floodplain. The ratio of the area of the smaller to the larger elements is approximately 250. The transition takes place within 4 to 5 elements, which implies an area ratio of 3 to 4 between two neighboring cells. A ratio of 1.5 to 2 is usually advised.

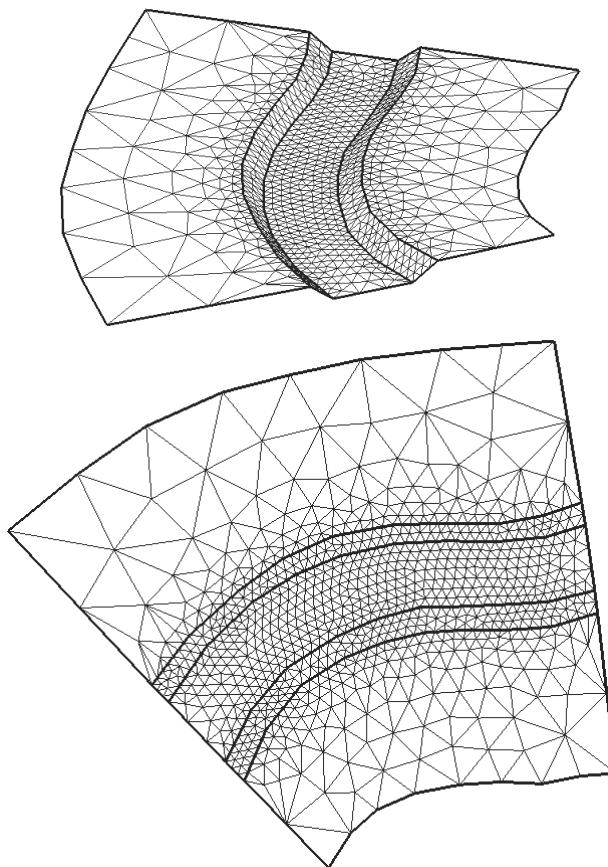


Figure 11.4. Example of a strongly irregular mesh.
Top: perspective view. Bottom: plan view

Mesh isotropy. Another issue in two-dimensional mesh generation is the meshing of long and narrow geometric features, such as dikes, roads, river embankments, etc. Using long and narrow elements may allow for a substantially reduced computational effort. The modeler is usually inclined to stretch the mesh in the direction of the flow in the main channel and along the river banks (Figure 11.5). The aspect ratio of the cells used to discretize the channel embankments in Figure 11.5 is between 7 and 10.

This meshing approach is traditionally justified with the argument that the Courant number, that is the key parameter to the quality of the numerical solution, should be the same in the longitudinal and transverse direction in order to maximize solution accuracy. If the purpose is to simulate passive scalar advection, the Courant number is determined by the flow velocity vector and stretching the mesh in the longitudinal direction is justified.

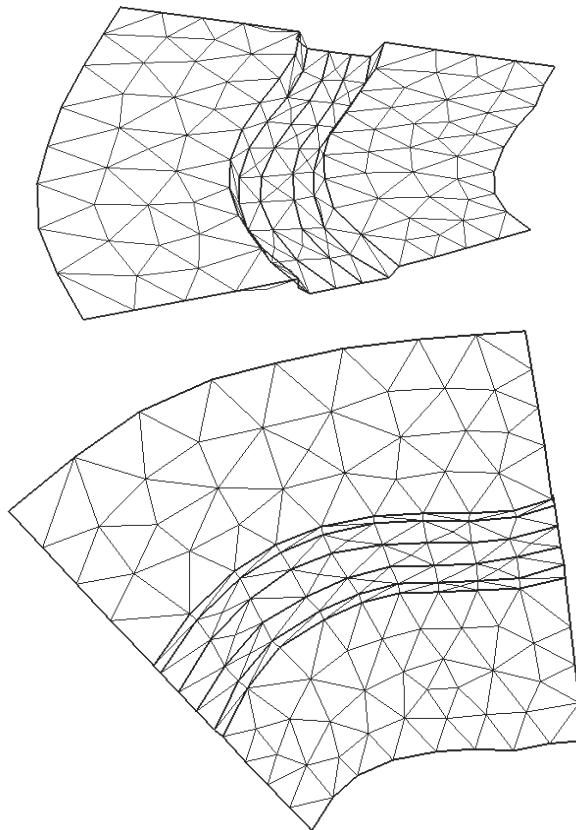


Figure 11.5. Example of a strongly anisotropic mesh.
Top: perspective view. Bottom: plan view

If the purpose is to simulate floodplain dynamics and two-dimensional free surface transients, however, the Courant number is determined from the propagation speed of the waves. As shown in Chapter 5, the domain of dependence of the solution is made of two surfaces in the phase space (Figure 11.6). The first surface is the curved line with tangent vector $(u, v, 1)$. The second surface is a conical surface,

expanding from the first surface at a speed $c = (gh)^{1/2}$. Two situations may be considered:

– Subcritical flow: the plan view shape of the dependence domain of the solution over a time step Δt is circular (Figure 11.6a). The domain of dependence of the point A in Figure 11.6a is a circle centered around B, that is shifted from A by a distance $(u^2 + v^2)^{1/2} \Delta t$. The domain of dependence is isotropic. Stretching the computational cell in the direction of the flow artificially increases the weight of lateral elements in the estimate of the gradients.

– Supercritical flow (Figure 11.6b): the point A is not included in the domain of dependence of the solution. The shape of the domain of dependence becomes narrower as the Froude number (the Mach number in the case of gas dynamics simulations) increases. Figure 11.6b shows the shape of the domain of dependence for a Froude number $Fr = 2.5$. In this case, the ratio of the longitudinal to transverse dimensions of the domain of dependence is $(2.5 + 1)/2 = 1.75$. Stretching the cells by a factor larger than 2 in the longitudinal direction is not justified.

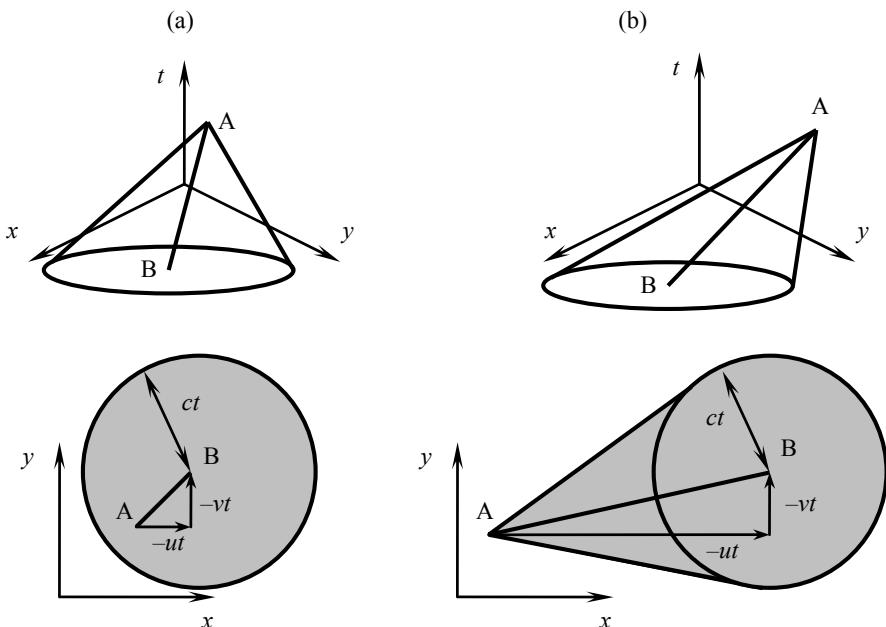


Figure 11.6. Domain of dependence of a point A. Subsonic/subcritical case (a), supersonic/supercritical case (b). Top: perspective view. Bottom: plan view

The longitudinal and transverse dimensions of the domain of dependence of point A are given by:

$$\left. \begin{aligned} D_L &= (u^2 + v^2)^{1/2} + c - \min[(u^2 + v^2)^{1/2} - c, 0] \\ D_T &= 2c \end{aligned} \right\} [11.3]$$

the minimum operator, $\min()$ allows a single expression to be obtained for both subcritical/subsonic and supercritical/supersonic conditions. The aspect ratio of the domain of dependence is thus obtained as:

$$\frac{D_L}{D_T} = \frac{1}{2} [\text{Fr} + 1 - \min(\text{Fr} - 1, 0)] = 1 + \frac{\text{Fr} - \min(\text{Fr}, 1)}{2} [11.4]$$

where $\text{Fr} = (u^2 + v^2)^{1/2} / c$ is the Froude number. In the field of gas dynamics, the Froude number is replaced with the Mach number M , thus yielding the following formula:

$$\frac{D_L}{D_T} = 1 + \frac{M - \min(M, 1)}{2} [11.5]$$

Equations [11.4] and [11.5] show that stretching the mesh in the longitudinal direction is not justified in the case of subcritical/subsonic flow configurations if the purpose is to solve the hydrodynamic equations. Strong mesh aspect ratios should be used only in the case of scalar transport.

Mesh anisotropy does not have the same consequences on solution accuracy for explicit and implicit schemes:

- if the numerical scheme used is explicit, numerical diffusion is higher when the Courant number is smaller (see the amplitude portraits shown in Appendix B). Consequently, gradient smearing is stronger in the direction the mesh is stretched;

- if an implicit numerical scheme is used, numerical diffusion is usually stronger for larger Courant numbers. Gradient smearing occurs preferentially in the direction perpendicular to mesh stretching. A classical consequence of this is the artificial polarization of the velocity field along the wider dimension of the mesh.

11.3. Boundary conditions

11.3.1. Number and nature of boundary conditions

The number of conditions to be prescribed at a model boundary is the number of characteristics entering the computational domain (see Chapters 1 and 2, [CUN 80]). For a one-dimensional configuration, the following holds:

- For all the scalar laws presented in this book, the propagation direction of the characteristics is that of the flow. Consequently, one boundary condition is needed at each inflowing boundary. Typically, the boundary condition is supplied in the form of a prescribed value (prescribed concentration in the case of contaminant transport, water saturation for the Buckley-Leverett model, flow velocity for the inviscid Burgers equation) or a prescribed flux F at the boundary.
- The water hammer equations (section 2.4) are characterized by two constant, opposite wave speeds. There is one incoming characteristic at each domain boundary. One boundary condition is needed at each end of the model. The most classically used conditions are prescribed pressure, prescribed discharge, and pressure-discharge relationships (e.g. pumps).
- In the Saint Venant equations (section 2.5), the number of boundary conditions is a function of the flow regime. Supercritical inflow requires two boundary conditions, supercritical outflow requires none. Both subcritical inflow and outflow require one boundary condition. Typical boundary conditions in free surface hydraulics are prescribed water level, prescribed discharge and stage-discharge relationships.
- In the Euler equations (section 2.6), the number of boundary conditions is also a function of the flow regime. Supersonic inflow requires three boundary conditions, supersonic outflow requires none. Subsonic inflow requires two boundary conditions, subsonic outflow requires only one. Classical types of boundary condition are prescribed pressure, prescribed flow velocity, prescribed density.

Almost all market-available simulation software packages are equipped to deal with all these types of boundary conditions, thus leaving the modeler with an unbounded number of possibilities to model a given situation. The modeler should be aware, however, that some combinations of boundary conditions must be used with extreme care, and that certain boundary condition types may not be used at all boundaries of the model. He should also be aware that the model may not be able to prescribe the desired boundary value in certain situations. These aspects are explored in the following sections.

11.3.2. Prescribed discharge/flow velocity

Prescribed discharge conditions are classical in pipe transient and free surface flow simulations. Gas dynamics simulations rather use prescribed flow velocity conditions. Prescribing such conditions at the upstream boundary of a computational domain usually poses no problem. Prescribing an outflowing discharge or velocity at the downstream boundary of a model may jeopardize the simulation. The reasons for this are the following:

(1) There exists a maximum possible value for the prescribed discharge or flow velocity at a downstream boundary. This is true for the water hammer equations, Saint Venant and Euler equations. This point is illustrated with the Saint Venant equations in a frictionless horizontal, rectangular channel. In this case, the characteristic form of the equations (see section 2.5) simplifies to:

$$\left. \begin{array}{l} u - 2c = \text{Cst} \\ u + 2c = \text{Cst} \end{array} \right\} \begin{array}{l} \text{for } \frac{dx}{dt} = u - c \\ \text{for } \frac{dx}{dt} = u + c \end{array} \quad [11.6]$$

Consider a channel with water depth and flow velocity h_0 and u_0 next to the right-hand boundary. Since the water is flowing out of the channel, the flow velocity u_0 is assumed positive. The second characteristic equation [11.6] leads to:

$$u_0 + 2(gh_0)^{1/2} = u_b + 2(gh_b)^{1/2} \quad [11.7]$$

where h_b and u_b the water depth and flow velocity at the downstream boundary. Equation [11.7] can be rewritten as:

$$u_b = u_0 + 2(gh_0)^{1/2} - 2(gh_b)^{1/2} \quad [11.8]$$

The unit discharge $q_b = h_b u_b$ at the downstream boundary is therefore:

$$q_b = u_b h_b = \left[u_0 + 2(gh_0)^{1/2} - 2(gh_b)^{1/2} \right] h_b \quad [11.9]$$

The function $q_b(h_b)$ is zero for $h_b = 0$ and for a depth h_{\max} defined as:

$$h_{\max} = \frac{\left[u_0 + 2(gh_0)^{1/2} \right]^2}{4g} \quad [11.10]$$

Figure 11.7 illustrates the behavior of the function $q_b(h_b)$.

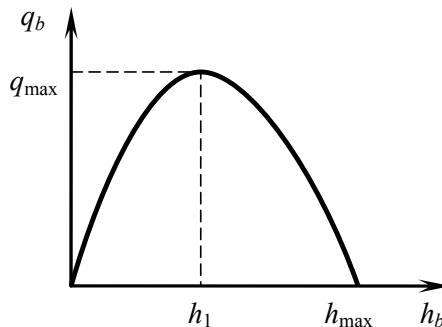


Figure 11.7. Unit discharge q_b at the downstream boundary of the domain as a function of the downstream water depth h_b

The function q_b is maximum for $h_b = h_1$ such that $\mathrm{d}q_b/\mathrm{d}h_b = 0$, that is:

$$h_1 = \frac{\left[u_0 + 2(gh_0)^{1/2}\right]^2}{9g} = \frac{4}{9}h_{\max} \quad [11.11]$$

The maximum possible prescribed discharge q_{\max} at the downstream boundary is given by:

$$q_{\max} = \left[u_0 + 2(gh_0)^{1/2} - 2(gh_1)^{1/2}\right] h_1 = \frac{1}{g} \left[\frac{u_0 + 2(gh_0)^{1/2}}{3} \right]^{3/2} \quad [11.12]$$

(2) It is not possible to prescribe an outflowing velocity yielding a Froude number (Mach number for gas dynamics simulations) larger than unity at the downstream boundary of the domain.

(3) Prescribing the discharge at both ends of a channel reach is not advised [CUN 80]. Indeed, this is an indirect way of prescribing the amount of fluid stored in the domain. This may lead to simulation problems if the outflowing discharge is larger than the inflowing discharge, because the net mass balance may exceed the quantity of water available in the model at some stage.

11.3.3. Prescribed pressure/water level

In free surface hydraulics, prescribed water level conditions are usually met at the downstream end of river models. An exception is the simulation of tidal flows,

where the water level may be prescribed at both ends of the modeled reach. In the field of water hammer and gas dynamics simulations, the pressure may be prescribed at any model boundary.

It is not always possible to prescribe the desired value of the pressure or water level at a domain boundary. This is because prescribing too low a pressure or level may trigger supercritical outflow, a situation where the flow variables at the boundary are entirely determined from the low conditions inside the domain. Since all the characteristics leave the domain, attempting to prescribe a pressure or water level in such situations is meaningless.

The example of the Saint Venant equations in a rectangular, horizontal channel presented in section 11.3.2 is used again. The Froude number at the boundary is obtained from equation [11.8] as:

$$\text{Fr}_b = \frac{u_b}{c_b} = \frac{u_0 + 2(gh_0)^{1/2} - 2(gh_b)^{1/2}}{(gh_b)^{1/2}} \quad [11.13]$$

It is easy to check that Fr_b is a decreasing function of h_b and that $\text{Fr}_b = 1$ for $h_b = h_1$ as defined in equation [11.11]. The behavior of the function $\text{Fr}_b(h_b)$ is illustrated in Figure 11.8:

- For $h_b > h_1$, Fr_b is smaller than unity. The characteristic $(u - c)$ enters the domain. A boundary condition may be prescribed.
- For $h_b < h_1$, Fr_b is larger than unity. The characteristics $(u - c)$ and $(u + c)$ leave the domain and h_b cannot be prescribed at the boundary.

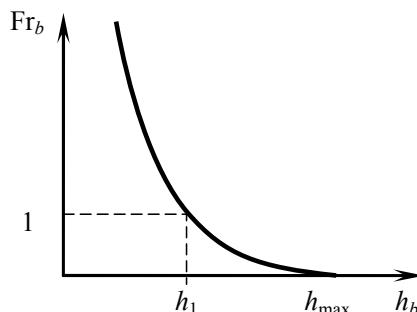


Figure 11.8. Froude number Fr_b at the right-hand boundary as a function of the downstream water depth h_b

The above reasoning remains valid for any hyperbolic system with wave speeds that may change sign, such as the Euler equations.

11.3.4. Stage-discharge and pressure-discharge relationships

In the field of pipe transient modeling, pressure-discharge relationships are used to represent head losses across a singularity, one side of which is connected to a node with a fixed pressure. They are also used to represent pumps taking water from a source with a known pressure. In the field of free surface flow modeling, stage-discharge relationships are often used to provide a condition at a boundary of the model where no measurement is available for the water level or discharge. In such a case, the stage-discharge relationship is a function of the channel geometry and is derived from specific assumptions on the flow regime: the assumption of a uniform or critical flow allows the discharge to be inferred from the water level. Using the assumption of uniform flow leads to equation [1.83], recalled here:

$$Q = \frac{1}{n_M} \frac{A^{5/3}}{\chi^{2/3}} S_0^{1/2}$$

where A is the cross-sectional area, n_M is Manning's friction coefficient, S_0 is the bed slope and χ is the wetted perimeter. The assumption of critical flow leads to:

$$Q = \left(\frac{g}{b} \right)^{1/2} A^{3/2} \quad [11.14]$$

where b is the top width of the channel.

Pressure-discharge and stage-discharge relationships must be used with care. In particular:

(1) the inflowing discharge at an upstream boundary should always be a decreasing function of the pressure or water level,

(2) the outflowing discharge at a downstream boundary should always be an increasing function of the pressure or water level.

The computational solution may become unstable if these rules are not satisfied. The reason for this is the following:

- Consider an inflowing (upstream) boundary for a water hammer model. Assume that the discharge is an increasing function of pressure. Any increase in the pressure at the boundary (such an increase may be triggered by a transient

propagating in the pipe) results in an increase in the discharge. From continuity, increasing the discharge at the upstream boundary triggers fluid compression, which triggers an increase in the pressure. This in turn generates an increase in the discharge. This cyclic behavior results in instability. Note that the reasoning also holds for a decrease in the discharge. Similar arguments may be used for the shallow water equations (with the difference that the pressure is replaced with the water level).

- Consider an outflowing boundary. Assume that the discharge is a decreasing function of pressure. Any increase in the pressure resulting from transients generated within the domain yields a decrease in the discharge at the downstream boundary. From continuity, this triggers fluid compression, thus leading to a pressure increase. This triggers a new decrease in the pressure. Repeating the cycle may lead to instability in this case too.

- In contrast, prescribing a decreasing pressure-discharge condition at the upstream boundary or an increasing pressure-discharge relationship at the downstream boundary contributes to stabilizing the solution.

