

# **OPTIMAL CONTROL PROBLEMS CONSTRAINED BY HYPERBOLIC CONSERVATION LAWS**



A THESIS SUBMITTED TO THE UNIVERSITY OF KWAZULU-NATAL  
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY  
IN THE COLLEGE OF AGRICULTURE, ENGINEERING & SCIENCE

By

Mohammed A. M. Tirab

School of Mathematics, Statistics & Computer Science

August 2021

# Contents

<b>Abstract</b>	<b>xii</b>
<b>Preface</b>	<b>xiii</b>
<b>Acknowledgements</b>	<b>xvii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Statement of the problem . . . . .	1
1.2 Method of solutions . . . . .	2
1.3 Literature review . . . . .	2
1.4 Aims and objectives . . . . .	5
1.5 Plan of the thesis . . . . .	6
<b>2 A review of hyperbolic conservation laws: Theory and Numerics</b>	<b>7</b>
2.1 Scalar conservation laws . . . . .	7
2.1.1 The linear Scalar case . . . . .	9
2.1.2 Non-linear scalar case . . . . .	9
2.1.3 Method of characteristics . . . . .	10

2.2	Weak Solutions . . . . .	13
2.2.1	Solutions to the Riemann problem . . . . .	13
2.2.2	Admissibility conditions for the weak solutions . . . . .	16
2.3	Systems of conservation laws . . . . .	20
2.3.1	Linear systems of conservation laws . . . . .	23
2.3.2	Non-linear systems of conservation laws . . . . .	27
2.3.3	Solution to the Riemann problem . . . . .	28
2.4	Finite volume method . . . . .	41
2.4.1	Derivation of the method . . . . .	41
2.4.2	Some numerical schemes . . . . .	44
2.5	Numerical results based on finite volume schemes . . . . .	49
2.5.1	Linear advection equation . . . . .	49
2.5.2	Burger's equation . . . . .	50
2.5.3	Shallow water equations . . . . .	54
2.5.4	System of Euler equations . . . . .	55
2.6	Relaxation systems . . . . .	57
2.7	Discretisation of the relaxation system . . . . .	59
2.7.1	Spatial discretisation . . . . .	59
2.7.2	Time integration . . . . .	62

2.8	Numerical results based on relaxation schemes . . . . .	65
2.8.1	Linear advection equation . . . . .	65
2.8.2	Burger’s equation . . . . .	66
2.8.3	Shallow water equations . . . . .	67
2.8.4	System of Euler equations . . . . .	68
2.9	Discontinuous Galerkin method . . . . .	70
2.10	Concluding remarks . . . . .	75
<b>3</b>	<b>Optimal control of 1D Scalar conservation laws using the relaxation method</b>	<b>76</b>
3.1	Introduction . . . . .	76
3.2	Problem formulation . . . . .	79
3.3	Derivation of an optimality system . . . . .	80
3.4	Numerical algorithm . . . . .	82
3.5	Discretisation techniques . . . . .	82
3.5.1	Spatial discretisation . . . . .	83
3.5.2	Time integration . . . . .	87
3.6	Numerical results and discussion . . . . .	89
3.6.1	Optimal control of advection equation . . . . .	90
3.6.2	Optimal control of Burger’s equation . . . . .	92

3.7	Tangent vectors approach . . . . .	96
3.7.1	Numerical scheme . . . . .	101
3.7.2	Optimisation algorithm . . . . .	106
3.8	Concluding remarks . . . . .	107
<b>4</b>	<b>Optimal control of multi-dimensional systems of conservation laws</b>	<b>108</b>
4.1	Problem formulation . . . . .	108
4.2	Multi-dimensional conservation laws . . . . .	109
4.3	Optimality conditions . . . . .	111
4.4	Space and time discretisation . . . . .	113
4.4.1	Forward equations . . . . .	113
4.4.2	Backward equations . . . . .	114
4.5	Three-dimensional relaxation approach . . . . .	115
4.5.1	Three-dimensional scheme . . . . .	120
4.6	Numerical algorithm . . . . .	124
4.7	Numerical results . . . . .	125
4.7.1	Solution of the flow equation . . . . .	126
4.7.2	Solution of the optimal control problem . . . . .	126
4.8	Concluding remarks . . . . .	131

<b>5</b>	<b>Optimal control problem governed by the multi-dimensional system of Euler equations</b>	<b>138</b>
5.1	Introduction . . . . .	138
5.2	Problem formulation . . . . .	141
5.3	A lattice Boltzmann approximation of the Euler equations . . . . .	142
5.4	Optimality conditions . . . . .	147
5.5	Numerical analysis and results . . . . .	156
5.5.1	Solution of the flow equations . . . . .	159
5.5.2	Solution to the optimal control problem . . . . .	159
5.6	Concluding remarks . . . . .	166
<b>6</b>	<b>Conclusion</b>	<b>167</b>
	<b>References</b>	<b>169</b>

# List of Tables

2.1	Error-norms for the Burger's equation (2.144) with initial data $U(0,x) = x$ , with grid points 100, 200 and 300. . . . .	52
2.2	Error-norms for the Burger's equation (2.144) with initial data (2.148), with grid points 100, 200, 300, 400 and 500. . . . .	54
2.3	Error-norms for the Linear advection equation (2.144) with smooth initial data $U(0,x) = x$ using relaxation schemes of different gridpoints 100, 200, 300, 400, 500 and 600. . . . .	67
2.4	Error-norms for the Burger's equation (2.144) with piecewise initial data (2.146) using relaxation schemes of different gridpoints 100, 200, 300, 400, 500 and 600. .	70
3.1	Computational time (in second) and the number of iterations (No. It.) for the inverse design in the advection equation (3.41) and Riemann data (3.43) obtained with the relaxation approach. . . . .	93
3.2	Computational time (CPU time in second) and the number of iterations (No. It.) for the inverse design in the Burgers equation (3.45) and Riemann data (3.47) obtained with the relaxation approach. . . . .	95

# List of Figures

2.1	The total amount of $U$ due to the flow across boundaries. . . . .	8
2.2	The height and velocity of water in a shallow channel. . . . .	33
2.3	Space discretisation for the finite volume. . . . .	41
2.4	Numerical flux across cell boundaries. . . . .	43
2.5	Numerical and exact solutions to the initial value problem (2.140) with $U(0,x) = x$ .	50
2.6	Solutions of linear advection equation (2.140) with Riemann initial data (2.142): Upwind scheme (Left) and Semi-discrete central Upwind scheme (Right). . . . .	51
2.7	Numerical and exact solutions to the Riemann problem (2.144) and (2.146). . . . .	52
2.8	Numerical and exact solutions to the Riemann problem (2.144) and (2.148). . . . .	53
2.9	Numerical results for the dam break problem (2.149) and (2.150). . . . .	55
2.10	Numerical results for the Shock tube problem (2.151) and (2.152). . . . .	56
2.11	Numerical and exact solutions to the Cauchy problem (2.140) and $U(0,x) = x$ using relaxation schemes: First-order (Left) and second-order (Right). . . . .	66
2.12	Solutions to the Riemann problem (2.140) and (2.142) using relaxation schemes: First-order (Left) and second-order (Right). . . . .	68
2.13	Numerical and exact solutions to the Cauchy problem (2.144) and $U(0,x) = x$ using relaxation schemes: First-order (Left) and second-order (Right). . . . .	68