11.4. Numerical parameters

11.4.1. Computational time step

Most numerical methods presented in Chapters 6 to 10 achieve optimal accuracy when the absolute value of the Courant number of the waves is close to one. When the flow and geometry are highly variable in space, it is not possible to maintain the same value of the Courant number at all points for all waves. The computational time step often results from a trade-off between solution accuracy and computational efficiency. Recall that:

- explicit schemes are subjected to a stability constraint. This constraint imposes that the absolute value of the Courant (also called CFL) number of the faster wave should not be larger than unity. Simulation packages using explicit methods may reduce the computational time step during the simulation so as to enforce the stability constraint. The computational time step actually used may not always be the computational time step requested by the user;

- implicit schemes are not subjected to stability constraints. Commercially available simulation packages classically use the time step requested by the user, even when this yields very large values of the Courant number. It is advised that the modeler estimate roughly the typical wave speeds to be encountered during the simulation, so as to specify a computational time step that will ensure an average Courant number close to one.

11.4.2. Scheme centering parameters

Implicit schemes such as the Preissmann scheme or the finite element schemes presented in Chapter 8 use a time-centering parameter θ . This parameter is used to weight the respective contributions of the time levels n and $n + 1$ in the calculation of the space derivatives. $\theta = 1/2$ give the same weight to the known time level n and the unknown time level $n + 1$. The larger θ , the larger the contribution of the unknown time level $n + 1$. Centered schemes, such as the Crank-Nicholson or Galerkin technique with symmetrical shape functions, give unstable solutions for $\theta < 1/2$. When θ is set to $1/2$, oscillations usually appear in the computed solutions when $C_r \neq 1$ because the numerical diffusion in the truncation error is set to zero, thus leaving room for numerical dispersion (see Appendix B for detailed considerations). Increasing θ allows numerical diffusion to be increased, thus leading to solution stabilization, profile smoothing and artificial wave damping.

The Preissmann scheme uses an additional, space centering parameter ψ . Such a parameter is also used or wave speed interpolation by the semi-implicit finite element techniques presented in Chapter 8. When hyperbolic systems are dealt with, with waves traveling in opposite directions, using $\psi = 1/2$ is advised because this allows the waves with positive and negative speeds to be treated in a symmetrical way.

11.4.3. Iteration control

Implicit schemes for the solution of nonlinear systems involve iterative procedures. Iterations are also needed in the case of the Alternate Directions Implicit (ADI) technique for the solution of multidimensional systems (see section 6.9.2). The question arises of the criteria used to assess the degree of convergence of the iterative procedure. Four options are available:

- the number of iterations is pre-defined by the user of the numerical technique. The same number of iterations is made at each computational time step, regardless of the degree of convergence of the solution at the end of the iterative loop;
- the user defines iteration stop criteria. Convergence may be checked by computing the residual of the system to be solved, or the difference between two successive values of the solution from one iteration to the next. Iterations are stopped when the residual or the difference between two successive iterations falls below a predefined threshold. Some packages also allow a maximum permissible number of iterations to be defined. The iterative procedure is stopped when this maximum number of iterations is reached, regardless of the state of convergence of the solution;

- the number of iterations and/or the convergence criteria are determined automatically by the software. Although interesting at first sight because this does not require the supervision of an experienced user, this option is questionable because the user loses any control on the degree of accuracy of the solution;
- some packages simply perform no iteration, which allows the computational cost of the solution procedure to be reduced dramatically. This, however, may result in strongly degraded solutions.

If a convergence criterion is to be specified by the user, it may be worth documenting the formula used. In general, two options are available in computational hydraulics packages:

- convergence is checked locally: the residual of the system to be solved (or the difference between two successive iterations) is computed at each point (or cell, or node) of the computational grid. Convergence is achieved if each of the computational points falls below the predefined iteration stop criteria. Convergence implies that the numerical solution is satisfied at all points with a satisfactory degree of accuracy. The drawback of this approach is the large number of iterations often required;
- convergence may be checked on average over the computational domain. For instance, an average value is computed for the residual over the entire domain; or the average value of the difference between two successive iterations is used for comparison with the convergence criteria. This latter approach is faster than the former because there are always areas in the computational domain where convergence is achieved faster than in other areas. The solution may be considered “converged” in an average sense even though there are regions in the domain where convergence is far from being achieved. The resulting error may propagate into the rest of the computational domain at later times. For this reason, specifying strict convergence criteria may prove beneficial in the long term, even though the solution process may be slowed down at early simulation stages.

11.5. Simplifications in the governing equations

11.5.1. *Rationale*

Some of the numerical techniques presented in Chapter 6 are not equipped to deal with transcritical flow configurations. This is the case with the Preissmann scheme presented in Chapter 6. Although transcritical versions of the scheme have been proposed in the literature [JOH 02], they do not seem to have been implemented in industrial packages. To overcome this problem, practical implementations of these schemes in engineering free surface flow modeling software often solve a simplified version of the equations. In these software

packages, the governing equations are simplified so as to guarantee that the waves $(u - c)$ and $(u + c)$ always propagate in opposite directions. Examples of such techniques are the Local Partial Inertia (LPI) and the Reduced Momentum Equation (RME) approaches. Practical consequences of these techniques are detailed in [NOV 10], only the broad lines of the techniques are given here.

11.5.2. The Local Partial Inertia (LPI) technique

The LPI approach [JIN 00] consists of multiplying the inertial terms $\partial Q / \partial t$ and $\partial(Q^2 / A) / \partial x$ in the Saint Venant equations by a coefficient ε that decreases from one to zero as the absolute value of the Froude number approaches unity:

$$\varepsilon \frac{\partial Q}{\partial t} + \varepsilon \frac{\partial}{\partial x} \left(\frac{Q^2}{A} \right) + \frac{\partial}{\partial x} \left(\frac{P}{\rho} \right) = (S_0 - S_f) g A \quad [11.15]$$

This amounts to dividing the other terms of the equation by ε .

$$\frac{\partial Q}{\partial t} + \frac{\partial}{\partial x} \left(\frac{Q^2}{A} \right) + \frac{1}{\varepsilon} \frac{\partial}{\partial x} \left(\frac{P}{\rho} \right) = \frac{1}{\varepsilon} (S_0 - S_f) g A \quad [11.16]$$

The Jacobian matrix A is modified into:

$$A = \begin{bmatrix} 0 & 1 \\ c^2/\varepsilon - u^2 & 2u \end{bmatrix} \quad [11.17]$$

The eigenvalues of A are:

$$\left. \begin{aligned} \lambda^{(1)} &= u - \varepsilon^{-1/2} c \\ \lambda^{(2)} &= \frac{u}{\varepsilon} + \varepsilon^{-1/2} c \end{aligned} \right\} \quad [11.18]$$

These eigenvalues are identical to the exact eigenvalues $u - c$ and $u + c$ only when $\varepsilon = 1$. When the Froude number tends to one, ε tends to zero and the eigenvalues $\lambda^{(1)}$ and $\lambda^{(2)}$ as given by equation [11.16] tend to infinity. For Froude numbers larger than one, $\varepsilon = 0$ and the diffusive wave approximation is obtained. The diffusive wave model is not a hyperbolic model. Figure 11.9 shows the

variations of $\lambda^{(1)}/c$ and $\lambda^{(2)}/c$ with the Froude number with a weighting function ε given by:

$$\varepsilon = \max(0, 1 - Fr^2) \quad [11.19]$$

The theoretical wave speeds are $\lambda^{(1)}/c = Fr - 1$ and $\lambda^{(2)}/c = Fr + 1$. As illustrated in Figure 11.9, the wave speeds in the LPI approach depart from the theoretical values as the absolute value of the Froude number approaches unity. This should be expected because a large Froude number value means that the inertial terms play a significant role in the momentum equation. Neglecting these terms can only lead to incorrect wave speed estimates.

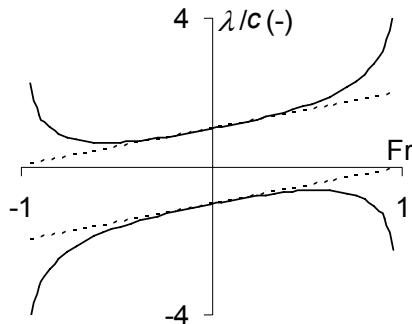


Figure 11.9. Wave speeds given by the LPI approach. Dashed lines: theoretical. Solid lines: equations [11.18–19]

11.5.3. The Reduced Momentum Equation (RME) technique

The RME approach [DHI 05] is similar in essence to the LPI approach, except that only the derivative $\partial(Q^2/A)/\partial x$ in the Saint Venant equations is multiplied by the weighting coefficient ε in the momentum equation:

$$\frac{\partial Q}{\partial t} + \frac{\partial}{\partial x} \left(\varepsilon \frac{Q^2}{A} + \frac{P}{\rho} \right) = (S_0 - S_f)gA \quad [11.20]$$

The behavior of ε with the Froude number is the same as in the LPI approach: ε decreases from one to zero when the absolute value of the Froude number

increases from 0 to one. Neglecting the variations in ε with Q and A , the following Jacobian matrix is obtained:

$$A = \begin{bmatrix} 0 & 1 \\ c^2 - \varepsilon u^2 & 2\varepsilon u \end{bmatrix} \quad [11.21]$$

The main difference with the LPI approach is that the system remains hyperbolic under supercritical conditions. The eigenvalues of A are:

$$\begin{aligned} \lambda^{(1)} &= \varepsilon u - \left[(\varepsilon - 1)\varepsilon u^2 + c^2 \right]^{1/2} \\ \lambda^{(2)} &= \varepsilon u + \left[(\varepsilon - 1)\varepsilon u^2 + c^2 \right]^{1/2} \end{aligned} \quad [11.22]$$

Figure 11.10 shows the variations of $\lambda^{(1)}/c$ and $\lambda^{(2)}/c$ with the Froude number with ε defined as in equation [11.19]. Note that the modified wave speeds do not change sign with the Froude number. This makes it impossible to reproduce supercritical flow conditions. For large Froude numbers, the wave speeds are equal to the propagation speeds of the waves in still water. As in the LPI technique, the inertial terms are neglected in the range of Froude numbers where they are predominant.

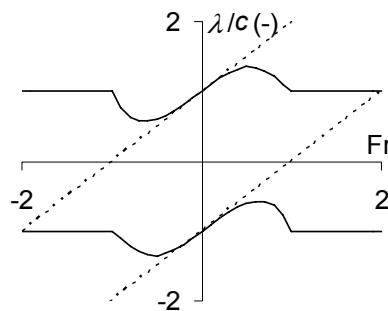


Figure 11.10. Wave speeds given by the RME approach.
Dashed lines: theoretical. Solid lines: equations [11.19–20]

11.5.4. Application examples

11.5.4.1. Steady flow over a bump

The RME technique is applied to the steady state shallow water test case presented in Chapter 9. The description and parameters of the test case can be found

in section 9.6. The reference solution is the solution obtained using the Auxiliary Variable-based Balancing (AVB) technique presented in section 9.5, with equations [9.91] and [9.101] for the estimate of Δh . This solution is plotted on the left-hand side of Figure 11.11.

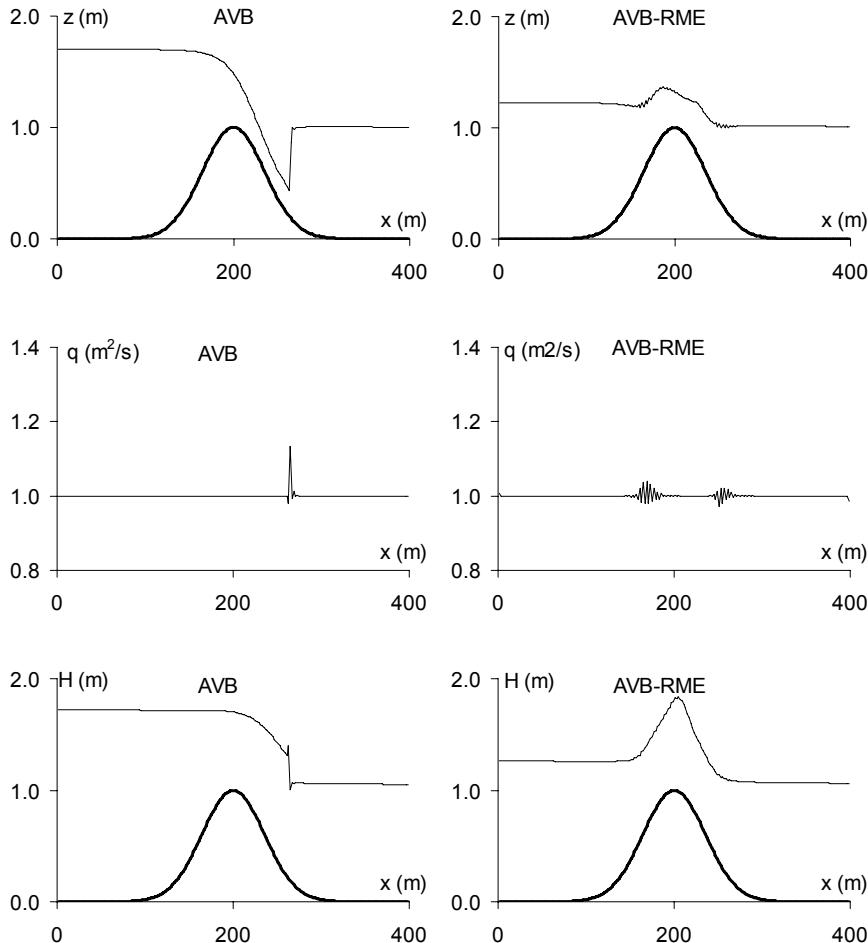


Figure 11.11. Steady flow over a bump computed using the Auxiliary Variable-based Balancing approach (AVB) and the AVB approach combined with the Reduced Momentum Equation approach (AVB-RME). Top: free surface elevation. Middle: unit discharge. Bottom: hydraulic head

The profiles on the right-hand side of Figure 11.11 are the water level, unit discharge and hydraulic head obtained by applying the RME approach and the AVB technique to the calculation of the fluxes. Note that the RME approach influences only the estimate of the momentum flux within the computational cells, by multiplying the term q^2/h with the coefficient ε .

The most striking feature is the impossibility for the RME approach to reproduce the hydraulic jump on the downstream side of the bump. This was to be expected because the RME approach does not allow supercritical conditions to be reproduced. The consequence is a local increase in the water level across the bump and a reduced head loss because friction is reduced due to the artificially increased water depth.

The peak discharge observed across the hydraulic jump in the original AVB approach is replaced with small amplitude oscillations on both sides of the bump. Another striking feature is the artificial increase in the hydraulic head induced by the RME approach across the bump. While the original AVB approach classically computes a decreasing head profile from upstream to downstream, applying the RME approach yields an increase on both sides of the bump. The singular head loss at the location of the jump is correctly identified by the AVB solution. Since the hydraulic jump cannot be represented in the RME approach, this singular head loss is not represented in the RME solution.

11.5.4.2. Dambreak problem

The RME technique is applied to the dambreak problem presented in section 10.3.4, with the parameters given in Table 1. Remember that the dambreak problem is a Riemann problem, for which an analytical solution is available (see Chapter 4, and more specifically section 4.3.3).

The numerical solution is obtained using a first-order finite volume scheme with a HLL Riemann solver, where the fluxes are computed using the RME approach. The cell size and time step are respectively $\Delta x = 1 \text{ m}$ and $\Delta t = 2 \times 10^{-2} \text{ s}$. Δt is approximately three times as small as the maximum permissible time step allowed by the stability constraint of the scheme. However, experience shows that using larger values for Δt yields sharp oscillations in the computed profiles in the transcritical region of the solution.

The numerical solution at $t = 30 \text{ s}$ is compared to the analytical solution in Figure 11.12. The formula of the flux being modified via the coefficient ε , the speed of the shock is modified in the numerical solution compared to the analytical solution. The solution in the intermediate region of constant state thus differs from the analytical solution.

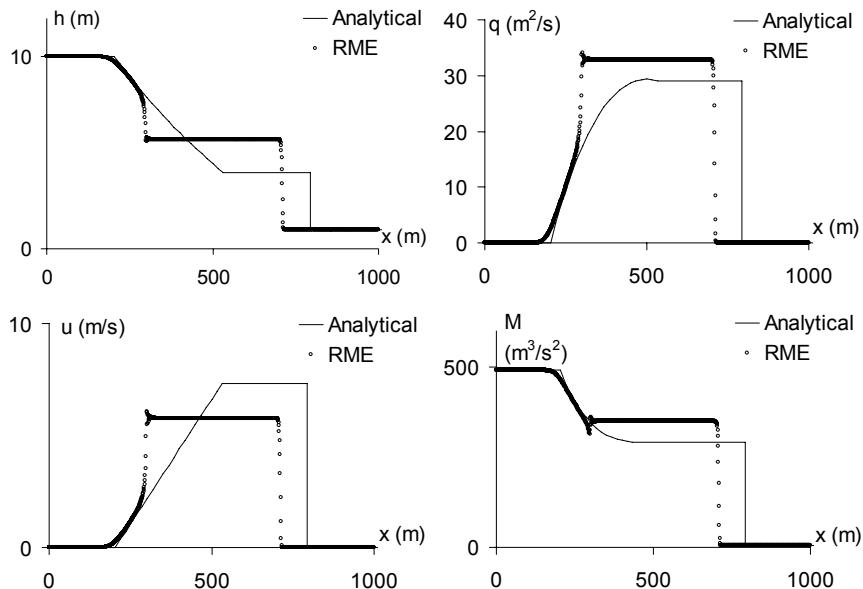


Figure 11.12. Dam-break problem. Analytical solution and numerical solution obtained using the RME approach

11.6. Numerical solution assessment

11.6.1. Software solution accuracy

The modeler's main concern is that the software package used should actually be able to solve the governing equations with a reasonable degree of accuracy. This leads to the key notion of numerical convergence: the numerical solution is said to converge to the analytical solution if it converges uniformly to it as the time step Δt and cell width Δx tend to zero. Solution convergence should be viewed as a necessary condition to model applicability in that it guarantees that the “actual” solution can be computed numerically with any arbitrary degree of accuracy provided that sufficient computational effort is spent in refining the discretization of space and time.

Convergence is related to the notions of consistency and stability. The purpose of this section is to give an overview of consistency, stability and convergence issues. Consistency and stability analysis techniques are presented in Appendix B.

11.6.2. Assessing solution convergence

A widespread way of assessing the convergence of the numerical solution consists of solving numerically a test case for which an analytical solution is available. In the case of a one-dimensional, scalar conservation law solved over a solution domain $[0, L]$, a numerical solution U_i^n and an analytical solution $U(x_i, t^n)$ is assumed to be available at all points x_i of the discretized domain $[0, L]$ for a given time t^n . The pointwise error, defined as the difference between the numerical and analytical solution, is computed as:

$$e_i^n = U_i^n - U(x_i, t^n) \quad [11.23]$$

The L_p -norm of the error (where p is an integer) is given by the numerical integral:

$$L_p = \left[\frac{1}{L} \sum_i^N |e_i^n|^p \Delta x_i \right]^{1/p} \quad [11.24]$$

The most widely used measure of error is the L_2 -norm. Sometimes, the L_∞ -norm is used. It is obtained as the limit expression of equation [11.24] when p tends to infinity as:

$$L_\infty = \max_i |e_i^n| \quad [11.25]$$

Solution convergence is achieved by computing the numerical solution at a given time for different values of Δx . Note that Δt must be decreased proportionally to Δx so as to preserve a constant value for the Courant number and to preserve the stability properties of the solution. The L_p -norm is computed for each of these values of Δx . If L_p tends to zero as Δx tends to zero, the numerical solution is convergent.

The order of convergence of the solution is said to be α if a power law can be fitted to the experimental pairs $(\Delta x, L_p)$:

$$L_p = K \Delta x^\alpha \quad [11.26]$$

where K is a constant. A very simple way of estimating α consists of plotting L_p and Δx along logarithmic axes. α is the slope of the straight line that can be fitted to the set of experimental points $(\log(\Delta x), \log(L_p))$. Figure 11.13 shows an example of such

a graph. The solution given by Scheme 1 (circular dots) can be approximated by a straight line with slope 1 in logarithmic coordinates (the L_2 -norm of the error is multiplied by 10^2 when Δx is multiplied by 10^2). The solution given by Scheme 2 (cross-shaped dots) can be approximated with a line with slope 2 (the L_2 -norm of the error is multiplied by 10^4 when Δx is multiplied by 10^2). Therefore, the orders of convergence of the solutions obtained by Schemes 1 and 2 are respectively 1 and 2.

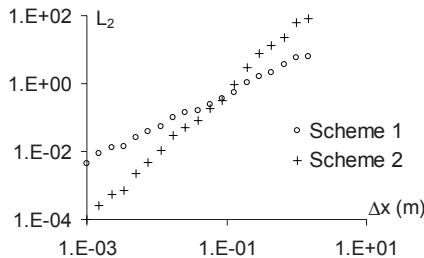


Figure 11.13. Examples of L_2 -norm of the error as a function of Δx for two different numerical schemes

11.6.3. Consistency analysis – numerical diffusion and dispersion

For the sake of simplicity, consider a scalar conservation law in the form [1.1], recalled here:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} = S$$

As shown in Appendix B (see section B.1), discretizing equation [1.1] (or its non-conservation, or characteristic form) leads us to solve a different equation that can be written in the form:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} = S + TE(\Delta x, \Delta t) \quad [11.27]$$

where the difference TE between the discretized equation and the original equation is called the truncation error. While the purpose is to solve equation [1.1], equation [11.27] is solved instead. The truncation error is made of an infinite sum of elementary terms formed by powers of Δx and Δt . These terms also include the higher-order derivatives of U with respect to x and t :

$$\begin{aligned}
 \text{TE}(\Delta x, \Delta t) &= \sum_{p=1}^{\infty} k_p \Delta x^{a_p} \Delta t^{b_p} \frac{\partial^{c_p} U}{\partial x^{d_p} \partial t^{c_p-d_p}} \\
 &= \sum_{p=1}^{\infty} k_p (\lambda \text{Cr})^{a_p} \Delta t^{a_p+b_p} \frac{\partial^{c_p} U}{\partial x^{d_p} \partial t^{c_p-d_p}} \\
 &= \sum_{p=1}^{\infty} k_p \frac{\Delta x^{a_p+b_p}}{(\lambda \text{Cr})^{b_p}} \frac{\partial^{c_p} U}{\partial x^{d_p} \partial t^{c_p-d_p}}
 \end{aligned} \tag{11.28}$$

where a_p , b_p , c_p and d_p are integer powers for the p th term in the truncation error and k_p is a coefficient that usually depends on the Courant number $\text{Cr} = \lambda \Delta t / \Delta x$.

Section B.1 in Appendix B shows how the expression for the truncation error is derived from the numerical scheme. When Δx and Δt tend to zero, the terms with smaller powers of Δx and Δt in equation [11.28] become predominant over the other terms. In the field of numerical methods for hyperbolic conservation laws, two main situations occur:

(1) Numerical diffusion occurs when the truncation error [11.28] can be written in the form:

$$\text{TE}(\Delta x, \Delta t) = D(\Delta x, \Delta t) \frac{\partial^2 U}{\partial x^2} + \text{HOT} \tag{11.29}$$

where HOT represents the sum of the higher-order terms in the sum [11.26]. D is a so-called numerical diffusion coefficient that usually depends on Δt , Δx and possibly the Courant number.

(2) Numerical dispersion is encountered when the truncation error [11.28] can be written in the form:

$$\text{TE}(\Delta x, \Delta t) = k_1(\Delta x, \Delta t) \frac{\partial^3 U}{\partial x^3} + k_2(\Delta x, \Delta t) \frac{\partial^3 U}{\partial x^2 \partial t} + \text{HOT} \tag{11.30}$$

where k_1 and k_2 are so-called dispersion coefficients.

Numerical diffusion and numerical dispersion influence the behavior of the numerical solution in very different ways (Figure 11.14). Numerical diffusion tends to smooth out the computed profiles (Figure 11.14a), thus leading to the damping of transients via dissipation of the energy contained in the solution signal. Numerical dispersion, in contrast, modifies the speed at which the various components of the solution propagate. Shifting these components with respect to each other introduces

oscillations in the computed profiles, especially in the neighborhood of steep gradients (Figure 11.14a).

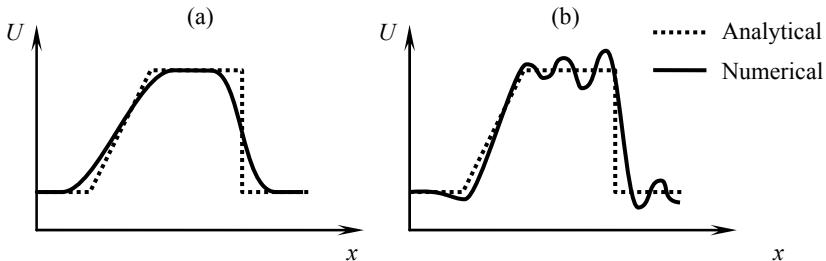


Figure 11.14. Typical effects of numerical diffusion (a) and numerical dispersion (b) on the numerical solution

Numerical dispersion destroys the TVD character of the numerical solution. In the field of linear PDEs or hyperbolic systems (passive scalar transport, see section 1.3; or water hammer problems, see section 2.4), numerical dispersion may simply result in non-physical results, such as negative concentrations or pressures. If the PDEs to be solved are nonlinear, numerical dispersion may result in nonlinear instability.

11.6.4. Stability analysis – phase and amplitude portraits

Linear (or harmonic) stability analysis provides valuable information on the performance of the numerical technique used. How amplitude and phase portraits should be derived for a given scheme is dealt with in section B.2.5 (Appendix B). The amplitude portrait of a scheme indicates how a harmonic component with wave length M in the solution is amplified from one time step to the next by the numerical scheme. The amplification factor $|A_N|$ is usually given as a function of the so-called wave number M :

$$M = \frac{L}{\Delta x} \quad [11.31]$$

The minimum possible value for M is 2 (at least two points are needed to represent one period of a harmonic component). Applying the numerical scheme over k computational time steps to a harmonic component of wave length L with a cell width Δx yields an amplification by a factor $|A_N|^k$. A steep front or discontinuity in the solution is represented by half a wavelength that is $L/2$.

Assume for instance that the explicit upwind scheme is used to compute the solution of the linear advection equation over 100 time steps ($k = 100$) with a Courant number $\text{Cr} = 0.5$. The amplitude portrait of the scheme is shown in Figure B.2. The numerical values of $|A_N|$ and $|A_N|^k$ ($k = 100$) are shown in Table 11.1 for various values of M .

M	$ A_N $	$ A_N ^{100}$	M	$ A_N $	$ A_N ^{100}$
2	0.0	0.0	30	0.995	0.58
10	0.951	6.6×10^{-3}	50	0.990	0.82
16	0.981	0.14	100	0.999	0.95
20	0.988	0.29	200	0.9999	0.99

Table 11.1. Amplification factor for $\text{Cr} = 0.5$ as a function of the wave number M

As the table indicates, if 16 cells are used to describe one wavelength of the harmonic component (8 cells for half a wavelength), only 14% of the initial amplitude remains after 100 time steps at $\text{Cr} = 0.5$. If 30 cells are used for one period (15 cells for half a wavelength), the harmonic component is damped by more than 40% after 100 computational time steps. The practical consequence of this is that an initially discontinuous profile is spread artificially over 8 to 15 cells within 100 computational time steps. Operating the explicit upwind scheme at $\text{Cr} = 0.5$ requires that any steep front in the initial profile be represented by at least 8 cells for the numerical solution to be reasonably accurate after 100 computational time steps. This is why the first-order upwind scheme is classically admitted to require approximately 10 cells to represent a steep front or a discontinuity.

11.7. Getting started with a simulation package

Sections 11.1 to 11.5 only provide examples of problems that may occur when industrial computational hydraulics packages are used. The list of issues raised in these sections is not exhaustive. The modeler's responsibility is to be aware of the limitations and weaknesses of the algorithms implemented in the software package he is using. Experience shows that the user of a modeling package quickly becomes used to this tool and very often makes the algorithms and solution techniques an integral part of his way of thinking and interpreting the modeled reality.

For this reason, the “getting started” phase every modeling software user must go through when learning to use a specific package is extremely important because it largely conditions the future perception of (and way of using) the software by the

user. It is believed that the following recommendations will help any modeler when learning to use a new modeling package:

(1) *Always read the manuals.* Very often, commercially available or public domain computational hydraulics/fluid dynamics packages are provided with a detailed documentation that includes a description of the governing equations, solution algorithms and user options available. Reading the documentation thoroughly most often allows the experienced user to identify the possible weaknesses and limitations of the modeling software package.

(2) *Always run the sample test cases provided with the package.* A complete package documentation should include sample test files, with test cases for which analytical solutions are available (dambreak problem for the open channel equations, sudden valve failure for the water hammer equations, etc.) Solving these test cases often provides a fair idea of the degree of accuracy of the solution techniques.

(3) *Assess the sensitivity of the package to the numerical parameters.* Complying with step (2) is not sufficient. Sample test cases are often chosen to demonstrate the ability of the software to reproduce theoretical solutions, hence proving its accuracy. It may happen, however, that modifying slightly some of the numerical parameters (time centering coefficient, computational time step, iteration convergence criteria, etc.) leads to a strong degradation in the quality of the numerical result. If the package is oversensitive to the numerical parameters used, it is always worth knowing.

(4) *Invent your own test cases.* Since the purpose of a software vendor is to sell the software packages, situations where the package performs poorly are usually not documented, and no sample test cases are provided for them. For this reason, it is advised not to restrict the test phase to step (2) and (3). The user should invent his/her own test cases and test the physical plausibility of the solution, even when no analytical solution is available for comparison.

The main prerequisite in modeling is that the modeler be critical about the numerical results provided by the software. Being critical is all the more difficult as user-friendly modeling tools increasingly put an emphasis on a realistic presentation of the simulation results, thus giving the impression that what is being displayed on the screen is reality. Modeling only provides an approximation of reality and numerical techniques are essentially inaccurate. For this reason, the modeler's critical judgment remains an essential feature of the modeling process. If one piece of advice should be given to a newcomer in the world of modeling, it may be the following: the model is and must remain a tool. Use it, never let it use you.

Appendix A

Linear Algebra

A.1. Definitions

A vector v is an ordered set of m numbers v_1, \dots, v_m , called the components of the vector. The components of the vector are arranged in a single column. The following notation is used:

$$v = \begin{bmatrix} v_1 \\ \vdots \\ v_i \\ \vdots \\ v_m \end{bmatrix} = [v_1 \quad \cdots \quad v_i \quad \cdots \quad v_m]^T = [v_i] \quad [\text{A.1}]$$

where v_i is the i th component of v and T is the transposition operator.

An $m \times n$ matrix A is formed by a set of numbers a_{ij} arranged in m rows and n columns. i and j are respectively the indices for the row and the column of the matrix. The following notation is used:

$$A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,j} & \cdots & a_{1,n} \\ \vdots & & \vdots & & \vdots \\ a_{i,1} & \cdots & a_{i,j} & \cdots & a_{i,n} \\ \vdots & & \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,j} & \cdots & a_{m,n} \end{bmatrix} = [a_{i,j}] \quad [\text{A.2}]$$

An $m \times n$ matrix may be viewed as a set of n vectors of size m arranged in a single line. The matrix A in equation [A.3] may be defined as:

$$A = \left[a^{(1)} \dots a^{(j)} \dots a^{(n)} \right] \quad [A.3]$$

where the vectors $a^{(j)}, j = 1, \dots, n$, are defined as:

$$a^{(j)} = [a_{1,j} \dots a_{i,j} \dots a_{m,j}]^T \quad [A.4]$$

Note that a vector is a single-columned matrix.

Also note the following, particular cases:

- A square matrix has the same number of rows and columns, $m = n$.
- A symmetric matrix is a matrix that is left invariant by transposition (a symmetric matrix is necessarily a square matrix):

$$a_{i,j} = a_{j,i} \quad \forall \begin{cases} i = 1, \dots, m \\ j = 1, \dots, m \end{cases} \quad [A.5]$$

- The identity matrix is a symmetric matrix, the elements of which are all zero, except the diagonal terms that are equal to one:

$$\begin{aligned} I &= [\delta_{i,j}] \\ \delta_{i,j} &= \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \end{aligned} \quad [A.6]$$

where $\delta_{i,j}$ is known as Kronecker's operator.

A.2. Operations on matrices and vectors

A.2.1. Addition

Let $A = [a_{i,j}]$ and $B = [b_{i,j}]$ be two $m \times n$ matrices. Adding A and B yields the matrix $C = [c_{i,j}]$ defined as follows:

$$\begin{aligned} C &= A + B \\ c_{i,j} &= a_{i,j} + b_{i,j} \quad \forall (i = 1, \dots, m; j = 1, \dots, n) \end{aligned} \quad [A.7]$$

The sum of two vectors is defined exactly in the same way:

$$\left. \begin{array}{l} \mathbf{w} = \mathbf{u} + \mathbf{v} \\ w_i = u_i + v_i \quad \forall i = 1, \dots, m \end{array} \right\} \quad [\text{A.8}]$$

Note that matrices or vectors may be added only if they have the same size.

A.2.2. Multiplication by a scalar

Let $\mathbf{A} = [a_{i,j}]$ and β be a matrix and a scalar respectively. Multiplying \mathbf{A} by β yields the matrix $\mathbf{B} = [b_{i,j}]$ defined as:

$$\left. \begin{array}{l} \mathbf{B} = \beta \mathbf{A} \\ b_{i,j} = \beta a_{i,j} \quad \forall (i = 1, \dots, m; j = 1, \dots, n) \end{array} \right\} \quad [\text{A.9}]$$

The product of a scalar and a vector is defined in the same way:

$$\left. \begin{array}{l} \mathbf{v} = \beta \mathbf{u} \\ v_i = \beta u_i \quad \forall i = 1, \dots, m \end{array} \right\} \quad [\text{A.10}]$$

A.2.3. Matrix product

Let $\mathbf{A} = [a_{i,j}]$ be an $m \times l$ matrix and $\mathbf{B} = [b_{i,j}]$ be an $l \times n$ matrix. The product of \mathbf{A} and \mathbf{B} is an $m \times n$ matrix \mathbf{C} defined as:

$$\left. \begin{array}{l} \mathbf{C} = \mathbf{AB} \\ c_{i,j} = \sum_{k=1}^l a_{i,k} b_{k,j} \quad \forall (i = 1, \dots, m; j = 1, \dots, n) \end{array} \right\} \quad [\text{A.11}]$$

A vector being nothing else than a matrix with only one column, the product between the matrix \mathbf{A} and the vector \mathbf{u} is defined as:

$$\left. \begin{array}{l} \mathbf{v} = \mathbf{Au} \\ v_i = \sum_{k=1}^n a_{i,k} u_k \quad \forall i = 1, \dots, m \end{array} \right\} \quad [\text{A.12}]$$

NOTE.— In contrast with scalar multiplication, the matrix product is not commutative. The product AB is not equal to the product BA in the general case.

A.2.4. Determinant of a matrix

Let A be a square matrix of size $m \times m$. The determinant of A , denoted by $\text{Det}(A)$, or $|A|$, is defined using the following recurrence relationship:

$$\begin{aligned} |A| &= \sum_{i=1}^m (-1)^{i+q} a_{i,q} |A_{i,q}| & \forall q = 1, \dots, m \\ &= \sum_{j=1}^m (-1)^{j+p} a_{p,j} |A_{p,j}| & \forall p = 1, \dots, m \end{aligned} \quad [A.13]$$

where the matrix $A_{i,q}$ is the $(m-1) \times (m-1)$ square matrix obtained from A by removing the row q and the column i . The final result is the same, regardless of the row q and the column i chosen in the sum [A.13].

The determinant verifies the following properties:

$$\left. \begin{array}{l} |AB| = |BA| = |A||B| \\ |A^T| = |A| \\ |I| = 1 \end{array} \right\} \quad [A.14]$$

A.2.5. Inverse of a matrix

Let A be an $m \times m$ square matrix. The inverse A^{-1} of A is an $m \times m$ matrix defined as:

$$A^{-1}A = AA^{-1} = I \quad [A.15]$$

The first relationship [A.14] indicates that a matrix has an inverse only if its determinant is non-zero. The third relationship implies that the determinant of the inverse of A is the inverse of the determinant of A .

A.3. Differential operations using matrices and vectors

A.3.1. Differentiation

Let $\mathbf{A} = [a_{ij}]$ be an $m \times n$ matrix. \mathbf{A} is differentiated with respect to a given parameter or variable t by differentiating all its components individually:

$$\frac{\partial \mathbf{A}}{\partial t} = \left[\frac{\partial a_{i,j}}{\partial t} \right] \quad [\text{A.16}]$$

This definition also applies to the particular case of a vector that can be seen as a single-columned matrix:

$$\frac{\partial \mathbf{u}}{\partial t} = \left[\frac{\partial u_i}{\partial t} \right] \quad [\text{A.17}]$$

A.3.2. Jacobian matrix

Let $\mathbf{u} = [u_i]$ be a vector of size m and $\mathbf{v} = [v_i]$ be a vector of size n . The Jacobian matrix \mathbf{A} of \mathbf{u} with respect to \mathbf{v} is an $m \times n$ matrix defined as:

$$\left. \begin{array}{l} \mathbf{A} = \frac{\partial \mathbf{u}}{\partial \mathbf{v}} \\ a_{i,j} = \frac{\partial u_i}{\partial v_j} \quad \forall (i = 1, \dots, m; j = 1, \dots, n) \end{array} \right\} \quad [\text{A.18}]$$

A.4. Eigenvalues, eigenvectors

A.4.1. Definitions

The scalar λ is an eigenvalue of the matrix \mathbf{A} if there is a non-zero vector \mathbf{v} , called and eigenvector, such that:

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v} \quad [\text{A.19}]$$

The characteristic polynomial of \mathbf{A} is defined as:

$$P(\lambda) = |\mathbf{A} - \lambda\mathbf{I}| \quad [\text{A.20}]$$

The eigenvalues of A are the roots of the characteristic polynomial:

$$P(\lambda) = 0 \quad [\text{A.21}]$$

The eigenvector v associated with a given eigenvalue λ is obtained by substituting the (known) value of λ into equation [A.19]. A linear algebraic system is obtained. Since at least one of the components of u is non-zero, it can be set to any arbitrary value, e.g. one, that serves as a basis in the computation of the remaining components of v .

A.4.2. Example

Consider the matrix A obtained for the Saint Venant equations (see section 2.5.3.1):

$$A = \begin{bmatrix} 0 & 1 \\ c^2 - u^2 & 2u \end{bmatrix}$$

The eigenvalues of A verify equations [A.20–21]:

$$\begin{vmatrix} -\lambda & 1 \\ c^2 - u^2 & 2u - \lambda \end{vmatrix} = 0 \quad [\text{A.22}]$$

which leads to:

$$-(2u - \lambda)\lambda - (c^2 - u^2) = 0 \quad [\text{A.23}]$$

Equation [A.23] can be rewritten as:

$$(\lambda - u)^2 = c^2 \quad [\text{A.24}]$$

which leads to the following two solutions:

$$\left. \begin{array}{l} \lambda^{(1)} = u - c \\ \lambda^{(2)} = u + c \end{array} \right\} \quad [\text{A.25}]$$

The eigenvector $\mathbf{K}^{(1)}$ associated with the first eigenvalue $\lambda^{(1)}$ verifies:

$$\begin{bmatrix} 0 & 1 \\ c^2 - u^2 & 2u \end{bmatrix} \begin{bmatrix} K_1^{(1)} \\ K_2^{(1)} \end{bmatrix} = (u - c) \begin{bmatrix} K_1^{(1)} \\ K_2^{(1)} \end{bmatrix} \quad [\text{A.26}]$$

that is:

$$\left. \begin{aligned} K_2^{(1)} &= (u - c)K_1^{(1)} \\ (c^2 - u^2)K_1^{(1)} + 2uK_2^{(1)} &= (u - c)K_2^{(1)} \end{aligned} \right\} \quad [\text{A.27}]$$

These two conditions can easily be checked to be equivalent. The first eigenvector is therefore:

$$\mathbf{K}^{(1)} = \begin{bmatrix} K_1^{(1)} \\ (u - c)K_1^{(1)} \end{bmatrix} \quad [\text{A.28}]$$

The vector $\mathbf{K}^{(1)}$ verifies equation [A.19] for any non-zero value of $K_1^{(1)}$. Using the obvious choice $K_1^{(1)} = 1$ leads to:

$$\mathbf{K}^{(1)} = \begin{bmatrix} 1 \\ u - c \end{bmatrix} \quad [\text{A.29}]$$

It is easy to check that the second eigenvector is given by:

$$\mathbf{K}^{(2)} = \begin{bmatrix} 1 \\ u + c \end{bmatrix} \quad [\text{A.30}]$$

Appendix B

Numerical Analysis

B.1. Consistency

B.1.1. Definitions

The notion of consistency is applicable to the discretized version of a differential equation (see Chapters 6 and 7). It is defined in common language as follows.