2.14	Numerical and exact solutions to the Riemann problem (2.144) and (2.146) using relaxation schemes: First-order (Left) and second-order (Right). . . . .	69
2.15	Numerical and exact solutions to the Riemann problem (2.144) and (2.148) using relaxation schemes: First-order (Left) and second-order (Right). . . . .	69
2.16	Numerical results for the dam break problem (2.149) and (2.150) using relaxation schemes: First-order (Left) and second-order (Right). . . . .	71
2.17	Numerical results for the Shock tube problem (2.151) and (2.152) using relaxation schemes: First-order (Left) and second-order (Right). . . . .	72
3.1	Initial, optimised and target values (Left), surface solution (Middle) and convergence history (Right) of the optimal control of the linear advection equation (3.41) with initial data (3.42) obtained at $T = 0.02$ with the relaxation scheme: First-order (Top) and Second-order (Bottom). . . . .	91
3.2	Initial, optimised and target values (Left), surface solution (Middle) and convergence history (Right) of the optimal control of the linear advection equation (3.41) with initial data (3.43) obtained at $T = 0.4$ with the relaxation scheme: First-order (Top) and Second-order (Bottom). . . . .	92
3.3	Initial, optimised and target values (Left), surface solution (Middle) and convergence history (Right) of the optimal control of the Cauchy problem (3.45) with initial data (3.46), obtained with the relaxation scheme: First-order (Top) and Second-order (Bottom). . . . .	93
3.4	Initial, optimised and target values (Left), surface solution (Middle) and convergence history (Right) of the optimal control of the Riemann problem (3.45) with initial data (3.47) that comprises a shock wave, obtained with the relaxation scheme: First-order (Top) and Second-order (Bottom). . . . .	94

3.5	Initial, optimised and target values (Left), surface solution (Middle) and convergence history (Right) of the optimal control of the Riemann problem (3.45) with initial data (3.49) that comprises a rarefaction wave, obtained with the relaxation scheme: First-order (Top) and Second-order (Bottom). . . . .	96
4.1	Solution of the Cauchy problem for Burger's equation (4.63) with the initial data (4.64) on the rectangle $[0, 1] \times [0, 1]$ with grid points $100 \times 100$ , the results obtained with the finite volume scheme (Right) and the initial data (Left). . . . .	127
4.2	Solution of the Cauchy problem for Burger's equation (4.63) with the initial data (4.64) on the rectangle $[0, 1] \times [0, 1]$ with grid points $100 \times 100$ , the results obtained with the relaxation scheme (Right) and the initial data (Left). . . . .	127
4.3	Solution of the Riemann problem for Burger's equation (4.63) with the initial data (4.65) on the rectangle $[0, 1] \times [0, 1]$ with grid points $50 \times 50$ , the results obtained with the finite volume scheme (Right) and the initial data (Left). . . . .	128
4.4	Solution of the Riemann problem for Burger's equation (4.63) with the initial data (4.65) on the rectangle $[0, 1] \times [0, 1]$ with grid points $50 \times 50$ , the results obtained with the relaxation scheme (Right) and the initial data (Left). . . . .	128
4.5	Results of the optimal solution (Top-Left), the desired solution (Top-Right), the gradient (Bottom-Left) and the initial for desired (Bottom-Right) of the optimal control problem (4.1) for the Cauchy case. The results are obtained using the finite volume scheme with grid points $100 \times 100$ and time $T = 0.5$ . . . . .	130
4.6	Results of the optimal solution (Top-Left), the desired solution (Top-Right), the gradient (Bottom-Left) and the initial for desired (Bottom-Right) of the optimal control problem (4.1) for the Cauchy case. The results are obtained using the relaxation scheme with grid points $100 \times 100$ and time $T = 0.5$ . . . . .	131

4.7	Convergence history for the solution of the optimal control problem based on the finite volume method with the tolerance $\varepsilon = 10^{-3}$ and the initial control (4.64). . . . .	132
4.8	Convergence history for the solution of the optimal control problem based on the relaxation method with the tolerance $\varepsilon = 10^{-3}$ and the initial control (4.64). . . . .	133
4.9	Plot of the optimal solution (Top-Left), the desired solution (Top-Right), the gradient (Bottom-Left) and the initial for desired (Bottom-Right) of the optimal control problem (4.1) for the Riemann case. The results are computed using the finite volume method with grid points $50 \times 50$ and time $T = 0.2$ . . . . .	134
4.10	Plot of the optimal solution (Top-Left), the desired solution (Top-Right), the gradient (Bottom-Left) and the initial for desired (Bottom-Right) of the optimal control problem (4.1) for the Riemann case. The results are computed using the relaxation method with grid points $50 \times 50$ and time $T = 0.2$ . . . . .	135
4.11	Convergence history for the solution of the optimal control problem related to the finite volume scheme with the tolerance $\varepsilon = 10^{-3}$ and the initial control (4.65). . . . .	136
4.12	Convergence history for the solution of the optimal control problem related to the relaxation scheme with the tolerance $\varepsilon = 10^{-3}$ and the initial control (4.65). . . . .	137
5.1	Two-dimensional, nine-velocities (D2Q9) square Lattice Boltzmann Model. . . . .	145
5.2	Contour plot of the density, pressure and energy for the solution of the Euler equations in two dimensions using the Lattice Boltzmann method. We use $500 \times 500$ grid points and the results are computed up to time $T = 0.0063$ . . . . .	160
5.3	Contour plot of the density, pressure and energy for the solution of the Euler equations in two dimensions using the finite volume method. We use $500 \times 500$ grid points and the initial condition (left) and the results (right) are computed up to time $T = 0.0063$ . . . . .	161

5.4	Contour plot of the optimal (left) and desired solutions (right) density, pressure and energy for the solution of the optimal control problem for the first example. The results are computed at time $T = 0.0005$ . . . . .	163
5.5	Convergence history for the solution of the optimal control problem computed with the tolerance $\varepsilon = 10^{-3}$ using the initial control (5.74). . . . .	164
5.6	Contour plot of the optimal (left) and desired solutions (right) density, pressure and energy for the solution of the optimal control problem for the second example. The results are computed at time $T = 0.0005$ . . . . .	165
5.7	Convergence history for the solution of the optimal control problem computed with the tolerance $\varepsilon = 10^{-3}$ using the initial control (5.76) . . . . .	166

# Abstract

This thesis deals with the solutions of optimal control problems constrained by hyperbolic conservation laws. Such problems pose significant challenges for mathematical analysis and numerical simulations. Those challenges are mainly because of the discontinuities that occur in the solutions of non-linear systems of conservation laws and become more acute when dealing with the multi-dimensional case.