A discretized equation is consistent with a differential equation if it becomes equivalent to it as the discretization space and time steps tend to zero. The “difference” between the original equation and the discretization is called the truncation error.

B.1.2. Principle of a consistency analysis

The following section explains how to carry out a consistency analysis. The example of the linear advection equation is used. The non-conservation form [1.48] of the linear advection is recalled here:

$$\frac{\partial C}{\partial t} + u \frac{\partial C}{\partial x} = 0$$

Consider the first-order upwind discretization of equation [1.48] (see Chapter 6):

$$\left. \begin{aligned} C_i^{n+1} &= Cr C_{i-1}^n + (1 - Cr) C_i^n \\ Cr &= \frac{u \Delta t}{\Delta x} \end{aligned} \right\} \quad [B.1]$$

For the sake of clarity, let $C_i^n = C$. The consistency of [B.1] to [1.48] is analyzed using a second-order Taylor series expansion:

$$\left. \begin{aligned} C_{i-1}^n &= C - \Delta x \frac{\partial C}{\partial x} + \frac{\Delta x^2}{2} \frac{\partial^2 C}{\partial x^2} + \varepsilon_1(\Delta x^3) \\ C_i^{n+1} &= C + \Delta t \frac{\partial C}{\partial t} + \frac{\Delta t^2}{2} \frac{\partial^2 C}{\partial t^2} + \varepsilon_2(\Delta t^3) \end{aligned} \right\} \quad [B.2]$$

Substituting equations [B.2] into equation [B.1] leads to:

$$\begin{aligned} C + \Delta t \frac{\partial C}{\partial t} + \frac{\Delta t^2}{2} \frac{\partial^2 C}{\partial t^2} + \varepsilon_2(\Delta t^3) &= (1 - Cr)C \\ + \left[C - \Delta x \frac{\partial C}{\partial x} + \frac{\Delta x^2}{2} \frac{\partial^2 C}{\partial x^2} + \varepsilon_1(\Delta x^3) \right] Cr \end{aligned} \quad [B.3]$$

Substituting definition [B.1] of the Courant number into equation [B.3] yields the following equation:

$$\frac{\partial C}{\partial t} + u \frac{\partial C}{\partial x} = -\frac{\Delta t}{2} \frac{\partial^2 C}{\partial t^2} - \varepsilon_2(\Delta t^3) + u \frac{\Delta x}{2} \frac{\partial^2 C}{\partial x^2} + \frac{\varepsilon_1}{\Delta x}(\Delta x^3) \quad [B.4]$$

The truncation error TE is defined as the difference between the discretized equation and the original equation. Comparing equations [1.48] and [B.4] leads to:

$$TE(\Delta x, \Delta t) = -\frac{\Delta t}{2} \frac{\partial^2 C}{\partial t^2} - \varepsilon_2(\Delta t^3) + u \frac{\Delta x}{2} \frac{\partial^2 C}{\partial x^2} + \frac{\varepsilon_1}{\Delta x}(\Delta x^3) \quad [B.5]$$

$TE(\Delta x, \Delta t)$ tends to zero when both Δt and Δx tend to zero. Discretization [B.1] is consistent with the advection equation [1.48].

NOTE.— The truncation error contains powers of Δt and Δx . In contrast with a well-admitted (and incorrect) practice in engineering studies, decreasing only Δx or Δt is not sufficient for the discretization to be accurate. Both the time step and the cell size should be decreased in order to increase the accuracy of the discretization.

B.1.3. Numerical diffusion, numerical dispersion

Numerical diffusion and dispersion are purely numerical phenomena that arise from the discretization process. As seen in the previous section, the truncation error is made of an infinite sum of terms that contain powers of Δt and Δx multiplied by the derivatives of the solution with respect to time and/or space. TE ($\Delta x, \Delta t$) may be expressed in general form as:

$$\text{TE}(\Delta x, \Delta t) = \sum_{p,q} \alpha_{p,q} \Delta t^{\beta_{p,q}} \Delta x^{\chi_{p,q}} \frac{\partial^{(p+q)} U}{\partial t^p \partial x^q} \quad [\text{B.6}]$$

where the indices p and q vary from zero to infinity. Comparing equations [B.5] and [B.6] yields the following expressions for the coefficients $\alpha_{p,q}$ and the exponents $\beta_{p,q}$ and $\chi_{p,q}$:

$$\left. \begin{array}{l} \alpha_{1,0} = \alpha_{0,1} = \alpha_{1,1} = 0 \\ \alpha_{2,0} = -\frac{\Delta t}{2} \\ \alpha_{0,2} = \frac{\Delta x}{2} \\ \beta_{2,0} = \beta_{0,2} = 1 \end{array} \right\} \quad [\text{B.7}]$$

In general, $\beta_{p,q}$ and $\chi_{p,q}$ increase with p and q . The consequence is that the terms that contain higher-order derivatives decrease faster than those containing lower-order derivatives when Δt and Δx decrease. The relative importance of the lower-order terms in the truncation error increases when the cell size and the time step decrease.

Numerical diffusion appears when:

$$\left. \begin{array}{l} \alpha_{1,0} = \alpha_{0,1} = \alpha_{2,0} = \alpha_{1,1} = 0 \\ \alpha_{0,2} \neq 0 \end{array} \right\} \quad [\text{B.8}]$$

Then the lowest-order derivative in the truncation error is a second-order derivative with respect to x . Such a term is classically attached to diffusion, hence the term “numerical diffusion”.

Numerical dispersion arises when the truncation error contains third-order derivatives with respect to space:

$$\left. \begin{array}{l} \alpha_{1,0} = \alpha_{0,1} = 0 \\ \alpha_{2,0} = \alpha_{1,1} = \alpha_{0,2} = 0 \\ \alpha_{3,0} = \alpha_{2,1} = \alpha_{1,2} = 0 \\ \alpha_{0,3} \neq 0 \end{array} \right\} \quad [B.9]$$

Example: the truncation error [B.5] induces numerical diffusion. This can be shown by eliminating the second-order terms with respect to time. To do so, equation [B.4] is differentiated with respect to time and space:

$$\left. \begin{array}{l} \frac{\partial^2 C}{\partial t^2} + u \frac{\partial^2 C}{\partial x \partial t} = -\frac{\Delta t}{2} \frac{\partial^3 C}{\partial t^3} - \varepsilon_4(\Delta t^3) + u \frac{\Delta x}{2} \frac{\partial^3 C}{\partial x^2 \partial t} + \frac{\varepsilon_3}{\Delta x} (\Delta x^3) \\ \frac{\partial^2 C}{\partial x \partial t} + u \frac{\partial^2 C}{\partial x^2} = -\frac{\Delta t}{2} \frac{\partial^3 C}{\partial x \partial t^2} - \varepsilon_6(\Delta t^3) + u \frac{\Delta x}{2} \frac{\partial^3 C}{\partial x^3} + \frac{\varepsilon_5}{\Delta x} (\Delta x^3) \end{array} \right\} \quad [B.10]$$

Eliminating the derivative $\partial^2 C / \partial x \partial t$ leads to a relationship between $\partial^2 C / \partial t^2$ and $\partial^2 C / \partial x^2$:

$$\begin{aligned} \frac{\partial^2 C}{\partial t^2} &= u^2 \frac{\partial^2 C}{\partial x^2} + \left(u \frac{\partial^3 C}{\partial x \partial t^2} - \frac{\partial^3 C}{\partial t^3} \right) \frac{\Delta t}{2} + \left(\frac{\partial^3 C}{\partial x^2 \partial t} - u \frac{\partial^3 C}{\partial x^3} \right) \frac{u \Delta x}{2} \\ &\quad + \varepsilon_7(\Delta t^3) + \varepsilon_8(\Delta x^2) \end{aligned} \quad [B.11]$$

Substituting equation [B.11] into equation [B.5] yields the following expression:

$$TE(\Delta x, \Delta t) = \left(\frac{u \Delta x}{2} - \frac{\Delta t}{2} u^2 \right) \frac{\partial^2 C}{\partial x^2} + [\varepsilon_9(\Delta t) + \varepsilon_{10}(\Delta x)] \Delta t \quad [B.12]$$

where the polynomials ε_9 and ε_{10} contain third- and higher-order terms with respect to time and space. The first term on the right-hand side of the equal sign becomes

predominant over the second term when Δt and Δx tend to zero, leading to the following equivalence:

$$\text{TE}(\Delta x, \Delta t) \underset{\substack{\Delta x \rightarrow 0 \\ \Delta t \rightarrow 0}}{\approx} \left(\frac{u\Delta x}{2} - \frac{\Delta t}{2} u^2 \right) \frac{\partial^2 C}{\partial x^2} \quad [\text{B.13}]$$

This is a diffusion term. The presence of this term in the truncation error indicates that the upwind scheme does not solve the linear advection [1.48] exactly, but an advection equation in the form:

$$\frac{\partial C}{\partial t} + u \frac{\partial C}{\partial x} - D_{\text{num}} \frac{\partial^2 C}{\partial x^2} \approx 0 \quad [\text{B.14}]$$

where the numerical diffusion coefficient D_{num} is given as:

$$D_{\text{num}} = \frac{u}{2} (\Delta x - u\Delta t) = (1 - \text{Cr}) \frac{u\Delta x}{2} \quad [\text{B.15}]$$

NOTE.— As indicated by the \approx sign in equation [B.14], the truncation error contains higher-order terms in the polynomials ε_9 and ε_{10} . Consequently, the right-hand side of the equation is not strictly zero.

B.2. Stability

B.2.1. Definition

The notion of stability applies to the solution of a differential equation. The solution may be analytical or numerical.

A solution is said to be stable over the time-space domain $[x_1, x_2] \times [t_1, t_2]$ if it is bounded over the domain. In other words, there are two values U_{\min} and U_{\max} such that:

$$\exists(U_{\min}, U_{\max}), \quad U_{\min} \leq U(x, t) \leq U_{\max} \quad \forall \begin{cases} x \in [x_1, x_2] \\ t \in [t_1, t_2] \end{cases} \quad [\text{B.16}]$$

For a numerical solution solved over a computational domain with M computational points for N computational time steps, the condition becomes:

$$\exists(U_{\min}, U_{\max}), \quad U_{\min} \leq U_i^n \leq U_{\max} \quad \forall \begin{cases} i = 1, \dots, M \\ n = 1, \dots, N \end{cases} \quad [\text{B.17}]$$

Stability is usually referred to under the implicit assumption that there is no limit to the time interval over which the solution is to be computed. In other words, t_2 is assumed to be infinite.

B.2.2. Principle of a stability analysis

The simplest existing stability analysis technique is the harmonic stability analysis, also known as Von Neumann analysis. This method is applicable to linear equations with constant coefficients. The purpose is to investigate the stability of a solution that is to be computed from a known initial condition. Assume that the governing equation is an m th-order equation in the form:

$$\sum_{p=0}^m a_p \frac{\partial^p U}{\partial t^p} + b_p \frac{\partial^p U}{\partial x^p} = 0 \quad [\text{B.18}]$$

where a_p and b_p are constant coefficients. The harmonic analysis consists of seeking solutions to equation [B.18] in the form of elementary harmonic solutions in the form:

$$U(x, t) = \sum_k u_k \exp(\omega_k t + \sigma_k x) = \sum_k U_k(x, t) \quad [\text{B.19}]$$

where u_k is a constant and the coefficients σ_k and ω_k take the form:

$$\left. \begin{aligned} \sigma_k &= j\sigma_{k,i} \\ \omega_k &= \omega_{k,r} + j\omega_{k,i} \end{aligned} \right\} \quad [\text{B.20}]$$

where j is the pure imaginary number, $j^2 = -1$. The numbers $\sigma_{k,r}$, $\omega_{k,r}$ and $\omega_{k,i}$ are real numbers. Therefore σ_k is a pure imaginary number, while ω_k is a complex number with real and imaginary parts. This is motivated by the following considerations. Substituting equation [B.20] into [B.19] leads to the following expression:

$$\begin{aligned} U_k(x, t) &= [\cos(\sigma_{k,i}x) + j \sin(\sigma_{k,i}x)] \times \\ &\quad [\cos(\omega_{k,r}t) + j \sin(\omega_{k,r}t)] \exp(\omega_{k,r}t) \end{aligned} \quad [\text{B.21}]$$

The elementary function U_k is a sinusoidal function of x (and is therefore periodic in space), the amplitude of which is a sinusoidal function of time multiplied by an exponential. The real part $\omega_{k,r}$ conditions the variation of the amplitude of the

solution in time. If $\omega_{k,r}$ is negative, the amplitude of the solution decreases with time and the solution is stable. If $\omega_{k,r}$ is positive, the amplitude of the solution increases exponentially with time and the solution is unstable. The stability analysis thus amounts to studying the variations of $\omega_{k,r}$, more particularly its sign. Substituting equation [B.19] into equation [B.18] gives:

$$\sum_{p=0}^m a_p \frac{\partial^p \sum_k U_k}{\partial t^p} + b_p \frac{\partial^p \sum_k U_k}{\partial x^p} = 0 \quad [\text{B.22}]$$

Swapping the sums and using the linearity property of the differentiation operator gives:

$$\sum_k \left(\sum_{p=0}^m a_p \frac{\partial^p U_k}{\partial t^p} + b_p \frac{\partial^p U_k}{\partial x^p} \right) = 0 \quad [\text{B.23}]$$

Since the exponential functions U_k form an orthogonal set, no exponential can be expressed as a linear combination of other exponentials. Consequently, equation [B.23] leads to the necessary condition:

$$\sum_{p=0}^m a_p \frac{\partial^p U_k}{\partial t^p} + b_p \frac{\partial^p U_k}{\partial x^p} = 0 \quad \forall k \quad [\text{B.24}]$$

In other words, equation [B.18] is applicable to each of the components U_k individually. Differentiating equation [B.19] with respect to time and space leads to:

$$\left. \begin{aligned} \frac{\partial^p U_k}{\partial t^p} &= \omega^p U_k \\ \frac{\partial^p U_k}{\partial x^p} &= \sigma^p U_k = (j\sigma_{k,i})^p U_k \end{aligned} \right\} \quad [\text{B.25}]$$

Substituting equation [B.25] into equation [B.24] leads to the following condition:

$$\sum_{p=0}^m a_p \omega^p + b_p \sigma^p = 0 \quad [\text{B.26}]$$

where the subscript k has been dropped for the sake of clarity. Solving equation [B.26] for a given value of σ yields R roots $\omega^{(r)}$ ($r = 1, \dots, R$). The solution U_k then takes the form:

$$U_k(x, t) = \sum_{r=1}^R \beta^{(r)} \exp[\omega^{(r)} t + \sigma x] \quad [\text{B.27}]$$

The solution is stable if and only if each of the exponentials in equation [B.27] are stable for all possible values of σ . In other words, the real part of each of the roots $\omega^{(r)}$ must be zero or negative.

B.2.3. Harmonic analysis of analytical solutions

B.2.3.1. The linear advection equation

The harmonic analysis of the advection equation in non-conservation form is carried out hereafter. The non-conservation form [1.48] is recalled:

$$\frac{\partial C}{\partial t} + u \frac{\partial C}{\partial x} = 0 \quad [\text{B.28}]$$

This equation can be written in the form [B.18] by letting:

$$\left. \begin{array}{l} m = 1 \\ a_1 = 1 \\ b_1 = u \end{array} \right\} \quad [\text{B.29}]$$

The solution is sought in the form [B.19]. Differentiating expression [B.19] with respect to time and space gives:

$$\left. \begin{array}{l} \frac{\partial U_k}{\partial t} = \omega U_k \\ \frac{\partial U_k}{\partial x} = j\sigma_i U_k \end{array} \right\} \quad [\text{B.30}]$$

Substituting equation [B.30] into equation [B.28] yields the following equation:

$$\omega = ju\sigma_i \quad [\text{B.31}]$$

The solution takes the form:

$$U_k = \beta \exp[(ut - x)j\sigma_i] \quad [B.32]$$

Two remarks can be made:

– The real part of the exponential is zero. The solution is stable, its amplitude is constant.

– The solution is invariant for $dx = u dt$. This is precisely the invariance property derived in Chapter 1. Moreover, the speed at which the solution travels does not depend on the wavelength.

These two remarks may be synthesized as follows: the solution is transported at the speed u . Neither its shape nor its amplitude are altered.

B.2.3.2. The diffusion equation

The diffusion equation can be written in the form:

$$\frac{\partial U}{\partial t} - D \frac{\partial^2 C}{\partial x^2} = 0 \quad [B.33]$$

where the diffusion coefficient D is a real positive number. Equation [B.33] is rewritten in the form [B.18] with:

$$\left. \begin{array}{l} m = 2 \\ a_1 = 1 \\ a_2 = b_1 = 0 \\ b_2 = -D \end{array} \right\} \quad [B.34]$$

The solution is sought in the form [B.19]. Differentiating equation [B.19] with respect to time and space gives:

$$\left. \begin{array}{l} \frac{\partial U_k}{\partial t} = \omega U_k \\ \frac{\partial^2 U_k}{\partial x^2} = -\sigma_i^2 U_k \end{array} \right\} \quad [B.35]$$

Substituting equation [B.35] into equation [B.33] leads to:

$$\omega = -D\sigma_i^2 \quad [B.36]$$

Hence the expression of the solution:

$$U_k = U \exp(-D\sigma_i^2 t + i\sigma_i x) \quad [B.37]$$

Two remarks can be made:

– The analytical solution of the diffusion equation is stable. This is because D is assumed to be positive. A negative coefficient D would lead to a positive ω , thus yielding an increasing exponential and an unstable solution.

– Large values of σ_i correspond to short wavelengths. The shorter the wavelength, the steeper the exponential. In other words, the amplitude of short waves decreases faster with time than the amplitude of long waves. This explains why steep fronts and sharp gradients are smoothed out faster than long waves and mild profiles.

B.2.3.3. The advection dispersion equation

The advection dispersion equation is a third-order PDE in the form:

$$\frac{\partial U}{\partial t} + u \frac{\partial C}{\partial x} + \Omega \frac{\partial^3 C}{\partial x^3} = 0 \quad [B.38]$$

where the coefficient Ω is called the dispersion coefficient. Equation [B.38] can be written in the form [B.18] by letting:

$$\left. \begin{array}{l} m = 3 \\ a_1 = 1 \\ a_2 = a_3 = b_2 = 0 \\ b_1 = u \\ b_3 = \Omega \end{array} \right\} \quad [B.39]$$

The solution is sought in the form [B.19]. Differentiating equation [B.19] with respect to time and space leads to:

$$\left. \begin{array}{l} \frac{\partial U_k}{\partial t} = \omega U_k \\ \frac{\partial U_k}{\partial x} = j\sigma_i U_k \\ \frac{\partial^3 U_k}{\partial x^3} = -j\sigma_i^3 U_k \end{array} \right\} \quad [B.40]$$

Substituting equations [B.40] into equation [B.38] leads to:

$$\omega = (u - \Omega\sigma_i^2)\sigma_i \quad [\text{B.41}]$$

The solution U_k takes the form:

$$U_k = U \exp \left\{ \left[(u - \Omega\sigma_i^2)t - x \right] j\sigma_i \right\} \quad [\text{B.42}]$$

Note that:

- the real part of the exponential is zero. The solution is stable, its amplitude is constant in time;
- the solution [B.42] verifies the following invariance property:

$$\frac{dU_k}{dt} = 0 \quad \text{for } \frac{dx}{dt} = u - \Omega\sigma_i^2 \quad [\text{B.43}]$$

The elementary solution U_k is invariant along the characteristic of speed $u - \Omega\sigma_i^2$. In contrast with the advection equation, the travelling speed of the solution of the dispersion equation depends on the wavelength. Although the amplitude of the waves does not change, the various waves in the solution are shifted gradually with respect to each other, leading to oscillations in the solution. The oscillations are usually stronger in the neighborhood of steep fronts because shorter waves are characterized by larger values of σ_i , thus leading to stronger shifts.