The problem is formulated as the minimisation of a flow matching cost functional constrained by multi-dimensional hyperbolic conservation laws. The control variable is the initial condition of the partial differential equations.

In our analysis of the problem, we review extensively the constraints equation and we consider successively the one-dimensional and the multi-dimensional cases. In all the cases, we derive the optimality conditions in the adjoint approach at the continuous level, which are then discretised to arrive at a numerical algorithm for the solution. In the derivation of the optimality conditions, we replace the non-linear conservation laws either by the relaxation equation or the Lattice Boltzmann equation. We illustrate our findings on examples related to the multi-dimensional Burger and the Euler equations.

# Preface

The studies carried out in this thesis were under the direct supervision of Dr. Jean Medard T. Ngnotchouye in the School of Mathematics, Statistics and Computer Science, College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Pietermaritzburg, from August 2016 to August 2021.

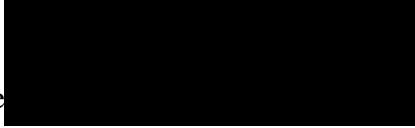
I hereby declare that the thesis represents the original work of the author except where duly acknowledged and referenced. No portion of the thesis has been submitted to another university or institution of learning for any degree or qualification.

Signature: 

Mohammed A. M. Tirab

.....08.11.2021.....

Date

Signature 

Dr. Jean Medard T. Ngnotchouye

.....16-11-2021.....

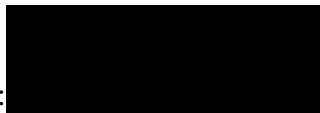
Date

# Declaration 1: Plagiarism

I, Mohammed A. M. Tirab, declare that:

- (i) the studies conducted in this thesis, except where duly indicated and acknowledged, represent my original effort;
- (ii) this thesis has not been submitted in full or in part for any degree or examination to any other university;
- (iii) this thesis does not contain data, pictures, graphs or other information that belongs to someone else unless where duly acknowledged;
- (iv) this thesis does not contain the writing of others unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
  - (a) their words have been re-written but the general information attributed to them has been referenced;
  - (b) where their exact words have been used, their writing has been placed inside quotation marks and referenced;
- (v) this thesis does not contain text, graphics or tables copied and pasted from the Internet unless specifically acknowledged and the source being detailed the thesis and the References sections.

Signature:



Mohammed A. M. Tirab

08.11.2021

Date

## **Declaration 2: Publication**

Jean Medard T. Ngnotchouye and Mohammed A. M. Tirab. A lattice Boltzmann approach to optimal control problems constrained by multidimensional Euler equations. Submitted to Networks and Heterogeneous Media Paper ID: 210707-Ngnotchouye. *Under review*



# Dedication

I dedicate this work:

To the memory of my late mother, may Allah mercy be upon her. Your words, prayers and love live on.

To my father, for his love, support and encouragement paved my way to success.

To my wife, your love, encouragement and support are always brightening our path.

To whom I am proud of them, my children, who have all been incredibly patient and understanding throughout my degree programme. Your love made our life going on with great happiness and forever.

To my brothers, sisters and their children for their support and encouragement.

To my uncles, aunts and cousins for their support and encouragement.

To all my students and my great teachers in different education levels.

# Acknowledgements

I thank Allah for providing me with the strength and the faith to begin and complete this thesis. He kept me throughout smooth and turbulent times and I emerged unscathed.

I would like to thank all those who assisted and encouraged me to accomplish this project.

My thanks are conveyed to my advisor and supervisor, Dr. Jean Medard T. Ngnotchouye, for his advice, recommendations and motivation during the period of my Ph.D. research.

I am grateful to Prof. Precious Sibanda for his arrangement with the supervisor of my Ph.D. programme, many thanks.

I would also like to thank staff members of the School of Mathematical Sciences of the University of KwaZulu-Natal, Pietermaritzburg.

Many thanks, especially to Dr. Absalom, for the assistance, advice and encouragement.

I would also like to thank the school administrator, Ms. C. Barnard, for making me comfortable throughout my programme.

Thanks also to the technical team for making equipment and facilities readily available for my every need.

I would also like to thank the academic higher degrees officers, Ms. Tamlyn Skye and Mrs. Jothimala Manickum for their assistance throughout the examination period.

I am thankful to my close teachers, Prof. Fathy, Dr. Hafiz, Dr. Abdulgadir, Dr. Ali and Mrs. Zenab. Your kind guidance, advice and encouragement are very much appreciated.

I am grateful to friends and colleagues, Dr. Olumuyiwa, Dr. Shina, Dr. Ghirmay, Dr. Orakwelu, Dr. Degoot, Ali and Salaheldin. Thank you very much.

My deepest thank you to Dr. Olumuyiwa and Dr. Shina for always being available to assist and edit many parts of my draft. Your time, feedback and encouragement are very much appreciated.

I want to extend my gratitude to my friends, Dr. Huzifa, Dr. Abdulmajid, Dr. Berimah, Dr. Amal, Dr. Omer, Dr. Alfadel, Dr. Yasir, Ahmed, Rashed, Mohammed, Adam, Makeya, Fayze, Mousa,

Ibrahim, Alzaky, Slah, Babker, Abdulgebar, Arif, Samirah, Aidah, Naemat, Rawiah, Eman and the rest of my beautiful people back in Sudan.

I would like to express my appreciation to my students in Sudan for always being in touch with me and their encouragement. Many thanks.

Lastly, I am incredibly grateful to my family in Sudan and abroad, including my dear father, dear wife, children, brothers, sisters, nephews, uncles, aunts and cousins. Without the prayers, patience, love and support from all of you, I would not have completed this journey.

# Chapter 1

## Introduction

### 1.1 Statement of the problem

In this study, we consider the mathematical analysis and numerical solutions of the optimal control problem constrained by multi-dimensional systems of conservation laws of the form

$$\frac{\partial U}{\partial t} + \nabla \cdot f(U) = 0, \quad (1.1)$$

where  $U = U(t, \mathbf{x}) \in \mathbb{R}^m$  is the conserved quantity which depends on time  $t \in \mathbb{R}_+$  and spatial coordinates  $\mathbf{x} \in \mathbb{R}^n$ ,  $f(U)$  is the flux function, which is a nonlinear mapping with components  $f_i : \mathbb{R}^m \rightarrow \mathbb{R}^m$  for  $i = 1, \dots, n$ , and  $\nabla = \left( \frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_d} \right)$ . The initial conditions at time  $t = 0$  are given as

$$U(0, \mathbf{x}) = U_0(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^n. \quad (1.2)$$

For a given desired state  $U_d$  prescribed at final time  $T$ , the dynamical optimisation problem is formulated as

$$\begin{aligned} & \underset{U_0}{\text{Minimise}} J(U(T, \cdot), U_0; U_d), \\ & \text{subject to equations (1.1) and (1.2),} \end{aligned} \quad (1.3)$$

where  $J$  denotes the cost functional, usually taken as a flow matching type.

## 1.2 Method of solutions

There are two main approaches for solving optimal control problems numerically, the continuous (optimise-then-discretise) approach and the discrete (discretise-then-optimise) approach [1–3]. In the continuous approach, we usually use the minimum principle to derive some optimality conditions that amount to the original equations, adjoint equations with the terminal conditions, and the gradient of the reduced cost functional. Then, these conditions are discretised to solve the problem. In the discrete approach, we first discretise the cost functional and the constraints which lead to an optimisation problem in high dimensional spaces. The problem is then solved using the so-called Karush–Kuhn–Tucker (KKT) conditions.

In a naive continuous approach to solve problem (1.3), we obtain an adjoint equation of the form

$$-\frac{\partial}{\partial t}P(t, \mathbf{x}) - Df(U)\nabla \cdot P(t, \mathbf{x}) = 0, \quad (1.4)$$

where  $P(t, \mathbf{x})$  is the adjoint variable and  $Df(U)$  is the Jacobian matrix of  $f(U)$  with respect to  $U$ . Solutions of the flow equation (1.1) have discontinuity, even if the initial conditions are smooth [4–8]. Therefore, the term  $Df(U)$  in equation (1.4) will be discontinuous and the product  $Df(U)\nabla \cdot P$  will be undefined.

In order to solve problem (1.3) numerically, we use an adjoint-based approach and in the Lagrangian formulation that leads to the optimality conditions, we replace the nonlinear system of conservation laws (1.1) by a semi-linear approximation by either a relaxation approximation [9] or a Lattice Boltzmann approximations [10]. The convective part of these approximations is linear, which will allow us to resolve the problem in equation (1.4).

## 1.3 Literature review

The existence and uniqueness of solutions of the optimal control problem (1.3) rely on the existence of solutions of (1.1) and the existence of a solution of constrained optimisation problems. In addition, the semi-group generated by a conservation law, in general, is non-differentiable in  $L^1$

even in the scalar, one-dimensional case [11, 12]. A differential structure on general bounded variation solutions for hyperbolic conservation laws in one space dimension was presented by Bressan and Marson [13], Stefano Bianchini [14] and Bianchini and Yu [15]. For one-dimensional Cauchy problems, Glimm [16] provided proof for the global existence of the entropy admissible solutions using the random choice method. The deterministic proof of existence and the uniqueness results of the weak solutions were proven by Bressan et al. [17–21]. Results on the existence and regularity of entropy solutions of the multi-dimensional system of conservation laws have been presented by Lions et al. [22], LeFloch [23], Neves [24], Panov [25], Chen [26], Crasta et al. [27], Dogbe and Bianca [28] and the reference therein.

The solution of optimal control problems related to the system of conservation laws has attracted a lot of attention in the published literature. Different methods have been proposed [29–34], among others, Morales-Hernández and Zuazua [35] considered a computational method for the two-dimensional inverse design of linear transport equations on unstructured grids. Lecaros and Zuazua [36] analysed an inverse design problem for the two-dimensional scalar conservation law in the presence of shock based on an alternating descent method and the finite difference method. Herty et al. [37] proposed an iterative algorithm for the solution to optimisation problems governed by scalar hyperbolic conservation laws. They analysed the convergence properties of adjoint and gradient approximations on an unbounded domain with a strictly convex flux. Schäfer Aguilar et al. [38] presented convergence of discretisation schemes for the adjoint equation arising in the adjoint-based derivative computation for optimal control problems related to the entropy solutions of conservation laws. Pfaff and Ulbrich [39, 40] considered the optimal control of initial-boundary value problems for entropy solutions of scalar hyperbolic conservation laws. They proved that the control-to-state mapping is differentiable in a certain generalised sense of differentiability. Hajian et al. [41] presented optimal control problems subject to a nonlinear scalar conservation law based on the discretise-then-optimise approach and the total variation diminishing Runge-Kutta schemes. Zeng et al. [42] proposed a nonlinear optimal control method related to the Saint-Venant PDEs with conservation laws via a control parameterisation approach. Frenzel and Lang [43] proposed

a third-order weighted essentially non-oscillatory method and used the discretise-then-optimize approach for solving optimal control problems subject to scalar nonlinear hyperbolic conservation laws. Hintermüller and Strogies [44] discussed the optimal control of scalar conservation laws with strong stability preserving Runge-Kutta methods. Zahr and Persson [45] considered an adjoint method to solve shape optimisation problems on deforming domains conservation law and applied the discontinuous Galerkin method to discretize the transformed equations. Herty et al. [46] investigated optimal control problems governed by hyperbolic systems and developed a numerical method for the solution to linear adjoint equations arising in the optimisation problem.

The numerical solution of the system of conservation laws was presented by Kurganov et al. [47], Wang et al. [48] and Gottlieb et al. [49, 50] as well as the studies in [32, 50–52]. These publications discussed finite volume methods in the sense of the Reconstruct-Evolve-Average (REA) algorithm [53] to obtain solutions over the control volumes. REA algorithm consists of reconstructing a piecewise a polynomial function of the computed solution arising at cell interfaces [4, 7, 54], evolve the equation exactly or approximately to obtain the solution to the next time level and then an average of solution over each grid cell to obtain the new cell averages [55].

Other approaches for the solution of conservation laws are the relaxation methods or the lattice Boltzmann methods. The convergence analysis of the relaxation approximation (commonly known as Jin - Xin relaxation approximations) is presented in [56–58]. The discrete kinetic model was introduced by Aregba-Driollet et al. [59], Natalini and Terracina [60] and R. Natalini [61], also known as a Bhatnagar–Gross–Krook (BGK) approximation or lattice Boltzmann approximation, see [62, 63] for example. These approximations preserve the hyperbolic structure with additional source terms and we can solve the problem numerically without introducing Riemann solvers.