B.2.4. Harmonic analysis of numerical solutions

This section deals with the harmonic analysis of numerical solutions. Consider the explicit upwind scheme presented in Chapter 6, defined as in equation [B.1], recalled hereafter:

$$\left. \begin{aligned} C_i^{n+1} &= \text{Cr} C_{i-1}^n + (1 - \text{Cr})C_i^n \\ \text{Cr} &= \frac{u\Delta t}{\Delta x} \end{aligned} \right\}$$

The stability of a harmonic component U_k of the numerical solution is analyzed. Noting that C_i^n is the value of the numerical solution at the abscissa $x_i = i \Delta x$ at the time $t^n = n \Delta t$, equation [B.19] leads to the following expression for $(U_k)_i^{n+1}$:

$$\begin{aligned}
 (U_k)_i^{n+1} &= u_k \exp(\omega t^{n+1} + \sigma x) \\
 &= u_k \exp[\omega(t^n + \Delta t) + \sigma x] \\
 &= (U_k)_i^n \exp(\omega \Delta t)
 \end{aligned} \tag{B.44}$$

while the following expression is obtained for $(U_k)_{i-1}^n$:

$$(U_k)_{i-1}^n = (U_k)_i^n \exp(-\sigma \Delta x) \tag{B.45}$$

with $s = j \sigma_i$. Substituting equations [B.44–45] into the numerical scheme [B.1], we obtain:

$$\exp(\omega \Delta t) = Cr \exp(-\sigma \Delta x) + 1 - Cr \tag{B.46}$$

The quantity $\exp(\omega \Delta t)$ is the factor by which the solution is multiplied from one time step to the next. It is referred to as the numerical amplification factor A_N :

$$A_N = \exp(\omega \Delta t) = \frac{(U_k)_i^{n+1}}{(U_k)_i^n} \tag{B.47}$$

The solution is stable if the modulus of the amplification factor is equal to or smaller than unity. If this is the case, the modulus of the solution decreases from one time step to the next and the solution is indeed stable. Conversely, if the modulus of the amplification factor is larger than unity the numerical solution is unstable. The stability analysis amounts to studying the variations in the modulus of A_N with σ . This is done using a graphical representation in the complex plane (Figure B.1).

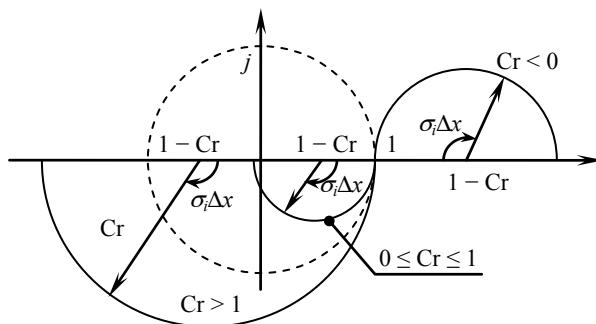


Figure B.1. Definition sketch for the numerical amplification factor in the complex plane

The expression for A_N is rewritten as:

$$A_N = 1 + [\exp(-\sigma_i \Delta x) - 1] Cr \quad [B.48]$$

The wave number M is introduced. M is the number of cells of width Δx needed to cover a period of the signal U_k :

$$M = \frac{2\pi}{\sigma_i \Delta x} \quad [B.49]$$

The minimum possible value for M being 2 (at least two cells are needed to describe a sine wave), the quantity σ_i lies between 0 and π . A_N is represented graphically in the complex plane as the circle of radius $|Cr|$ that is tangent to the unit circle at the point $z = 1$ (Figure B.1). The center of the circle is located at $z = 1 - Cr$. Quite obviously, A_N is located outside the unit circle for $Cr < 0$ and $Cr > 1$. The solution is unstable. A_N is located inside the unit circle when Cr is between 0 and 1. In this case the numerical solution is stable.

B.2.5. Amplitude and phase portraits

Amplitude and phase portraits are graphical representations of the performance of numerical schemes. The amplitude and phase portrait display the modulus of the amplification factor and the numerical wave speed as functions of the wave number M respectively. The numerical wave speed is derived by noting that the solution component U_k is constant if:

$$\omega_{k,i} dt + \sigma_{k,i} dx = 0 \quad [B.50]$$

In other words, U_k is an invariant along the characteristic line:

$$\frac{dx}{dt} = -\frac{\omega_{k,i}}{\sigma_{k,i}} = -\frac{\arg(A_N)}{\Delta t \sigma_{k,i}} = -\frac{M \Delta x \arg(A_N)}{2\pi \Delta t} \quad [B.51]$$

The speed dx/dt is usually referred to as the phase velocity (as opposed to the group velocity, see [VIC 82]). The ratio C_N of the numerical wave speed to the analytical wave speed is given by:

$$C_N = -\frac{M \Delta x \arg(A_N)}{2\pi u \Delta t} = -\frac{M \arg(A_N)}{2\pi Cr} \quad [B.52]$$

The amplification factor and the ratio of the wave speeds for the numerical solution of the advection equation must tend to one when the wave number tends to infinity. This is an indication that the numerical solution tends to behave as the analytical one when an infinity of points (hence and very small time step and cell width) are used to compute the numerical solution.

The amplitude portrait of the explicit upwind scheme [B.1] is obtained from equation [B.48] (see Figure B.2). (see Figure B.2).

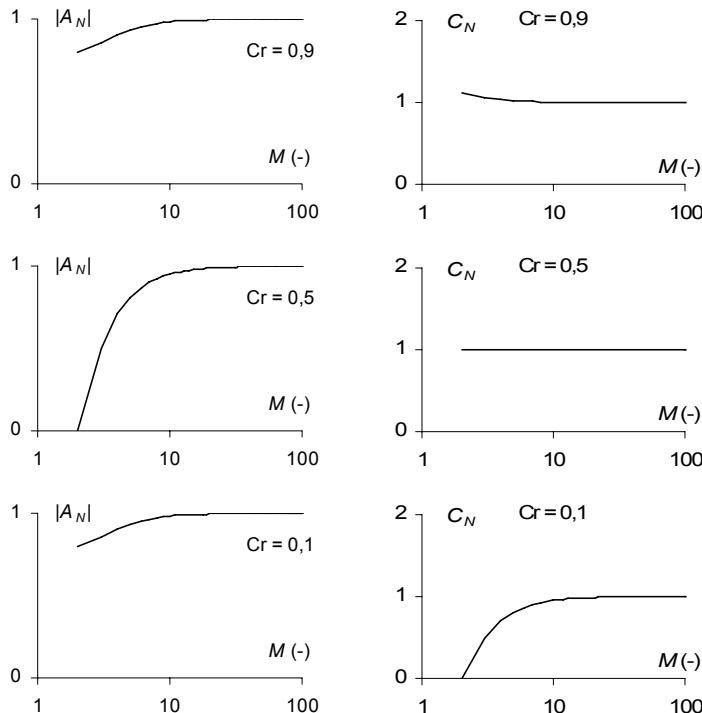


Figure B.2. Amplitude and phase portrait for the explicit upwind scheme

$$\begin{aligned}
 |A_N| &= |1 + [\cos(\sigma\Delta x) - 1 - j \sin(\sigma\Delta x)] Cr| \\
 &= \left\{ [1 - Cr + Cr \cos(\sigma\Delta x)]^2 + [Cr \sin(\sigma\Delta x)]^2 \right\}^{1/2} \\
 &= \left\{ [1 - Cr + Cr \cos(2\pi / M)]^2 + [Cr \sin(2\pi / M)]^2 \right\}^{1/2}
 \end{aligned} \tag{B.53}$$

The phase portrait is obtained from equations [B.48] and [B.52] (see Figure B.2):

$$C_N = -\frac{M \arg(A_N)}{2\pi i} = \frac{M \arccos(\operatorname{Re} A_N / |A_N|)}{2\pi i} \quad [\text{B.54}]$$

Note that:

- short waves are characterized by small wave numbers. The amplification factor being an increasing function of M , short waves are damped more quickly than long waves. After a certain amount of time, the shorter waves are eliminated from the numerical solution, only the longer waves remain. The most unfavorable configuration is encountered for $\operatorname{Cr} = 1/2$. In this case $A_N = 0$ for $M = 2$ and the waves $M = 2$ are eliminated from the numerical solution after the first time step. A first-order expansion in $1/M$ indicates in contrast that the amplification factor tends to one as M tends to infinity, which illustrates the convergence of the numerical solution toward the analytical solution;
- the numerical wave speed is larger than the analytical wave speed for Courant numbers larger than $1/2$. It is smaller than the analytical wave speed for Courant numbers smaller than $1/2$. C_N tends to unity as M tends to infinity, which is another indication that the numerical solution tends to the analytical one.

B.2.6. Extension to systems of equations

Harmonic analysis can be extended to systems of equations. When systems of equations are to be solved the solution is a vector variable and the amplification factor becomes a matrix. The eigenvalues of the matrix are complex in the general case. The solution is stable when the absolute value of each of the eigenvalues of the amplification factor is larger than one.

The stability analysis is carried out for the water hammer equations without source term. The non-conservation form [2.5] of the water hammer equations is recalled hereafter:

$$\frac{\partial \mathbf{U}}{\partial t} + \mathbf{A} \frac{\partial \mathbf{U}}{\partial x} = 0$$

where \mathbf{A} and \mathbf{U} are defined as in equations [2.68–69]:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ c^2 & 0 \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} \rho A \\ \rho Q \end{bmatrix}$$

The solution $U(x, t)$ is sought as the sum of elementary components U_k in the form:

$$U_k(x, t) = u_k \exp(\omega_k t + \sigma_k x) \quad [B.55]$$

where u_k is a constant vector and the coefficients ω_k , σ_k are given as in equation [B.20]:

$$\left. \begin{aligned} \sigma_k &= j\sigma_{k,i} \\ \omega_k &= \omega_{k,r} + j\omega_{k,i} \end{aligned} \right\}$$

Differentiating equation [B.55] with respect to time and space gives:

$$\left. \begin{aligned} \frac{\partial U_k}{\partial t} &= \omega_k U_k \\ \frac{\partial U_k}{\partial x} &= j\sigma_{k,i} U_k \end{aligned} \right\} \quad [B.56]$$

Substituting equations [B.56] into equation [2.5] leads to:

$$\omega_k \begin{bmatrix} \rho A \\ \rho Q \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ c^2 & 0 \end{bmatrix} \sigma_k \begin{bmatrix} \rho A \\ \rho Q \end{bmatrix} = 0 \quad [B.57]$$

Equation [B.57] is rewritten as:

$$\begin{bmatrix} \omega_k & \sigma_k \\ c^2 \sigma_k & \omega_k \end{bmatrix} \begin{bmatrix} \rho A \\ \rho Q \end{bmatrix} = 0 \quad [B.58]$$

Equation [B.58] must hold for all possible values of U . This is true only if:

$$\begin{vmatrix} \omega_k & \sigma_k \\ c^2 \sigma_k & \omega_k \end{vmatrix} = 0 \quad [B.59]$$

that is:

$$\omega_k = \pm c \sigma_k \quad [B.60]$$

The solution U_k is the sum of two exponential functions:

$$U_k(x, t) = u_k^{(1)} \exp[(x - ct)i\sigma_{k,i}] + u_k^{(2)} \exp[(x + ct)i\sigma_{k,i}] \quad [B.61]$$

where $u_k^{(1)}$ and $u_k^{(2)}$ are constant vectors. Equation [B.61] allows the basic properties of the solution of the water hammer to be retrieved:

- The solution is the sum of two signals propagating at the speeds $-c$ and $+c$.
- The arguments of the exponentials are pure imaginary numbers. Therefore the amplitude of the signals is constant. The solution is stable.
- The speed at which each of the signals propagates is independent of the wavelength of the elementary solution. The signals propagate without deformation in the pipe.

B.3. Convergence

B.3.1. *Definition*

The numerical solution of a differential equation is said to be convergent if it tends to the analytical solution as both the computational time step and the cell width tend to zero.

Engineers and modelers implicitly assume that convergence is true when they use software packages to solve the partial differential equations of engineering. The purpose indeed is that the numerical solution be as close to the exact solution as deemed appropriate given the objectives of the engineering project. This is achieved by decreasing the computational time step and the cell width until the numerical solution is considered accurate enough.

B.3.2. *Lax's theorem*

Convergence proofs are difficult to establish. They use notions in functional analysis that go beyond the usual mathematical apparatus accessible to engineers. Lax's theorem for linear equations with constant coefficients allows convergence to be related to consistency and stability. The theorem may be formulated as follows.

Consistency and stability are sufficient and necessary conditions to convergence.

In other words, if the governing equations are discretized in a consistent way and if the numerical solution is stable, then the numerical solution converges to the exact solution.

Appendix C

Approximate Riemann Solvers

C.1. The HLL and HLLC solvers

C.1.1. The HLL solver

C.1.1.1. Principle

The HLL solver is named after the initials of Harten, Lax and Van Leer [HAR 83b]. The HLL solver was developed for 2×2 systems of hyperbolic conservation laws. In most applications of fluid mechanics, the two equations solved are the continuity and the momentum equations.

The HLL solver is based on the *a priori* assumption that the solution is made of two discontinuities moving at speeds λ^- and λ^+ that separate the left and right states of the Riemann problem from an intermediate region of constant state (Figure C.1). The discontinuities do not need to be physically permissible (i.e. the entropy principle does not need to be verified). Assuming that estimates can be provided for the speeds λ^- and λ^+ , the Rankin-Hugoniot conditions can be written across the two waves, thus providing two relationships between the intermediate region of constant state and the states of the Riemann problem.

Writing the jump relationship across the wave λ^- gives:

$$F_L - F_* = (U_L - U_*)\lambda^- \quad [C.1]$$

where U_L and U_R are the values of U and F in the left state of the Riemann problem respectively, and U^* and F^* are the values of U and F in the intermediate region respectively.

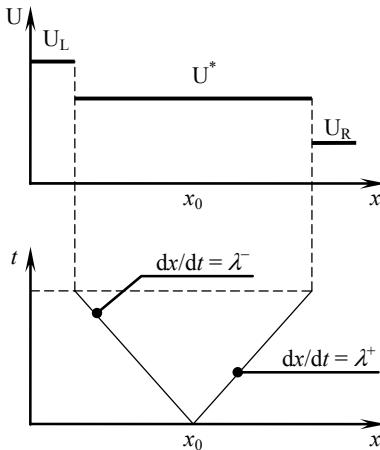


Figure C.1. Principle of the HLL solver. Definition sketch in the physical space (top) and in the phase space (bottom)

Writing the jump relationship across the wave λ^+ leads to:

$$F^* - F_R = (U^* - U_R) \lambda^+ \quad [C.2]$$

where U_R and F_R are the values of U and F in the right state of the Riemann problem. Solving equations [C.1–2] for U^* and F^* yields:

$$\left. \begin{aligned} U^* &= \frac{\lambda^+ U_R - \lambda^- U_L}{\lambda^+ - \lambda^-} + \frac{F_L - F_R}{\lambda^+ - \lambda^-} \\ F^* &= \frac{\lambda^+ F_L - \lambda^- F_R}{\lambda^+ - \lambda^-} - \frac{\lambda^- \lambda^+}{\lambda^+ - \lambda^-} (U_L - U_R) \end{aligned} \right\} \quad [C.3]$$

F^* as given by the second equation [C.3] is the approximate value of the flux in the intermediate region of constant state. For the sake of comparison with other approximate solvers (see sections C.2 and C.3), the second equation [C.3] can be rewritten as:

$$F^* = \frac{F_L + F_R}{2} + \left[\frac{1}{2} \frac{\lambda^- + \lambda^+}{\lambda^+ - \lambda^-} A + \frac{\lambda^- \lambda^+}{\lambda^+ - \lambda^-} I \right] (U_L - U_R) \quad [C.4]$$

where I and \tilde{A} are respectively the identity matrix and Roe's matrix, that is, the matrix such that $F_L - F_R = \tilde{A}(U_L - U_R)$ (see section 6.8 for details). Equation [C.4] indicates that the HLL flux can be seen as the sum of a centered estimate of the flux (that is unconditionally unstable when a first-order explicit method is used, see section 6.5.2) and a diffusive flux that contributes to stabilize the numerical solution. The diffusive flux is often called an artificial viscosity term.

The flux F_0 at the location of the initial discontinuity is equal to F^* if $\lambda^- < 0 < \lambda^+$, that is, for a subsonic/subcritical flow configuration. For supersonic/supercritical situations, F_0 cannot be taken equal to F^* , unless the estimate of the wave speeds λ^- and λ^+ is adapted as described in section C.1.1.2.

C.1.1.2. Wave speed estimates

Davis [DAV 88] proposed the following formula for the estimates of λ^- and λ^+ :

$$\left. \begin{array}{l} \lambda^- = \min[\lambda^{(1)}(U_L), \lambda^{(1)}(U_R), 0] \\ \lambda^+ = \max[\lambda^{(2)}(U_L), \lambda^{(2)}(U_R), 0] \end{array} \right\} \quad [C.5]$$

with $\lambda^{(1)} < \lambda^{(2)}$. Formula [C.4] allows subcritical conditions as well as supercritical conditions to be accounted for in a single formula. Assume indeed a supercritical flow to the left, $\lambda^{(1)} < \lambda^{(2)} < 0$. Equations [C.5] become:

$$\left. \begin{array}{l} \lambda^- = \min[\lambda^{(1)}(U_L), \lambda^{(1)}(U_R)] \\ \lambda^+ = 0 \end{array} \right\} \quad [C.6]$$

and equation [C.3] leads to:

$$F_0 = F^* = F_R \quad [C.7]$$

This is not the correct expression for the flux F^* in the intermediate region of constant state. However, this is the correct formula for the value F_0 of F at the location of the initial discontinuity.

Einfeldt [EIN 88] shows that the following estimates lead to a better resolution of shock waves:

$$\left. \begin{array}{l} \lambda^- = \min[\lambda^{(1)}(U_L), \tilde{\lambda}^{(1)}(U_R, U_R), 0] \\ \lambda^+ = \max[\tilde{\lambda}^{(2)}(U_R, U_R), \lambda^{(2)}(U_R), 0] \end{array} \right\} \quad [C.8]$$

where the wave speeds $\tilde{\lambda}^{(p)}(U_L, U_R)$ are the eigenvalues of Roe's matrix \tilde{A} obtained from the states U_L and U_R (see section 6.8 for details). An alternative option proposed in [EIN 88] consists of using the property, proved in [LAX 57] for convex conservation laws, that the propagation speed for a shock is approximated by the average of the wave speeds on both sides of the shocks:

$$\left. \begin{aligned} \lambda^- &= \min \left[\frac{\lambda^{(1)}(U_L) + \lambda^{(1)}(U_R)}{2}, 0 \right] \\ \lambda^+ &= \max \left[\frac{\lambda^{(2)}(U_L) + \lambda^{(2)}(U_R)}{2}, 0 \right] \end{aligned} \right\} \quad [\text{C.9}]$$

C.1.2. The HLLC solver

The HLLC solver is an extension of the HLL solver to 3×3 hyperbolic systems of conservation laws where the middle wave is a contact discontinuity. Examples of such systems are the Euler equations of gas dynamics seen in section 2.6, or the one-dimensional restriction of the two-dimensional shallow water equations seen in section 5.4. Combining the linear advection equation seen in section 1.2 to hydrodynamic equations such as the water hammer equations (section 2.4) or the Saint Venant equations (section 2.5).

Note that the HLL solver may also be applied to 3×3 systems because equation [C.3] does not use any particular assumption as to the number of components of U and F . However, the contact discontinuity in 3×3 systems moves at the speed u , while the speeds of the other two waves are $u - c$ and $u + c$. The contact discontinuity is located in the intermediate region of constant state. Equation [C.3] that leads us to overestimate the speed of the contact discontinuity induces strong numerical diffusion, thus causing unacceptable smoothing of the contact discontinuity in the numerical solution. The more accurate HLLC solver [TOR 94] was developed to better account for the propagation speed of the contact discontinuity.

The HLLC solver is applied to the one-dimensional restriction of the two-dimensional shallow water equations seen in section 5.4. Remember that the one-dimensional restriction of two-dimensional equations in the direction normal to the interfaces between the computational cells is the key tool to the finite volume solution of two-dimensional hyperbolic conservation laws (see section 7.1.2). For

the sake of simplicity, the source term is assumed to be zero in what follows. Equation [2.2] is to be solved, with the following definitions for U, F and S:

$$U = \begin{bmatrix} h \\ hu \\ hv \end{bmatrix}, \quad F = \begin{bmatrix} hu \\ hu^2 + gh^2 / 2 \\ huv \end{bmatrix}, \quad S = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad [C.10]$$

where g is the gravitational acceleration, h is the water depth, u and v are the flow velocities in the x - and y -directions respectively. The first two components of the vectors account for the continuity and x -momentum equations, that yield the waves $u - c$ and $u + c$. The third component of the vectors U and F accounts for the conservation of the y -momentum. It is easy to check that the following characteristic equation holds for the third wave:

$$\frac{dv}{dt} = 0 \quad \text{for} \quad \frac{dx}{dt} = u \quad [C.11]$$

The y -velocity v is a Riemann invariant along the characteristic of speed u . The x -velocity u being independent from v , the propagation speed of the invariant v is not a function of it and the wave is a contact discontinuity.

The wave pattern of the solution is the following (see Figure C.2). The central wave is a contact discontinuity moving at a speed u . The left and right waves, the speeds of which are $u - c$ and $u + c$ respectively, may be rarefaction waves or shock waves depending on the left and right states of the Riemann problem.

The intermediate region of constant state is divided into two parts. The subregion to the left of the contact discontinuity is denoted by the superscript $*,1$, while the subregion on the right-hand side of the contact discontinuity is denoted by the superscript $*,2$. The first two components F_1 and F_2 of the flux are the same in both subregions. They are computed using equation [C.3] as follows:

$$\left. \begin{aligned} F_1^{*,1} &= F_1^{*,2} = \frac{\lambda^+ F_{1,L} - \lambda^- F_{1,R}}{\lambda^+ - \lambda^-} - \frac{\lambda^- \lambda^+}{\lambda^+ - \lambda^-} (h_L - h_R) \\ F_2^{*,1} &= F_2^{*,2} = \frac{\lambda^+ F_{2,L} - \lambda^- F_{2,R}}{\lambda^+ - \lambda^-} - \frac{\lambda^- \lambda^+}{\lambda^+ - \lambda^-} [(hu)_L - (hu)_R] \end{aligned} \right\} \quad [C.12]$$

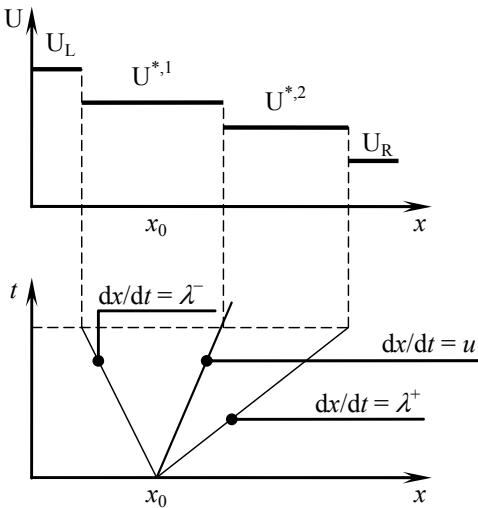


Figure C.2. Principle of the HLLC solver in the physical space (top) and in the phase space (bottom)

where $F_1 = hu$ and $F_2 = hu^2 + gh^2/2$. The values of v on the left- and right-hand sides of the contact discontinuity are v_L and v_R respectively. The expression of v in the two subregions is:

$$\left. \begin{aligned} v^{*,1} &= v_L \\ v^{*,2} &= v_R \end{aligned} \right\} \quad [C.13]$$

which leads to the following expression for the third component of the flux vector:

$$\left. \begin{aligned} F_3^{*,1} &= F_1^{*,1} v_L \\ F_3^{*,2} &= F_1^{*,2} v_R \end{aligned} \right\} \quad [C.14]$$

The flux F_3 at the interface is equal to $F_3^{*,1}$ if $F_1^{*,1}$ is positive. It is equal to $F_3^{*,2}$ if $F_1^{*,1}$ is negative.

C.2. Roe's solver

C.2.1. Principle

Roe's solver [ROE 81] is based on the non-conservation form [2.5] of the governing equations:

$$\frac{\partial \mathbf{U}}{\partial t} + \mathbf{A} \frac{\partial \mathbf{U}}{\partial x} = \mathbf{S}'$$

where the matrix \mathbf{A} is defined so as to enforce conservation, as explained in section 6.8.2. The coefficients of Roe's matrix are constant coefficients, functions of the left and right states of the Riemann problem. Therefore the solution of the Riemann problem with the non-conservation form [2.5] of the equations is made of m contact discontinuities separating intermediate regions of constant state (Figure C.3).

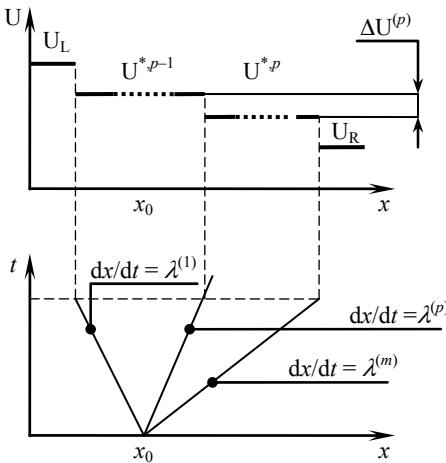


Figure C.3. Principle of Roe's solver. Definition sketch in the physical space (top) and in the phase space (bottom)

The eigenvalues and eigenvectors of \mathbf{A} being known, the difference between the left and right states can be written in the base of eigenvectors of \mathbf{A} as the sum of elementary jumps across the various waves:

$$\mathbf{U}_R - \mathbf{U}_L = \sum_{p=1}^m \Delta \mathbf{U}^{(p)} \quad [\text{C.15}]$$

where the quantity $\Delta U^{(p)}$ is the variation in U across the p th wave. The wave strengths $\alpha^{(p)}$ are introduced as:

$$\Delta U^{(p)} = \alpha^{(p)} K^{(p)} \quad [C.16]$$

where $K^{(p)}$ is the p th eigenvector of A . The p th wave moves at the (known) speed $\lambda^{(p)}$, which allows the jump in U to be related to the jump in F across the wave via the Rankin-Hugoniot condition:

$$\Delta F^{(p)} = \lambda^{(p)} \Delta U^{(p)} \quad [C.17]$$

Substituting equation [C.16] into equation [C.17] yields:

$$\Delta F^{(p)} = \lambda^{(p)} \alpha^{(p)} K^{(p)} \quad [C.18]$$

Applying equation [C.18] across the waves with negative speeds leads to the following relationship between the flux F_L and the flux F_0 at the location x_0 of the initial discontinuity:

$$F_0 - F_L = \sum_{\lambda^{(p)} < 0} \Delta F^{(p)} = \sum_{\lambda^{(p)} < 0} \lambda^{(p)} \alpha^{(p)} K^{(p)} \quad [C.19]$$

Applying equation [C.18] across the waves with positive speeds leads to the following relationship between F_R and F_0 :

$$F_R - F_0 = \sum_{\lambda^{(p)} > 0} \Delta F^{(p)} = \sum_{\lambda^{(p)} > 0} \lambda^{(p)} \alpha^{(p)} K^{(p)} \quad [C.20]$$

Subtracting equation [C.19] from equation [C.20] yields:

$$F_0 = \frac{F_L + F_R}{2} + \frac{1}{2} \sum_{\lambda^{(p)} < 0} \lambda^{(p)} \alpha^{(p)} K^{(p)} - \frac{1}{2} \sum_{\lambda^{(p)} > 0} \lambda^{(p)} \alpha^{(p)} K^{(p)} \quad [C.21]$$

Equation [C.21] can be rewritten in condensed form as:

$$F_0 = \frac{F_L + F_R}{2} - \frac{1}{2} \sum_{p=1}^m |\lambda^{(p)}| \alpha^{(p)} K^{(p)} \quad [C.22]$$

The original algorithm for Roe's solver is the following:

- 1) Compute Roe's matrix, its eigenvalues and its eigenvectors.

2) Compute the wave strengths $\alpha^{(p)}$ as the coordinates of ΔU in the base of eigenvectors of A .

3) Apply equation [C.19].

C.2.2. Algorithmic simplification

The algorithm provided in section C.2.1 requires the computation of the eigenvectors of the matrix \tilde{A} and the subsequent computation of the wave strengths. This can be avoided by noticing that equations [C.15–16] can be rewritten as:

$$\Delta U = K \alpha \quad [C.23]$$

where K and α are respectively the matrix of eigenvectors of \tilde{A} and the vector formed by the wave strengths $\alpha^{(p)}$. Conversely, equation [C.22] can be written as:

$$F_0 = \frac{F_L + F_R}{2} - \frac{1}{2} K |\Lambda| \alpha \quad [C.24]$$

where $|\Lambda|$ is the diagonal matrix formed by the absolute values of the eigenvalues of \tilde{A} . Equation [C.24] is rewritten as:

$$F_0 = \frac{F_L + F_R}{2} - \frac{1}{2} K |\Lambda| K^{-1} K \alpha = \frac{F_L + F_R}{2} - \frac{1}{2} K |\Lambda| K^{-1} \Delta U \quad [C.25]$$

Introducing the matrix $|\tilde{A}| = K |\Lambda| K^{-1}$, equation [C.25] becomes:

$$F_0 = \frac{F_L + F_R}{2} + \frac{1}{2} |\tilde{A}| (U_L - U_R) \quad [C.26]$$

This flux formula is to be compared with equation [C.4] for the HLL solver. As the HLL solver, Roe's flux may be seen as the sum of a centered flux and a diffusive flux (artificial viscosity) that contributes to stabilize the solution.

As mentioned in [EIN 88], Roe's diffusion flux is smaller than the diffusion flux in the HLL solver. This yields problems at sonic/critical points because the numerical diffusion is not strong enough at such points. The treatment of such points is dealt with in the next section.

C.2.3. Entropy violation and fixes

Roe's solver is well-known to yield local violations of the entropy condition in the neighborhood of sonic/critical points. This results in the appearance of “rarefaction shocks” that violate the entropy condition. Various methods have been introduced to eliminate the problem.

1) An entropy fix described in [TOR 97] consists of checking the presence of a sonic/critical point between the left (or right) state of the Riemann problem and the intermediate region of constant state. If a sonic/critical point is detected between the left state U_L and the intermediate region of constant state U_* , the solver is applied to the Riemann problem (U_L, U_*) . The new value of the flux is used as F_0 . Conversely, if a sonic point is detected between U_* and U_R , the solver is applied to the problem (U_*, U_R) . The new value of the flux computed from this Riemann problem is used as U_0 .

2) Another option consists of modifying slightly the calculation of the matrix $|\tilde{A}|$. Indeed the entropy violation at sonic/critical points is due to the fact that one of the eigenvalues $\tilde{\lambda}^{(p)}$ of Roe's matrix is close to zero. Consequently, the stabilizing, diffusive flux is not large enough, which triggers a local increase in the gradient of U . A simple solution consists of imposing that the absolute values of the eigenvalues of $|\Lambda|$ should not be smaller than a given threshold ε :

$$|\Lambda| = \begin{bmatrix} \ddots & & 0 \\ & \max\left(\left|\tilde{\lambda}^{(p)}\right|, \varepsilon\right) & \\ 0 & & \ddots \end{bmatrix} \quad [C.27]$$

C.2.4. Application example: the shallow water equations

Roe's solver is applied to the one-dimensional shallow water equations. These equations can be written in the form [2.2], with U , F , S and K defined as:

$$U = \begin{bmatrix} h \\ hu \end{bmatrix}, \quad F = \begin{bmatrix} hu \\ hu^2 + gh^2/2 \end{bmatrix}, \quad S = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad K = \begin{bmatrix} 1 & 1 \\ u-c & u+c \end{bmatrix} \quad [C.28]$$

The matrix $|\Lambda|$ is given by:

$$|\Lambda| = \begin{bmatrix} |u-c| & 0 \\ 0 & |u+c| \end{bmatrix} \quad [C.29]$$

which leads to the following expression for $|\tilde{A}|$:

$$K|\Lambda|K^{-1} = \frac{1}{2} \begin{bmatrix} (u+c)|Fr - 1| & |Fr + 1| - |Fr - 1| \\ + (c-u)|Fr + 1| & \\ (c^2 - u^2)|Fr + 1| & (c-u)|Fr - 1| \\ - (c^2 - u^2)|Fr - 1| & + (u+c)|Fr + 1| \end{bmatrix} \quad [C.30]$$

If the entropy fix 2) described in section C.2.3 is to be used, $|Fr \pm 1|$ must be replaced with $\max(|Fr \pm 1|, \epsilon)$ in equation [C.30].

C.3. The Lax-Friedrichs solver

The Lax-Friedrichs solver [LAX 54] is used in a number of recently proposed Discontinuous Galerkin (DG) techniques (see Chapter 8 and more specifically section 8.4). It combines solution robustness with the simplicity of the formulation. The flux F_0 at the location x_0 of the initial discontinuity is given by:

$$F_0 = \frac{F_L + F_R}{2} + \frac{1}{2} \frac{\Delta x}{\Delta t} (U_L - U_R) \quad [C.31]$$

This formula is equal to the sum of a centered flux (that leads to unconditionally unstable solutions when used with first-order explicit schemes) and a diffusive flux, or artificial viscosity term. The diffusion coefficient $\Delta x/(2 \Delta t)$ is the maximum possible value that preserves solution stability.

In [COC 98], a so-called “local” Lax-Friedrichs method, also known as the Rusanov flux, is used:

$$\left. \begin{aligned} F_0 &= \frac{F_L + F_R}{2} + \frac{\lambda_{\max}}{2} (U_L - U_R) \\ \lambda_{\max} &= \max_p \left(|\lambda_L^{(p)}|, |\lambda_R^{(p)}| \right) \end{aligned} \right\} \quad [C.32]$$

C.4. Approximate-state solvers

C.4.1. Principle

Approximate-state solvers use a property of hyperbolic systems of conservation laws so that the Riemann invariants (or generalized Riemann invariants) provide an approximation of the Rankin-Hugoniot conditions across shocks [LAX 57]. This means that the approximate solution should be a satisfactory approximation of the exact solution of the Riemann problem even though the nature of the waves present in the Riemann solution has been wrongly guessed.

This property is used by approximate-state solvers to make an *a priori* guess on the nature of the waves in the solution, which influences directly the nature of the system of equations to be solved. The assumption on the nature of the waves is not revised, even if the calculation of the variables in the intermediate zone of constant state shows that the initial guess was wrong. Although formally invalid, the method is shown to give good results when applied to a number of hyperbolic systems. There are two main families of approximate-state Riemann solvers:

- shock wave-based solvers, that use the assumption that the solution is made of shocks across which the Rankin-Hugoniot conditions are applied;
- rarefaction wave-based solvers, where the solution is assumed to be made of rarefaction waves across which the Riemann invariants or generalized Riemann invariants are applied.

C.4.2. Shock-based solvers

The earliest type of approximate state Riemann solvers are two-shock solvers applied to the Euler equations [COL 82]. Assuming that the left and right states are separated from the intermediate region of constant state by two shocks, applying the Rankin-Hugoniot relationships across the shocks gives:

$$\left. \begin{aligned} F_L - F^* &= (U_L - U^*) c_{s,L} \\ F_R - F^* &= (U_R - U^*) c_{s,R} \end{aligned} \right\} \quad [C.33]$$

where $c_{s,L}$ and $c_{s,R}$ are respectively the propagation speeds of the leftward and rightward shocks. In contrast with the HLL/HLLC solver or Roe's solver, these propagation speeds are not known *a priori*. For a 2×2 system of equations, system [C.33] leads to 4 equations, with the two components of U^* and the shock speeds $c_{s,L}$ and $c_{s,R}$ as unknowns.

System [C.33] is nonlinear. Its solution usually requires an iterative procedure. Dukowicz [DUK 85] introduced a simplification for a number of hyperbolic systems of conservation laws (including the Euler equations) whereby the jump relationships are approximated with second-degree polynomials, for which an analytical solution is available.

C.4.3. Rarefaction wave-based solvers

C.4.3.1. Principle

Two rarefaction wave-based Riemann solvers have been applied successfully to various types of hyperbolic systems of conservation laws such as the water hammer equations and 2×2 simplified systems describing two-phase flows in pipes [GUI 00, GUI 01a, GUI 01b], as well as the two-dimensional shallow water equations [GUI 03a, LHO 07] and the two-dimensional shallow water equations with porosity [FIN 10] for the large scale simulation of urban floods. These solvers use the assumption that the solution is made of rarefaction waves. Different options have been explored in the literature:

1) The Riemann invariants applied across the waves are written as functions of the conserved variable U . As shown in section 2.2.1, using the Riemann invariants allows the solution U to be determined uniquely in the intermediate region(s) of constant state. Knowing the solution U at all points allows the flux F_0 at the abscissa x_0 of the initial discontinuity to be computed. This approach can be used for pipe transients in pipes with uniform geometry and in the computation of the shallow water equations on flat bottoms [GUI 00, GUI 01a, GUI 01b, GUI 03a]. The main drawback of this method is that balancing of the source terms in the momentum equations is not straightforward.

2) A second option consists of expressing the Riemann invariants as functions of the flux F . The source terms are included in the characteristic form of the equations, which allows their influence to be taken into account in the determination of the flux F_0 . The broad lines of the approach is given hereafter. The method is presented briefly hereafter. More details on the practical application of the solver to the water hammer, shallow water equations and shallow water equations with porosity can be found in [LHO 07, FIN 10].

C.4.3.2. Computation of the intermediate state

The conservation form [2.2] is recalled:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} = S$$

The purpose is to transform equation [2.2] into an equation in characteristic form where the Riemann invariants are expressed as functions of the flux F. This is done by multiplying equation [2.2] with the Jacobian matrix $A = \partial F / \partial U$:

$$A \frac{\partial U}{\partial t} + A \frac{\partial F}{\partial x} = AS \quad [C.34]$$

Noting that $A dU = dF$, equation [C.34] is rewritten as:

$$\frac{\partial F}{\partial t} + A \frac{\partial F}{\partial x} = AS \quad [C.35]$$

The Riemann invariants appear by left-multiplying equation [C.35] by the inverse of the matrix K (see section 2.1.3):

$$\frac{\partial W}{\partial t} + \Lambda \frac{\partial W}{\partial x} = S' \quad [C.36]$$

where Λ is the diagonal matrix formed by the eigenvalues of A, and W and S' are defined as:

$$\left. \begin{aligned} dW &= K^{-1} dF \\ S' &= K^{-1} AS \end{aligned} \right\} \quad [C.37]$$

Equation [C.36] can be rewritten as:

$$\frac{dW_p}{dt} = S'_p \quad \text{for } \frac{dx}{dt} = \lambda^{(p)}, \quad p = 1, \dots, m \quad [C.38]$$

Solving the $m \times m$ system [C.38] for the components F_p yields directly the expression of F at the abscissa x_0 of the initial discontinuity.

C.4.3.3. Wave pattern update and flux calculation

Knowing the flow variables in the intermediate region of constant state allows the wave pattern to be updated [LHO 07]. This is done in two steps:

(1) The speeds λ^- and λ^+ of the waves are computed for the left, intermediate and right state.