Recently, there has been some development on solutions of optimal control problems constrained by the relaxation approximations to systems of conservation laws by Nørsgaard et al [64] for the solution of a shape and topology optimisation problem. Yohana and Banda [65] and Steffensen et al. [66] considered high-order relaxation approaches for the solution of the one-dimensional optimal control problems and used the adjoint method. Banda and Herty [67] used continuous

and discrete schemes for optimisation problems subject to nonlinear, scalar hyperbolic conservation laws. They constructed adjoint implicit-explicit-based schemes for control problems. Li et al. [68] considered a problem of airfoil design optimisation and combined the Lattice Boltzmann methods and the adjoint method. Herty et al. [69] considered a computational method for the two-dimensional problems and used an implicit-explicit scheme for the time stepping. They presented results related to the two-dimensional inviscid Burger's equation. Ngnotchouye et al. [70] proposed relaxation approaches to the optimal control of the Euler equations. They used an adjoint method and the one-dimensional, five velocities lattice Boltzmann approximations. Herty and Piccoli [71] presented a numerical method for solving optimal control problems based on a combination of the relaxation approach and numerical scheme of the tangent vectors. Albi et al. [72] proposed a linear multistep method for the solution of optimal control problems and combined semi-Lagrangian approximations with the one-dimensional hyperbolic relaxation systems.

## 1.4 Aims and objectives

For the solution of our optimal control problem (1.3), we propose in the continuous framework a set of optimality conditions that leads to a numerical algorithm. The algorithm involves the solution of the system of conservation laws forward in time, the solution of the adjoint equation backward in time and an update of the control using the gradient of the reduced cost functional. Precisely, we focus on the numerical algorithm for the solution of optimal control problems related to the multi-dimensional case. The system of conservation laws and the adjoint equations have to be solved on the same grid. Further, we contribute to a new optimisation algorithm based on the multi-dimensional lattice Boltzmann method. Also, we extend the algorithm related to the two-dimensional relaxation schemes discussed by Herty et al. [69] to the three-dimensional case. We obtain efficient algorithms that perform well for many problems of interest.



## 1.5 Plan of the thesis

The thesis is outlined as follows:

In Chapter 2, we review the general properties of the solutions of systems of conservation laws. Due to discontinuities that arise in the solution of conservation laws, the weak solutions are rigorously defined and the solution of the so-called Riemann problem is presented. Then, we discuss two approaches for the numerical solutions of the conservation laws namely finite volume and relaxation methods.

In Chapter 3, we discuss optimal control problems governed by scalar hyperbolic conservation laws. We derive the optimality conditions in the adjoint framework by replacing the original equations with the relaxation approximations. The optimality conditions are later discretised and an algorithm for the solutions of the problem is presented.

In Chapter 4, we extend the results of Chapter 3 to the multi-dimensional case. The results on the existence and uniqueness of an entropy solution of multi-dimensional systems of conservation laws are presented. We extend to the three-dimensional case some optimality conditions obtained by Herty et al. [69] for the two-dimensional case.

Chapter 5 deals with optimal control problems constrained by the multi-dimensional Euler equations of gas dynamics. We use a Lattice Boltzmann approximation of the Euler equation to derive the optimality conditions at the kinetic level. We illustrate our results on the two-dimensional nine velocities (2D9Q) lattice Boltzmann approximation of the Euler equations. Moreover, we present numerical results and compare them with those obtained using the finite volume method.

Chapter 6 contains the general conclusion of the thesis.

# Chapter 2

## A review of hyperbolic conservation laws: Theory and Numerics

This chapter deals with the general theory on systems of conservation laws and their numerical approximations. In particular, an investigation is conducted on obtaining exact and approximate solutions to conservation laws in one-dimensional space. Most of the materials of this chapter are taken from [4, 6, 73–75] for scalar conservation laws, from [4, 5, 76–80] for systems of conservation laws, from [9, 57, 81, 82] for the relaxation system, and from [4, 7, 76, 83–87] for the numerical simulations.

### 2.1 Scalar conservation laws

In this section, we are interested in scalar conservation laws. The system of conservation laws in one-dimensional is formulated as

$$\frac{\partial U}{\partial t} + \frac{\partial f(U)}{\partial x} = 0, \quad x \in \mathbb{R}, \quad t > 0, \quad (2.1)$$

where  $U$  is called the conserved quantity,  $f(U)$  is the flux function, the partial derivatives  $\frac{\partial U}{\partial t}$  and  $\frac{\partial f(U)}{\partial x}$  are with respect to time ( $t$ ) and space ( $x$ ), respectively.

Equation (2.1) is considered as fundamental laws of nature and has applications in fluid models like the shallow water equation and Euler equation to formulate a model for the flow in canals and pipes, respectively, to mention a few.

Consider equation (2.1) to be the scalar conservation law, where  $U \in \mathbb{R}$  with  $f(U) : \mathbb{R} \mapsto \mathbb{R}$

is the flux function that can be continuously differentiable. Integrating (2.1) over an interval of  $[a, b] \subset \mathbb{R}$ , one obtains

$$\begin{aligned} \frac{d}{dt} \int_a^b U(t, x) dx &= \int_a^b \frac{\partial}{\partial t} U(t, x) dx = - \int_a^b \frac{\partial}{\partial x} f(U(t, x)) dx \\ &= f(U(t, a)) - f(U(t, b)) = [\text{inflow at } a] - [\text{outflow at } b]. \end{aligned} \quad (2.2)$$

Equation (2.2) implies that the quantity of  $U$  is neither generated nor destroyed; the total amount of  $U$  stored within any given interval  $[a, b]$  will change only due to the flow of  $U$  through boundary points (see figure 2.1).

Equation (2.1) can be written in quasi-linear form using the chain rule, as

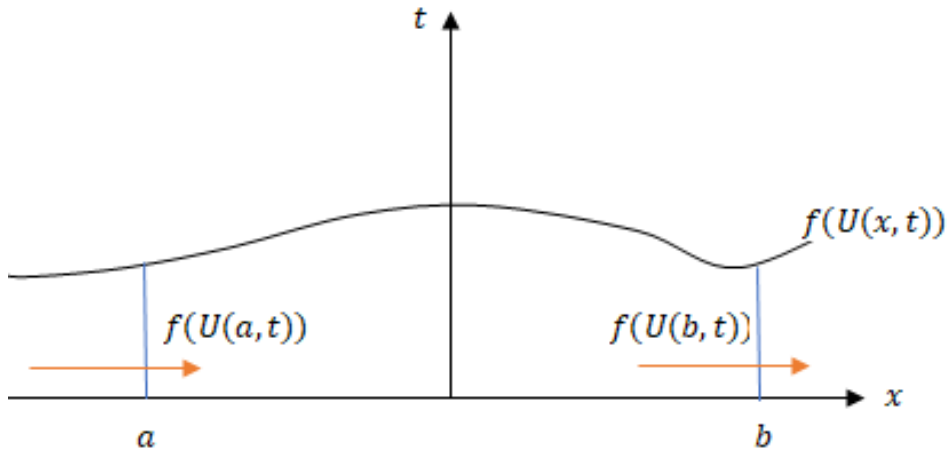


Figure 2.1: The total amount of  $U$  due to the flow across boundaries.

$$\frac{\partial U}{\partial t} + f'(U) \frac{\partial U}{\partial x} = 0, \quad (2.3)$$

where  $f'(U)$  represents the derivative of  $f(U)$  with respect to  $U$ .