(2) The wave pattern is identified as follows:

(2.1) if $\lambda_L^- < 0 < \lambda_*^-$, the point x_0 is a critical/sonic point in a rarefaction wave.

A critical point formula is used for the flux;

(2.2) if $\lambda_L^- \leq \lambda_*^- \leq 0 \leq \lambda_*^+$, the point x_0 is located within the intermediate region of constant state computed as in section C.4.3.2;

(2.3) if $\lambda_*^+ < 0 < \lambda_R^+$, the point x_0 is a critical/sonic point in a rarefaction wave.

A critical point formula is used for the flux;

(2.4) if $\lambda_L^- > \lambda_*^-$, the left-hand wave is a shock. The shock speed c_s is computed using the jump relationships between the left and intermediate states.
 $F_0 = F_L$ if $c_s > 0$, $F_0 = F_*$ otherwise;

(2.5) if $\lambda_*^+ > \lambda_R^+$, the right-hand wave is a shock. The shock speed c_s is computed using the jump relationships between the intermediate and right states.
 $F_0 = F_R$ if $c_s < 0$, $F_0 = F_*$ otherwise.

Practical applications to the water hammer and shallow water equations indicate that the flux formulae given by such solvers bear similarities with those given by the HLL solver, with enhanced balancing properties (see Chapter 9 for balancing issues of schemes in the presence of source terms).

Appendix D

Summary of the Formulae

This appendix gives an overview of the conservation, non-conservation and characteristic forms of the various conservation laws presented in this book.

Form	Equation	Equation number, section
Conservation form	$\frac{\partial}{\partial t}(AC) + \frac{\partial}{\partial x}(AuC) = 0$	[1.39], 1.3.1
Non-conservation form	$\frac{\partial C}{\partial t} + u \frac{\partial C}{\partial x} = 0$	[1.48], 1.3.2
Characteristic form	$\frac{dC}{dt} = 0 \quad \text{for } \frac{dx}{dt} = u, \quad C = \text{Const} \quad \text{for } \frac{dx}{dt} = u$	[1.49] and [1.50], 1.3.2

Table D.1. *The various forms of the linear advection equation*

Form	Equation	Equation number, section
Conservation form	$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} \left(\frac{u^2}{2} \right) = 0$	[1.69], 1.4.2
Non-conservation form	$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = 0$	[1.66], 1.4.2
Characteristic form	$\frac{du}{dt} = 0 \quad \text{for } \frac{dx}{dt} = u, \quad u = \text{Const} \quad \text{for } \frac{dx}{dt} = u$	[1.67] and [1.68], 1.4.2

Table D.2. The various forms of the inviscid Burgers equation

Form	Equation	Equation number, section
Conservation form	$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0, \quad Q = K_{\text{Str}} \frac{A^{5/3}}{\chi^{2/3}} S_0^{1/2}$	[1.83-84], 1.5.1
Non-conservation form	$\frac{\partial A}{\partial t} + \lambda \frac{\partial A}{\partial x} = S', \quad \frac{\partial Q}{\partial t} + \lambda \frac{\partial Q}{\partial x} = \lambda S',$ $\lambda = \frac{\partial Q}{\partial h} \left(\frac{\partial A}{\partial h} \right)^{-1}$	[1.87] and [1.91], 1.5.2 [1.93], 1.5.3
Characteristic form	$\frac{dA}{dt} = S' \quad \text{for } \frac{dx}{dt} = \lambda, \quad \frac{dQ}{dt} = \lambda S' \quad \text{for } \frac{dx}{dt} = \lambda$	[1.88] and [1.92], 1.5.2

Table D.3. The various forms of the kinematic wave equation

Form	Equation	Equation number, section
Conservation form	$\frac{\partial s}{\partial t} + \frac{\partial F}{\partial x} = 0, F = \frac{s^2}{s^2 + (1-s)^2 b_{BL}} V_d$	[1.111] and [1.115], 1.6.1
Non-conservation form	$\frac{\partial s}{\partial t} + \lambda \frac{\partial s}{\partial x} = 0,$ $\lambda = 2 \frac{(1-s)s}{[s^2 + (1-s)^2 b_{BL}]^2} b_{BL} V_d$	[1.116] and [1.117], 1.6.2
Characteristic form	$\frac{ds}{dt} = 0 \quad \text{for } \frac{dx}{dt} = \lambda, s = \text{Const} \quad \text{for } \frac{dx}{dt} = \lambda$	[1.118] and [1.119], 1.6.2

Table D.4. The various forms of the Buckley-Leverett equation

Form	Equation	Equation number, section
Conservation form	$\frac{\partial M}{\partial t} + \frac{\partial}{\partial x} \left(\frac{V_d}{\theta R_F} M \right) = 0, R_F = 1 + \frac{\rho_A C_A}{\theta C_T}$	[1.131] and [1.132], 1.7.1
Non-conservation form	$\frac{\partial M}{\partial t} + \lambda \frac{\partial M}{\partial x} = 0,$ $\frac{\partial C_T}{\partial t} + \lambda \frac{\partial C_T}{\partial x} = 0 \quad \lambda = \left(\frac{1}{R_F} - \frac{M}{R_F^2} \frac{dR_F}{dM} \right) V_d$	[1.136], [1.137] and [1.139], 1.7.2
Characteristic form	$\frac{dM}{dt} = \frac{dC_T}{dt} = 0 \quad \text{for } \frac{dx}{dt} = \lambda$	[1.140], 1.7.2

Table D.5. The various forms of the advection equation with adsorption-desorption

Form	Equation	Equation number, section
Conservation form	$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} = \mathbf{S},$ $\mathbf{U} = \begin{bmatrix} \rho A \\ \rho Q \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} \rho Q \\ Ap \end{bmatrix},$ $\mathbf{S} = \begin{bmatrix} 0 \\ p \frac{\partial A}{\partial x} - \rho g A \sin \theta - k u u \end{bmatrix}$	[2.2] and [2.68], 2.4.2
Non-conservation form	$\frac{\partial \mathbf{U}}{\partial t} + \mathbf{A} \frac{\partial \mathbf{U}}{\partial x} = \mathbf{S}', \quad \mathbf{S}' = \mathbf{S}$ $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ c^2 & 0 \end{bmatrix}$	[2.5] and [2.69], 2.4.3
Characteristic form	$\frac{dp}{dt} - \frac{\rho c}{A} \frac{dQ}{dt} = (k u u + \rho g A \sin \theta) \frac{c}{A} \quad \text{for } \frac{dx}{dt} = -c$ $\frac{dp}{dt} + \frac{\rho c}{A} \frac{dQ}{dt} = (-k u u - \rho g A \sin \theta) \frac{c}{A} \quad \text{for } \frac{dx}{dt} = c$	[2.79], 2.4.3

Table D.6. The various forms of the water hammer equations

Form	Equation	Equation number, section
Conservation form	$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} = \mathbf{S},$ $\mathbf{U} = \begin{bmatrix} A \\ Au \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} Au \\ \frac{Q^2}{A} + \frac{P}{\rho} \end{bmatrix},$ $\mathbf{S} = \begin{bmatrix} 0 \\ (S_0 - S_f)gA + I_p \end{bmatrix}$	[2.2] and [2.118], 2.5.2.4
Non-conservation form	$\frac{\partial \mathbf{U}}{\partial t} + \mathbf{A} \frac{\partial \mathbf{U}}{\partial x} = \mathbf{S}', \quad \mathbf{S}' = \mathbf{S}$ $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ c^2 - u^2 & 2u \end{bmatrix}$	[2.5] and [2.119], 2.5.3.1
Characteristic form	$\frac{du}{dt} - \frac{c}{A} \frac{dA}{dt} = (S_0 - S_f)g \quad \text{for } \frac{dx}{dt} = u - c$ $\frac{du}{dt} + \frac{c}{A} \frac{dA}{dt} = (S_0 - S_f)g \quad \text{for } \frac{dx}{dt} = u + c$	[2.140], 2.5.3.3

Table D.7. The various forms of the Saint Venant equations

Form	Equation	Equation number, section
Conservation form	$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} = S,$ $U = \begin{bmatrix} \rho \\ \rho u \\ E \end{bmatrix}, \quad F = \begin{bmatrix} \rho u \\ \rho u^2 + p \\ (E + p)u \end{bmatrix}$	[2.2] and [2.197], 2.6.2.5
Non-conservation form	$\frac{\partial U}{\partial t} + A \frac{\partial U}{\partial x} = 0$ $A = \begin{bmatrix} 0 & 1 & 0 \\ c^2 - u^2 & 2u & 0 \\ uc^2 - (E + p)u / \rho & (E + p) / \rho & u \end{bmatrix}$	[2.198] and [2.200], 2.6.3
Characteristic form	$\frac{dp}{dt} - \rho c \frac{du}{dt} = 0 \quad \text{for } \frac{dx}{dt} = u - c$ $\frac{ds}{dt} = 0 \quad \text{for } \frac{dx}{dt} = u$ $\frac{dp}{dt} + \rho c \frac{du}{dt} = 0 \quad \text{for } \frac{dx}{dt} = u + c$ $\frac{d}{dt}(u - \beta_1 p^{\beta_2}) = 0 \quad \text{for } \frac{dx}{dt} = u - c$ $\frac{ds}{dt} = 0 \quad \text{for } \frac{dx}{dt} = u - c$ $\frac{d}{dt}(u + \beta_1 p^{\beta_2}) = 0 \quad \text{for } \frac{dx}{dt} = u + c$ $\beta_1 = \frac{2\gamma}{\gamma+1} \left(\frac{\gamma p_0}{p_0^{1/\gamma}} \right)^{1/2}$ $\beta_2 = \frac{3\gamma+1}{2\gamma}$	[2.209], [2.217], [2.215], 2.6.3

Table D.8. The various forms of the Euler equations

Bibliography

- [AIZ 02] AIZINGER V., DAWSON C., “A discontinuous Galerkin method for two-dimensional flow and transport in shallow water”, *Advances in Water Resources*, 25, pp. 67–84, 2002.
- [AIZ 03] AIZINGER V., DAWSON C., Erratum to “A discontinuous Galerkin method for two-dimensional flow and transport in shallow water”, *Advances in Water Resources*, 26, pp. 349, 2003.
- [ALC 01] ALCRUDO F., BENKHALDOUN F., “Exact solutions to the Riemann problem with a bottom step”, *Computers and Fluids*, 30, pp. 643–671, 2001.
- [ALC 92] ALCRUDO F., GARCIA-NAVARRO P., SAVIRON J.M., “Flux difference splitting for the 1D open channel flow equations”, *International Journal for Numerical Methods in Fluids*, 14, pp. 1009–1018, 1992.
- [ALL 03] ALLIEVI L., *Teoria generale del moto perturbato dell'acqua nei turbi in pressione*. Am. Soc. Ing. Arch. Italiana, 1903. Translated to French by Allievi (1904). Translated into English by E.E. Halmos (1904): *The Theory of Water Hammer*, ASME, NY.
- [AUD 05] AUDUSSE E., BRISTEAU M.O., “A well-balanced positivity preserving “second-order” scheme for shallow water flows on unstructured meshes”, *Journal of Computational Physics*, 206, pp. 311–333, 2005.
- [BAR 02] BARDOS C., PIRONNEAU O., “A formalism for the differentiation of conservation laws”, *Comptes Rendus de l'Académie des Sciences Paris*, Ser. I 335, pp. 836–845, 2002.
- [BER 94] BERMÚDEZ A., VÁZQUEZ M.E., “Upwind methods for hyperbolic conservation laws with source terms”, *Computers and Fluids*, 23, pp. 1049–1071, 1994.
- [BER 98] BERMÚDEZ A., DERVIEUX A., DESIDERI J.A., VÁZQUEZ M.E., “Upwind schemes for the two-dimensional shallow water equations with variable depth using unstructured meshes”, *Computational Methods in Applied Mechanical Engineering*, 155, pp. 49–72, 1998.
- [BRU 02] BRUFAU P., VÁZQUEZ-CENDÓN M.E., GARCÍA-NAVARRO P., “A numerical method for the flooding and drying of irregular domains”, *International Journal for Numerical Methods in Fluids*, 39, pp. 247–275, 2002.

- [BUC 42] BUCKLEY S.E., LEVERETT M.C., "Mechanisms of fluid displacement in sands", *Transactions AIME*, 146, 107-116, 1942.
- [BUR 01] BURGUETE J., GARCÍA-NAVARRO P., "Efficient construction of high-resolution TVD conservative schemes for equations with source terms: application to shallow water flows", *International Journal for Numerical Methods in Fluids*, 37, pp. 209–248, 2001.
- [BUR 04] BURGUETE J., GARCÍA-NAVARRO P., "Improving simple explicit methods for unsteady open channel and river flow", *International Journal for Numerical Methods in Fluids*, 45, pp. 125–156, 2004.
- [BUR 48] BURGERS J.M. "A mathematical model illustrating the theory of turbulence", *Advances in Applied Mechanics*, 1, pp. 171-199, 1948.
- [CAC 03] CACUCI D.G., *Sensitivity and Uncertainty Analysis. I. Theory*, Chapman & Hall/crc, 2003.
- [CAR 72] CARLIER M., *Hydraulique Générale et Appliquée*, Eyrolles, 1972.
- [CHA 03] CHACÓN REBOLLO T., DOMÍNGUEZ DELGADO A., FERNÁNDEZ NIETO E.D., "A family of stable numerical solvers for the shallow water equations with source terms", *Computer Methods in Applied Mechanical Engineering*, 192, pp. 203–225, 2003.
- [CHO 73] CHOW V.T., *Open-Channel Hydraulics*, McGraw-Hill International Editions, Civil Engineering Series, 1973.
- [COC 01] COCKBURN B., SHU C.W., "Runge-Kutta discontinuous Galerkin methods for convection-dominated problems", *Journal of Scientific Computing*, 16, pp. 173–261, 2001.
- [COC 89a] COCKBURN B., LIN S.Y., SHU C.W., "TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws III: One-dimensional systems", *Journal of Computational Physics*, 84, pp. 90–113, 1989.
- [COC 89b] COCKBURN B., SHU C.W., "TVB Runge-Kutta local projection discontinuous Galerkin finite element method for scalar conservation laws II: General framework", *Mathematics of Computation*, 52, pp. 411–435, 1989.
- [COC 90] COCKBURN B., HOU S., SHU C.W., "TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws IV: The multidimensional case", *Math. Comp.*, 54, pp. 545–581, 1990.
- [COC 98] COCKBURN B., SHU C.W., "The Runge-Kutta discontinuous Galerkin method for conservation laws V: Multidimensional systems", *Journal of Computational Physics*, 141, pp. 199–224, 1998.
- [COL 51] COLE J.D., "On a quasilinear equation occurring in aerodynamics", *Quarterly Applied Mathematics*, 9, pp. 225-236, 1951.
- [COL 82] COLELLA P., "Glimm's method for gas dynamics", *SIAM Journal of Scientific and Statistical Computing*, 3, pp. 76–110, 1982.

- [COL 84] COLELLA P., WOODWARD P.R., “The Piecewise Parabolic Method (PPM) for gas-dynamical simulations”, *Journal of Computational Physics*, 54, pp. 174–201, 1984.
- [COU 48] COURANT R., FRIEDRICH K.O., *Supersonic Flow and Shock Waves*, Interscience, NY, London, 1948.
- [COU 52] COURANT R., ISAACSON E., REES M., “On the solution of nonlinear hyperbolic differential equations by finite differences”, *Communications on Pure and Applied Mathematics*, 5, pp. 243–255, 1952.
- [CUN 80] CUNGE J.A., HOLLY F.M. JR, VERWEY A., *Practical Aspects of Computational River Hydraulics*, Pitman Publ., reprinted by Iowa University Press, 1980.
- [DAU 67] DAUBERT A., GRAFFE O., “Quelques aspects des écoulements presque horizontaux à deux dimensions en plan et non permanents”, *La Houille Blanche*, 8, pp. 847–860, 1967.
- [DAV 88] DAVIS S.F., “Simplified second-order Godunov-type methods”, *SIAM Journal of Statistical Science and Computing*, 9, pp. 445–473, 1988.
- [DAW 02] DAWSON C., PROFT J., “Discontinuous and coupled continuous/discontinuous Galerkin methods for the shallow water equations”, *Computer Methods in Applied Mechanics and Engineering*, 191, pp. 4721–4746, 2002.
- [DEL 08] DELENNE C., GUINOT V., CAPPELAERE B., “Direct sensitivity computation for the Saint-Venant equations with hydraulic jumps”, *Comptes Rendus Mécanique* 336, pp. 766–771, 2008.
- [DHI 05] DHI, Mike 11 reference manual. DHI Water and environment, Hørsholm, 2005.
- [DIC 85] DICK E., “A multigrid method for the Cauchy-Riemann equations based on flux-difference splitting and its extension to the steady Euler equations”, *Journal of Computational and Applied Mathematics*, 12&13, pp. 247–263, 1985.
- [DUK 85] DUKOWICZ J., “A general, non-iterative Riemann solver for Godunov’s method”, *Journal of Computational Physics*, 61, pp. 119–137, 1985.
- [EIN 88] EINFELDT B., “On Godunov-type methods for gas dynamics”, *SIAM Journal of Numerical Analysis*, 25(2), pp. 294–318, 1988.
- [ELI 07] ELIZONDO D., CAPPELAERE B., FAURE C., “Automatic versus manual model differentiation to compute sensitivities and solve non-linear inverse problems”, *Computers and Geosciences*, 28, pp. 309–326, 2002.
- [FIN 10] FINAUD-GUYOT P., DELENNE C., LHOMME J., GUINOT V., LLOVEL C., “An approximate-state Riemann solver for the two-dimensional shallow water equations with porosity”, *International Journal for Numerical Methods in Fluids*, 2010.
- [FRO 68] FROMM J.E., “A method for reducing dispersion in convective difference schemes”, *Journal of Computational Physics*, 3, pp. 176–189, 1968.
- [GAL 03] GALLOUËT T., HÉRARD J.M., SEGUIN N., “Some approximate Godunov schemes to compute shalllow-water equations with topography”, *Computers and Fluids*, 32, pp. 479–513, 2003.

- [GAR 00] GARCIA-NAVARRO P., VAZQUEZ-CENDON M.E., “On numerical treatment of the source term in the shallow water equations”, *Computers and Fluids*, 29, pp. 951–979, 2000.
- [GER 82] GERRITSEN H., Accurate boundary treatment in shallow water flow computations, PhD thesis, University of Twente, The Netherlands, 1982.
- [GHO 09] GHOSTINE R., KESSERWANI G., MOSÉ R., VAZQUEZ J., GHENAIM A., “An improvement of classical slope limiters for high-order discontinuous Galerkin method”, *International Journal for Numerical Methods in Fluids*, 59, pp. 423–442, 2009.
- [GOD 59] GODUNOV S.K., “A difference method for calculation of discontinuous solutions of hydrodynamics”, *Matematicheski Sbornik*, 1959.
- [GOD 99] GODUNOV S.K., “Reminiscences about difference schemes”, *Journal of Computational Physics*, 153, pp. 6-25, 1999.
- [GOU 77] GOURLAY S., “Splitting methods for time dependent partial differential equations”, *The state of the art in numerical analysis, Proceedings of the conference: The State of the art in Numerical Analysis*, (ed.) D.A.H. JACOBS, New York University, 12-15 April 1976, Academic Press, 1977.
- [GRE 96] GREENBERG J.M., LEROUX A.Y., “A well-balanced scheme for the numerical processing of source terms in hyperbolic equations”, *SIAM Journal on Numerical Analysis*, 33, pp. 1–16, 1996.
- [GUI 00] GUINOT V., “Riemann solvers for water hammer simulations by Godunov method”, *International Journal for Numerical Methods in Engineering*, 49, pp. 851–870, 2000.
- [GUI 01a] GUINOT V., “Numerical simulation of two-phase flow in pipes using Godunov method”, *International Journal for Numerical Methods in Engineering*, 50, pp. 1169–1189, 2001.
- [GUI 01b] GUINOT V., “The discontinuous profile method for simulating two-phase flow in pipes using the single component approximation”, *International Journal for Numerical Methods in Fluids*, 37, pp. 341–359, 2001.
- [GUI 03a] GUINOT V., *Godunov-type Schemes, An Introduction for Engineers*, Elsevier, 2003.
- [GUI 03b] GUINOT V., “Riemann solvers and boundary conditions for two-dimensional shallow water simulations”, *International Journal for Numerical Methods in Fluids*, 41, pp. 1191-1219, 2003.
- [GUI 06] GUINOT V., SOARES-FRAZÃO S., “Flux and source term discretization for two-dimensional shallow water equations with porosity on unstructured grids”, *International Journal for Numerical Methods in Fluids*, 50, pp. 309–345, 2006.
- [GUI 07] GUINOT V., LEMÉNAGER M., CAPPELAERE B., “Sensitivity equations for hyperbolic conservation law-based flow models”, *Advances in Water Resources*, 30, pp. 1943–1961, 2007.