To obtain smooth solutions, equations (2.1) and (2.3) have to be fully equivalent. Nevertheless, if  $U$  has a jump at a point  $\xi$ , the left-hand side of (2.3) will include the product of a discontinuous function  $a(U) := f'(U)$  with the distributional derivative  $\frac{\partial U}{\partial x}$ , which in this illustration, contains a Dirac mass at the point  $\xi$ . Universally, such a product is not well-defined and equation (2.3) is meaningful only within a class of continuous functions. On the other hand, working with the equation in divergence form, equation (2.1) allows us to consider discontinuous solutions when

interpreted in a distributional sense. More precisely, a locally integrable function  $U = U(t, x)$  is a weak solution of (2.1) provided that

$$\iint \left[ U \frac{\partial \phi}{\partial t} + f(U) \frac{\partial \phi}{\partial x} \right] dx dt = 0, \quad (2.4)$$

for every differentiable function with compact support  $\phi \in C_c^1$ . Notice that equation (2.4) is only meaningful if both  $U$  and  $f(U)$  are locally integrable in the  $x, t$ - plane. Two cases of scalar conservation laws that depend on the conserved quantity variable  $U$  and we will consider both cases in the subsequent sections.

### 2.1.1 The linear Scalar case

The linear homogeneous Cauchy problem with constant coefficients is of the form

$$\frac{\partial U}{\partial t} + \lambda \frac{\partial U}{\partial x} = 0, \quad U(0, x) = \bar{U}(x), \quad (2.5)$$

with  $\lambda \in \mathbb{R}$ . Here, the solution to the Cauchy problem can be written explicitly. If  $\bar{U} \in C^1$ , it is easy to verify that the travelling wave

$$U(t, x) = \bar{U}(x - \lambda t), \quad (2.6)$$

provides a classical solution to equation (2.5). Where the initial condition  $\bar{U}$  is not differentiable and has just  $\bar{U} \in L^1_{loc}$ , the function  $U$  defined by equation (2.6) can be regarded as a solution in a distributional sense.

Equation (2.5) is called the advection equation and when  $\bar{U} = x$ , the solution

$$U(t, x) = x - \lambda t, \quad (2.7)$$

is called a travelling wave with speed  $\lambda$ .

### 2.1.2 Non-linear scalar case

In this section, the coefficients of the Cauchy problem (2.1) depend on the solution  $U$  and we assume that the flux function  $f(U)$  is differentiable. In this case, the most important property is

that the characteristic curves of solutions will intersect (even for smooth data) after a small time interval due to non-linearity.

### 2.1.3 Method of characteristics

Consider the scalar Cauchy problem

$$\frac{\partial U}{\partial t} + \frac{\partial f(U)}{\partial x} = 0, \quad U(0, x) = \bar{U}(x). \quad (2.8)$$

With smooth solutions, the equation can be written in the quasilinear form as

$$\frac{\partial U}{\partial t} + f'(U) \frac{\partial U}{\partial x} = 0. \quad (2.9)$$

Geometrically, this implies that the directional derivative of  $U(t, x)$  in the direction of the vector  $(1, f'(U))$  vanishes. We assume that a curve  $x(t)$  solves the ordinary differential equation

$$\frac{dx}{dt} = U(t, x(t)).$$

Let  $y(t) = U(t, x(t))$ , then  $dy/dt = 0$ . Notice that

$$\frac{d}{dt} U(t, x(t)) = \frac{\partial U}{\partial t} + \frac{\partial U}{\partial x} \frac{dx}{dt}. \quad (2.10)$$

Characteristic equations are found using equation (2.9) with equation (2.10), as

$$\begin{aligned} \frac{dU}{dt} &= 0, \quad \text{and} \\ \frac{dx}{dt} &= f'(U), \\ x(0) &= x_0. \end{aligned} \quad (2.11)$$

where  $x_0$  are the initial points on the characteristic curves  $x(t)$ . Equation (2.11) shows that we can use the initial data  $\bar{U}(x)$  to propagate the solution in the  $tx$ -plane.

Hence  $U$  is constant on each line of the form  $\{(t, x) : x = x_0 + t f'(U(x_0))\}$ . For each  $x_0 \in \mathbb{R}$ , we have, thus

$$U(x_0 + t f'(U(x_0)), t) = \bar{U}(x_0). \quad (2.12)$$

(2.12) is indeed the solution to the first order PDE (2.9) provided by the classical method of characteristics [74].

Furthermore, for any starting points  $x_0$  and  $x_1$ , with  $x_0 < x_1$  for two characteristic equations, we have

$$x_0 + t f'(U(x_0)) = x_1 + t f'(U(x_1)).$$

Solving for  $t$  gives

$$t = \frac{x_0 - x_1}{f'(U(x_1)) - f'(U(x_0))},$$

this equation rewritten as

$$t = \frac{-1}{\frac{f'(U(x_1)) - f'(U(x_0))}{x_1 - x_0}}.$$

Using the mean value theorem, there exists  $\xi \in (x_0, x_1)$  such that

$$f'(\bar{U}(\xi))\bar{U}'(\xi) = \frac{f'(U(x_1)) - f'(U(x_0))}{x_1 - x_0},$$

thus  $t = -1/f'(\bar{U}(\xi))\bar{U}'(\xi)$ . Since  $x_0$  and  $x_1$  are arbitrary, there must be an interval for which  $f'(\bar{U}(\xi))\bar{U}'(\xi) < 0$ , which guarantees that  $t \geq 0$ . Hence, we can define the finite time  $t_c$  from the classical solution of (2.8) as

$$t_c = -\frac{1}{\min_{x \in \mathbb{R}} \{f'(\bar{U}(x))\bar{U}'(x)\}} > 0,$$

which is the minimum time at which the derivative of the solution  $U$  with respect to  $x$  and  $t$  becomes infinite. The time  $t_c$  is referred to as the critical time (the smallest non-negative number for which the characteristics intersect). In general, beyond a finite time  $t_c$ , the map

$$x_0 \longmapsto x_0 + t f'(U(x_0)),$$

is no longer one-to-one and the implicit equation (2.12) does not define a single-valued function  $U = U(t, x)$ . At critical time  $t_c$ , a shock is formed and the solution can be extended for  $t > t_c$  in the weak sense, as in (2.4). Precisely, we can see that in the following example of the shock formation in the scalar Cauchy problem of Burger's equation.

**Example 2.1.1.** Consider the inviscid Burger's equation

$$\frac{\partial U}{\partial t} + \frac{\partial}{\partial x} \left( \frac{U^2}{2} \right) = 0, \quad (2.13)$$

with initial condition

$$U(0, x) = \bar{U}(x) = \sin(x). \quad (2.14)$$

For  $t > 0$  small, the solution can be found by the method of characteristics. Indeed, if  $U$  is smooth, (2.13) is equivalent to

$$\frac{\partial U}{\partial t} + U \frac{\partial U}{\partial x} = 0. \quad (2.15)$$

From equation (2.15), the directional derivative of the function  $U = U(t, x)$  along the vector  $(1, U)$  vanishes. Therefore,  $U$  must be constant along the characteristic lines in the  $tx$ - plane:

$$t \mapsto (t, x + t\bar{U}(x)) = (t, x + t\sin(x)).$$

Moreover, the characteristics line of (2.13)- (2.14) is  $x = x_0 + t\sin(x_0)$ . Given any two points  $x_0 = -\frac{\pi}{2}$  and  $x_1 = \frac{\pi}{2}$ , we can find the critical time  $t_c = \frac{\pi}{2}$ .

For  $t < t_c = \frac{\pi}{2}$ , these lines do not intersect. The solution to our Cauchy problem is thus given implicitly by

$$U(t, x + t\sin(x)) = \sin(x). \quad (2.16)$$

Conversely, when  $t > t_c = \frac{\pi}{2}$ , the characteristic lines intersect. As a result, the map

$$x \mapsto x + t\sin(x),$$

is not one-to-one, and (2.16) no longer defines a single-valued solution of our Cauchy problem.

Thus, the smooth solution of conservation laws can either blow up in time or can be multi-valued. Also, smooth solutions are not global in time and require a new solution, which is called a weak solution.

## 2.2 Weak Solutions

In the classical solutions described above, we may have a loss of regularity of the solutions in the sense that the derivatives can become infinite in finite time or  $U(t, x)$  can have multi-value functions where the characteristics meet. Therefore, to solve the problems globally, we have to look for weak solutions.

**Definition 2.2.1.** Let  $U(t, x)$  be a measurable function defined on an open set  $\Omega \subseteq \mathbb{R} \times \mathbb{R}$  with values in  $\mathbb{R}$  and let  $f(U)$  be a smooth vector field function from  $\mathbb{R}$  to  $\mathbb{R}$ , we say that  $U$  is a distributional solution of the scalar conservation laws (2.1) if

$$\iint_{\Omega} \left[ U \frac{\partial \phi}{\partial t} + f(U) \frac{\partial \phi}{\partial x} \right] dx dt = 0, \quad (2.17)$$

for every  $C^1$  function  $\phi$  from  $\Omega$  to  $\mathbb{R}$  with compact support.

From the definition above,  $U$  is not necessarily a continuous function while  $U$  and  $f(U)$  must be locally integrable functions in  $\Omega$ .

**Definition 2.2.2.** Let  $U$  as in Definition 2.2.1 and an initial data defined such that  $U(0, x) = \bar{U}(x)$ , we say that  $U$  is a distributional solution to the Cauchy problem (2.1) with  $\bar{U}(x)$ , if for every  $C^\infty$  function  $\phi$  with compact support in the strip  $[0, T] \times \mathbb{R}$ ,  $U$  satisfy the condition

$$\int_0^T \int_{-\infty}^{\infty} \left[ U \frac{\partial \phi}{\partial t} + f(U) \frac{\partial \phi}{\partial x} \right] dx dt + \int_{-\infty}^{\infty} \bar{U} \phi(0, x) dx = 0, \quad (2.18)$$

for all  $i = 1, \dots, n$  and  $\bar{U} \in L^1_{loc}(\mathbb{R}; \mathbb{R})$ .

**Definition 2.2.3.** A function  $U : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$  is a weak solution of the Cauchy problem (2.1) with  $U(0, x) = \bar{U}(x)$ , if  $U$  is a continuous function from  $[0, T]$  into  $L^1_{loc}$ , the initial data  $\bar{U}(x)$  holds and the restriction of  $U$  to an open strip  $]0, T[ \times \mathbb{R}$  is a distributional solution of (2.1).

### 2.2.1 Solutions to the Riemann problem

In this section, the general construct of solutions to the Riemann problem for scalar conservation laws is discussed. Consider the scalar conservation law (2.1) with the initial piecewise constant



data, given as

$$U(0, x) = \bar{U}(x) = \begin{cases} U^- & \text{if } x < 0, \\ U^+ & \text{if } x > 0, \end{cases} \quad (2.19)$$

for a given  $U^-, U^+ \in \mathbb{R}$ . Two types of weak solutions can be expected namely the shock waves and the rarefaction waves and we will consider both.

**Shock waves.** This occurs from the Riemann problem (2.1), (2.19) with  $U^- > U^+$ . Here, the characteristic curves intersect and a smooth solution cannot be constructed. Hence, the weak solution takes the form

$$U(t, x) = \begin{cases} U^- & \text{if } x < \lambda t, \\ U^+ & \text{if } x > \lambda t, \end{cases} \quad (2.20)$$

where the shock speed  $\lambda$  satisfies the **Rankine-Hugoniot condition**, given by the following Lemma

**Lemma 2.2.1.** *Given  $U^+, U^- \in \Omega \subseteq \mathbb{R}$  and  $\lambda \in \mathbb{R}$ . Consider the piecewise constant function of the form as in (2.20), if the function  $U(t, x)$  described in (2.20) is a weak solution of the Riemann problem (2.1) and (2.19), then*

$$f(U^+) - f(U^-) = \lambda(U^+ - U^-), \quad (2.21)$$

*is satisfied and it is called the Rankine-Hugoniot condition.*

(See [78] for proof of the lemma). From the above lemma, when the Rankine-Hugoniot condition (2.21) holds, the solution (2.20) is called the shock wave solution. Thus, equation (2.20) shows that the characteristic curves flow into the shock.

**Definition 2.2.4.** Let  $U^-$  and  $U^+$  be two states separated by a shock moving at the speed  $\lambda$ , given by the Rankine-Hugoniot condition (2.21). The characteristics for the conservation law (2.1) with convex flux function  $f(U)$  flow into the shock if the solution satisfies the following condition

$$f'(U^-) > \lambda > f'(U^+), \quad (2.22)$$

is called **Lax entropy condition**. Where  $f'(U^-)$  represents the characteristic speed at the upstream condition,  $f'(U^+)$  is the characteristic speed downstream condition.

Moreover, the function (2.20) is prescribed by the initial condition and satisfies the Lax entropy condition (2.22).