- [GUI 09a] GUINOT V., “Upwind finite volume solution of sensitivity equations for hyperbolic systems of conservation laws with discontinuous solutions”, *Computers and Fluids*, 38, pp. 1697–1709, 2009.
- [GUI 09b] GUINOT V., CAPPELAERE B., DELENNE C., “Finite-volume solution of one-dimensional shallow-water sensitivity equations”, *Journal of Hydraulic Research*, 47, pp. 811–819, 2009.
- [GUI 09c] GUINOT V., DELENNE C., CAPPELAERE B., “An approximate Riemann solver for sensitivity equations with discontinuous solutions”, *Advances in Water Resources*, 32, pp. 61–77, 2009.
- [GUN 99] GUNZBURGER M.D., “Sensitivities, adjoint and flow optimization”, *International Journal for Numerical Methods in Fluids*, 31, pp. 53–78, 1999.
- [HAL 82] HALL M.C.G., CACUCI D.G., SCHLESINGER M.E., “Sensitivity analysis of a radiative-convective model by the adjoint method”, *Journal of Atmospheric Sciences*, 39, pp. 2038–2050, 1982.
- [HAL 83] HALL M.C.G., CACUCI D.G., “Physical interpretation of adjoint functions for sensitivity analysis of atmospheric models”, *Journal of Atmospheric Sciences*, 40, pp. 2537–2546, 1983.
- [HAR 83a] HARTEN A., “High resolution schemes for hyperbolic conservation laws”, *Journal of Computational Physics*, 49, pp. 357–393, 1983.
- [HAR 83b] HARTEN A., LAX P.D., VAN LEER B., “On upstream differencing and Godunov-type schemes for hyperbolic conservation laws”, *Journal of Computational Physics*, 50, pp. 235–269, 1983.
- [HAR 84] HARTEN A., “On a class of high resolution total-variation-stable finite-difference schemes”, *SIAM Journal of Numerical Analysis*, 21, pp. 1–23, 1984 (with Appendix by P.D. Lax).
- [HER 07] HERVOUËT J.M., *Hydrodynamics of Free Surface Flows: Modelling with the Finite Element Method*, Wiley, 2007.
- [HOL 77] HOLLY F.M. JR, PREISSMANN A., “Accurate calculation of transport in two dimensions”, *Journal of the Hydraulics Division*, ASCE, 103, pp. 1259–1277, 1977.
- [HOP 50] HOPF E., *Communications in Pure and Applied Mathematics*, 3, 201, 1950.
- [HUB 00] HUBBARD M.E., GARCIA-NAVARRO P., “Flux difference splitting and the balancing of source terms and flux gradients”, *Journal of Computational Physics*, 165, pp. 89–125, 2000.
- [JAE 33] JAEGER C., *Théorie Générale du Coup de Bélier*, Dunod, Paris, 1933.
- [JAE 77] JAEGER C., *Fluid Transients in Hydro-Electric Engineering Practice*, Blackie and Sons, Glasgow London, 1977.

- [JIN 00] JIN M., FREAD D.L., "Discussion on the application of relaxation scheme to wave-propagation simulation in open channel networks", *Journal of Hydraulic Engineering*, 126, pp. 89–91, 2000.
- [JOH 02] JOHNSON T.C., BAINES M.J., SWEBY P.K., "A box scheme for transcritical flow", *International Journal for Numerical Methods in Engineering*, 55, pp. 895–912, 2002.
- [JOU 98] JOUKOWSKI N., "Water hammer", English translation O. Simin (1904), *Proc. American Waterworks Association*, 341, 1898.
- [KAT 77] KATOPODES N.K., Unsteady two-dimensional flow through a breached dam by the method of characteristics, PhD thesis, University of California, USA, 1977.
- [KAT 79] KATOPODES N., STRELKOFF T., "Two-dimensional shallow water-wave models", *Journal of the Engineering Mechanics Division (ASCE)*105, pp. 317–334, 1979.
- [KES 09] KESSERWANI G., MOSÉ R., VAZQUEZ J., GHENAIM A., "A practical implementation of high-order RKDG models for the 1D open-channel flow equations", *International Journal for Numerical Methods in Fluids*, 29, pp. 1389–1409, 2009.
- [KES 10] KESSERWANI G., LIANG Q., VAZQUEZ J., MOSÉ R., "Well-balancing issues related to the RKDG2 scheme for the shallow water equations", *International Journal for Numerical Methods in Fluids*, 62, pp. 428–448, 2010.
- [KRI 07] KRIVODONOVA L., "Limiters for high-order discontinuous Galerkin methods", *Journal of Computational Physics*, 226, pp. 276–296, 2007.
- [KUB 06] KUBATKO E.J., WESTERINK J.J., DAWSON C., "An unstructured grid morphodynamic model with a discontinuous Galerkin method for bed evolution", *Ocean Modelling*, 15, pp. 71–89, 2006.
- [LAX 54] LAX P.D., "Weak solutions of nonlinear hyperbolic equations and their computation", *Communications on Pure and Applied Mathematics*, 7(1), pp. 159–193, 1954.
- [LAX 57] LAX P.D., "Hyperbolic systems of conservation laws II", *Communications on Pure and Applied Mathematics*, 10, pp. 537–566, 1957.
- [LAX 60] LAX P.D., WENDROFF B., "Systems of conservation laws", *Communications on Pure and Applied Mathematics*, 13, pp. 217–237, 1960.
- [LED 86] LE DIMET F.X., TALGRAND O., "Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects", *Tellus*, 38A, pp. 97–110, 1986.
- [LEE 10] LEE S.H., WRIGHT N.G., "Simple and efficient solution of the shallow water equations with source terms", *International Journal for Numerical Methods in Fluids*, 63, pp. 313–340, 2010.
- [LEN 96] LENCASTRE A., *Hydraulique Générale*, Eyrolles, 1996.
- [LEV 02] LEVEQUE R.J., *Finite Volume Methods for Hyperbolic Problems*, Cambridge University Press, 2002.

- [LEV 98] LEVEQUE R.J., “Balancing source terms and flux gradients in high-resolution Godunov methods: the quasi-steady wave propagation algorithm”, *Journal of Computational Physics*, 146, pp. 346–365, 1998.
- [LHO 07] LHOMME J., GUINOT V., “A general approximate-state solver for hyperbolic systems of conservation laws with source terms”, *International Journal for Numerical Methods in Fluids*, 53, pp. 1509–1540, 2007.
- [LIA 09] LIANG Q., MARCHE F., “Numerical resolution of well-balanced shallow water equations with complex source terms”, *Advances in Water Resources*, 32, pp. 873–884, 2009.
- [LIG 55] Lighthill M.J., WHITHAM G.B., “On kinematic waves. I. Flood movement in long rivers”, *Proceedings of the Royal Society, A* 229, pp. 281–316, 1955.
- [LIU 75] LIU T.P., “The Riemann problem for general systems of conservation laws”, *Journal of Differential Equations*, 18, pp. 218–234, 1975.
- [LIU 76] LIU T.P., “The entropy condition and the admissibility of shocks”, *Journal of Mathematical Analysis and Applications*, 53, pp. 78–88, 1976.
- [LU 07] LU S., SAGAUT P., “Direct sensitivity analysis for smooth unsteady compressible flow using complex differentiation”, *International Journal for Numerical Methods in Fluids*, 53, pp. 1863–1886, 2007.
- [LYN 67] LYNESS J.N., MOLER C.B., “Numerical differentiation of analytic functions”, *SIAM Journal of Numerical Analysis*, 4, pp. 202–201, 1967.
- [MAR 95] MARCHUK G.I., *Adjoint Equations and Analysis of Complex Systems*, Kluwer Academic, Boston, 1995.
- [MES 97] MESELHE E., HOLLY F.M. JR., “Invalidity of the Preissmann scheme for transcritical flow”, *Journal of Hydraulic Engineering*, ASCE, 123, pp. 605–614, 1997.
- [NOV 10] NOVAK P., GUINOT V., JEFFREY A., REEVE D., *Hydraulic Modeling. Principles, Methods, Applications*, Spon Press, Abingdon, New York, 2010.
- [NUJ 95] NUJIC M., “Efficient implementation of non-oscillatory schemes for the computation of free-surface flows”, *Journal of Hydraulic Research*, 33, pp. 101–111, 1995.
- [PAN 89] PANCHANG V.G., O'BRIEN J.J., “On the determination of hydraulic model parameters using the adjoint state formulation”, *Modeling Marine Systems*, Boca Raton (Ed. A.M. Davies), pp. 6–18, 1989.
- [PRE 61a] PREISSMANN A., “Propagation des intumescences dans les canaux et rivières”, *Premier Congrès de l'Association Française du Calcul*, Grenoble, September 1961.
- [PRE 61b] PREISSMANN A., CUNGE J.A., “Calcul des intumescences sur machines électroniques”, *Proceedings of the 9th IAHR Conference*, Dubrovnik, 1961.
- [PRE 61c] PREISSMANN A., CUNGE J.A., “Calcul du mascaret sur machine électronique”, *La Houille Blanche*, 5, pp. 588–596, 1961.

- [ROE 81] ROE P.L., "Approximate Riemann solvers, parameter vectors and difference schemes", *Journal of Computational Physics*, 43, pp. 357–372, 1981.
- [SAN 00] SANDERS B.F., KATOPODES N.D., "Adjoint sensitivity analysis for shallow-water wave control", *Journal of Engineering Mechanics*, 126, pp. 909–919, 2000.
- [SAN 99] SANDERS B.F., KATOPODES N.D., "Control of canal flow by adjoint sensitivity method", *Journal of Irrigation and Drainage Engineering*, 125, pp. 287–297, 1999.
- [SAU 66] SAUERWEIN H., "The method of characteristics for three-dimensional unsteady magnetofluid dynamics of a multi-component medium", *Journal of Fluid Mechanics*, 25, pp. 17–41, 1966.
- [SAU 67] SAUERWEIN H., "Numerical calculations of multidimensional and unsteady flows by the method of characteristics", *Journal of Computational Physics*, 150, pp. 425–467, 1967.
- [SAV 93] SAVIC L., HOLLY F.M. JR., "Dambreak flood waves computed by modified Godunov method", *Journal of Hydraulic Research*, 31, pp. 187–204, 1993.
- [SOA 07] SOARES-FRAZÃO S., GUINOT V., "An eigenvector-based linear reconstruction scheme for the shallow water equations on two-dimensional unstructured meshes", *International Journal for Numerical Methods in Fluids*, 53, pp. 23–55, 2007.
- [STE 81] STEGER J.L., WARMING R.F., "Flux-vector splitting of the inviscid gasdynamic equations with application to finite-difference methods", *Journal of Computational Physics*, 40, pp. 263–293, 1981.
- [STO 57] STOKER J.J., *Water Waves*, Interscience Publishers Inc., New York, 1957.
- [STR 68] STRANG G., "On the construction and comparison of difference schemes", *SIAM Journal of Numerical Analysis*, 5 (3), pp. 506–517, 1968.
- [SWE 84] SWEBY P.K., "High resolution schemes using flux limiters for hyperbolic conservation laws", *SIAM Journal of Numerical Analysis*, 21, pp. 995–1011, 1984.
- [TOR 07] TORO E.F., GARCÍA-NAVARRO P., "Godunov-type methods for free-surface shallow flows: A review", *Journal of Hydraulic Research*, 45, pp. 736–751, 2007.
- [TOR 94] TORO E.F., SPRUCE M., SPEARES W., "Restoration of the contact surface in the HLL-Riemann solver", *Shock Waves*, 4, pp. 25–34, 1994.
- [TOR 97] TORO E.F., *Riemann Solvers and Numerical Methods for Fluid Dynamics*, Springer, Berlin Heidelberg, New York, 1997.
- [VAN 77] VAN LEER B., "Toward the ultimate conservative difference scheme. IV. A new numerical approach to convection", *Journal of Computational Physics*, 23, pp. 276–299, 1977.
- [VAN 79] VAN LEER B., "Toward the ultimate conservative difference scheme. V. A second-order sequel to Godunov's scheme", *Journal of Computational Physics*, 32, pp. 101–136, 1979.

- [VAZ 99] VÁZQUEZ-CENDÓN M.E., “Improved treatment of source terms in upwind schemes for the shallow water equations in channels with irregular geometry”, *Journal of Computational Physics*, 148, pp. 497–526, 1999.
- [VIC 82] VICHNEVETSKY R., BOWLES J.B., *Fourier Analysis of Numerical Approximations of Hyperbolic Equations*, SIAM, Philadelphia, 1982.
- [WAR 76] WARMING R., BEAM R., “Upwind second-order difference schemes and applications in aerodynamics flows”, *AIAA Journal*, 14, pp. 1241-1249, 1976.
- [WYL 77] WYLIE E.B., STREETER V.L. *Fluid Transients*, McGraw-Hill, New York, 1977.
- [ZHO 01] ZHOU J.G., CAUSON D.M., MINGHAM C.G., INGRAM D.M., “The surface gradient method for the treatment of source terms in the shallow water equations”, *Journal of Computational Physics*, 168, pp. 1–25, 2001.

Index

A, B

adsorption/desorption, 42, 51
 Freundlich's model, 43
 Langmuir's model, 43
 linear model, 43
advection equation, 14, 42, 165, 354
alternate directions
 explicit, 284
 implicit, 286
amplification factor, 476, 498
amplitude portrait, 476, 499
artificial viscosity, 507, 513-515
Bergeron's method, 82
boundary conditions, 12, 459
Buckley-Leverett equation, 35, 168
Burgers equation, 21, 305, 359

C

calibration, 411
characteristic
 form, 7, 16, 22, 30, 38, 45, 57, 66,
 75, 91, 113
 foot of a, 61
 method of, 227
 surface, 200
compressible flow, 63

computational domain, 61
concentration
 mass, 44
 volume, 14
conservation, 450
conservation form, 1, 14, 21, 28, 35,
 42, 63, 70, 85, 109
 derivation of, 3
conservation law
 concave, 419
 convex, 419
 hyperbolic system of, 1, 55, 195,
 196, 235
 non-convex, 421
 scalar, 1
conserved variable, 1, 53
consistency, 474, 487
 analysis, 487
 condition, 376
continuity equation, 71, 86, 110, 210,
 450
C–property, 377

D, E

dambreak problem, 177, 427, 453
Darcy's law, 35
dependence (domain of), 59, 80, 102,
 119, 204, 218, 457

differentiation

- analytical, 412
- code, 412
- complex, 412
- empirical, 412

diffusion

- equation, 495

flux, 2

numerical, 475, 489

Dirac's function, 418

discretization, 223, 330, 345

- multidimensional, 224

dispersion

- equation, 496
- numerical, 475, 489

Doppler effect, 221

eigenvalue, 483

eigenvector, 483

energy

- equation, 112

internal gas, 109

line, 29, 88

slope, 88

total, 108

entropy, 109

- condition, 152

- fix, 514

- violation, 514

equal area rule, 134

equation of state, 63, 109

error, 473

Euler equations, 108, 183

F, G

finite

- difference methods, 223

- element methods, 329

- volume methods, 293, 441

flow

- critical, 101

- sonic, 118

- subcritical, 101

subsonic, 118

supercritical, 101

supersonic, 118

transonic, 118

flux, 1, 53,

- function, 2

- splitting, 271

friction law

- Chezy's, 88

- Manning's, 88

- Strickler's, 28, 88

Galerkin technique, 335, 341

- discontinuous, 347

grid

- Cartesian, 224

- curvilinear, 224

- structured, 224

- unstructured, 224

H, I, J

Harten's criterion, 266

head (hydraulic), 36, 149-150

hydrocarbon, 35

implicitation parameter, 251, 332

influence (domain of), 59, 204, 218

initial conditions, 12

interpolation,

- first-order, 228

- second-order, 232

irreversibility, 154

iteration, 465

Jacobian matrix, 483

Joukowski's relationships, 124

jump (hydraulic), 149

jump relationships, 146, 416

K, L, M, N

kinematic wave, 28

Lagrange multiplier, 437

Legendre polynomial, 349

- limiter
 MC, 268-270
 minmod, 268-270, 352
 slope, 314-317
 Superbee, 268-270
 Van Leer, 268-270
- limiting, 351
 linearly degenerate, 415
 local partial inertia technique, 467
 mass matrix, 334
 mesh, 454
 momentum equation, 72, 87, 111, 210-212
 momentum flux, 64, 74, 91
 node, 330
 non-conservation form, 6, 30
 norm, 473
 number
 Courant, 226-229
 Froude, 101
 Mach, 118
- O, P,Q**
 objective function, 412, 436
 Petrov-Galerkin technique, 336, 342
 SUPG, 337, 345
- phase
 portrait, 476, 499
 space, 9
 velocity, 499
 pressure, 36, 63
 Quasi-Newton method, 436-439
- R**
 Rankine-Hugoniot conditions, 146
 approximations, 156
 reconstruction, 313-316
 reduced momentum equation
 technique, 468
 retardation factor, 45
- Riemann invariant, 7, 57-59, 75, 91, 96, 113, 216
 adjoint, 440
 generalized, 143
 sensitivity, 425
- Riemann problem, 161, 315
 generalized, 162
 sensitivity, 426
- Roe
 matrix, 280
 method, 281
 solver, 511
- Runge-Kutta time stepping, 261, 353
- S**
 Saint Venant equations, 84, 177
 scheme
 beam-warming, 266
 centered, 261
 CIR, 235
 conservative, 280
 Crank-Nicholson, 260
 explicit, 226
 Fromm, 266
 Godunov, 299
 Holly-Preissmann, 228
 implicit, 226, 245
 Lax-Wendroff, 266
 multidimensional, 288
 MUSCL, 316
 Preissmann, 250
 TVD, 263
 upwind, 244
- Secant plane approach, 202, 213
 sensitivity
 adjoint equations, 435
 empirical 412, 433
 forward equations, 413, 422
- shallow water equations
 one-dimensional, 323, 342, 387
 sensitivity, 427
 two-dimensional, 208

- shape function, 331, 345
 - shock tube, 183
 - solution
 - convergence, 473
 - discontinuous, 131, 416
 - existence, 61
 - monotony, 453
 - self-similar, 163
 - uniqueness, 61
 - weak, 144
 - Solver (Riemann)
 - approximate-state, 516
 - HLL, 443, 505
 - HLLC, 508
 - Lax-Friedrichs, 515
 - source term, 1, 53, 372
 - auxiliary variable-based balancing, 395
 - estimation of, 233
 - hydrostatic pressure reconstruction, 393
 - quasi-steady wave algorithm, 386
 - upwinding, 377
 - well-balanced schemes, 390
 - specific force, 65, 91, 375, 382, 428
 - splitting (time), 284, 342
 - stability, 491
 - analysis, 476, 492
 - stabilizing coefficient, 337
 - support function, 414
 - Sweby's criterion, 266
- T, W**
- Taylor series expansion, 414
 - test function, 329, 345
 - time step, 464
 - total variation, 263
 - truncation error, 474
 - water hammer, 68, 174, 237, 274-276, 308
 - wave
 - compound, 141
 - contact discontinuity, 140
 - number, 476, 499
 - rarefaction, 140
 - shock, 138
 - simple, 142
 - speed, 6, 8
 - strength, 512
 - weak form, 329, 344
 - weighted residuals, 330
 - wide channel approximation, 34