**Rarefaction waves.** This occurs from the Riemann problem (2.1), (2.19) with  $U^- < U^+$ . In this case, the weak solution (2.20) does not satisfy the Lax entropy condition (2.22) and there exists a continuous solution. To construct a continuous solution to (2.1), we note that replacing  $x, t$  by  $\lambda_x, \lambda_t$  keeps the equation invariant in the sense that a solution of one solves the other. More precisely, we consider self-similar solutions that only depend on the ratio  $\frac{x}{t}$ . Given

$$U(t, x) = \tilde{U}\left(\frac{x}{t}\right), \quad (2.23)$$

define the similarity variable  $\zeta = \frac{x}{t}$ . Thus, we have

$$\begin{aligned} \frac{\partial \tilde{U}}{\partial t} + \frac{\partial f(\tilde{U})}{\partial x} &= \frac{\partial \tilde{U}(\zeta)}{\partial t} + f'(\tilde{U}(\zeta)) \frac{\partial \tilde{U}(\zeta)}{\partial x} \\ &= \frac{\partial \tilde{U}}{\partial \zeta} \frac{\partial \zeta}{\partial t} + f'(\tilde{U}(\zeta)) \frac{\partial \tilde{U}}{\partial \zeta} \frac{\partial \zeta}{\partial x} \\ &= -\frac{x}{t^2} \frac{\partial \tilde{U}}{\partial \zeta} + f'(\tilde{U}(\zeta)) \frac{\partial \tilde{U}}{\partial \zeta} \cdot \frac{1}{t} \\ &= 0 \end{aligned}$$

or

$$\left(f'(\tilde{U}(\zeta)) - \frac{x}{t}\right) \frac{\partial \tilde{U}}{\partial \zeta} = 0.$$

In the nontrivial case of  $\frac{\partial \tilde{U}}{\partial \zeta} \neq 0$ , the above identity and the fact that  $f'$  is strictly increasing leads to the expression

$$\tilde{U}\left(\frac{x}{t}\right) = (f')^{-1}\left(\frac{x}{t}\right). \quad (2.24)$$

A self-similar solution in the form (2.24) is called a rarefaction wave.

Now, we can use equation (2.24) to construct the solution in the form

$$U(t, x) = \begin{cases} U^- & \text{if } x \leq f'(U^-)t, \\ (f')^{-1}\left(\frac{x}{t}\right) & \text{if } f'(U^-)t \leq x \leq f'(U^+)t, \\ U^+ & \text{if } x \geq f'(U^+)t. \end{cases} \quad (2.25)$$

Clearly, the solution (2.25) is a weak solution that satisfies the Lax entropy condition (2.22).

**Example 2.2.1.** Assume that the Burger's equation (2.13) with flux function  $f(U) = \frac{U^2}{2}$ , and the piecewise initial condition

$$U(0, x) = \bar{U}(x) = \begin{cases} U^- & \text{if } x < 0, \\ U^+ & \text{if } x > 0, \end{cases} \quad (2.26)$$

for a given  $U^-, U^+ \in \mathbb{R}$ . By Rankine-Hugoniot equations (2.21), we can find the shock speed in the form

$$\lambda(U) = \frac{1}{2}(U^+ + U^-).$$

Moreover, the weak solutions of the Riemann problem (2.13), (2.26) takes the form

$$U(t, x) = \begin{cases} U^- & \text{if } x < \frac{1}{2}(U^+ + U^-)t, \\ U^+ & \text{if } x > \frac{1}{2}(U^+ + U^-)t, \end{cases} \quad \forall U^- > U^+. \quad (2.27)$$

We use a discontinuous solution to derive some conditions that must be satisfied at the jump points in the following subsections.

## 2.2.2 Admissibility conditions for the weak solutions

Consider the Burger's equation as in (2.13) with initial data

$$U(0, x) = \begin{cases} 1 & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases} \quad (2.28)$$

For every constant  $\mu$ , where  $0 < \mu < 1$ , the weak solution is

$$U_\mu(t, x) = \begin{cases} 0 & \text{if } x \leq \frac{\mu t}{2}, \\ \mu & \text{if } \frac{\mu t}{2} < x < \frac{(1+\mu)t}{2}, \\ 1 & \text{if } x \geq \frac{(1+\mu)t}{2}. \end{cases} \quad (2.29)$$

As observed, the solution form (2.29) is a multiple weak solution. The piecewise constant function  $U_\mu$  trivially satisfies the equation outside the jumps. Moreover, the Rankine-Hugoniot conditions hold along the two lines of discontinuity  $\{x = \frac{\mu t}{2}\}$  and  $\{x = \frac{(1+\mu)t}{2}\}$ , for all  $t > 0$ .

The following conditions will be added to achieve the uniqueness of solutions and their continuous dependency on the initial data.

**I. Vanishing Viscosity.** A weak solution  $U$  of (2.1) is admissible in the vanishing viscosity sense if there is a sequence of smooth solutions  $U^\varepsilon$  to

$$\frac{\partial U^\varepsilon}{\partial t} + \frac{\partial f(U^\varepsilon)}{\partial x} = \varepsilon \frac{\partial^2 U^\varepsilon}{\partial x^2}, \quad (2.30)$$

which converges to  $U$  in  $L^1_{loc}$  as  $\varepsilon \rightarrow 0^+$ . However, it is very difficult to provide priori estimates to solutions to (2.1) and characterise the corresponding limits as  $\varepsilon \rightarrow 0^+$ .

## II. Entropy conditions.

**Definition 2.2.5. (Entropy - entropy flux).** A continuously differentiable function  $\eta : \Omega \subseteq \mathbb{R} \rightarrow \mathbb{R}$  is called an entropy for the general conservation laws (2.1), with entropy flux  $q : \Omega \subseteq \mathbb{R} \rightarrow \mathbb{R}$ , if

$$\eta'(U)f'(U) = q'(U), \quad \text{for all } U \in \Omega. \quad (2.31)$$

Coupling  $\eta$  and  $q$  gives an entropy-entropy flux. Moreover, from (2.31), if  $U$  is a  $C^1$  solution of (2.1), then

$$\frac{\partial \eta(U)}{\partial t} + \frac{\partial q(U)}{\partial x} = 0. \quad (2.32)$$

**Example 2.2.2.** Consider Burger's equation (2.13) with the flux  $f(U) = \frac{U^2}{2}$ . Taking  $\eta(U) = U^3$  and  $q(U) = \frac{3U^4}{4}$ , we can possibly check that the equation (2.31) is satisfied. Therefore,  $\eta$  is an entropy and  $q$  is the corresponding entropy flux. For given  $U^- = 1$  and  $U^+ = 0$  are the left and right states, respectively, we find that the function

$$U(t, x) = \begin{cases} 1 & \text{if } x < \frac{t}{2}, \\ 0 & \text{if } x \geq \frac{t}{2}, \end{cases} \quad (2.33)$$

is a weak solution of (2.13), where  $\lambda = \frac{1}{2}$  is the speed given by the Rankine-Hugoniot condition. However, in the distribution sense, it does not satisfy (2.32). Indeed

$$\frac{1}{2} = \lambda[\eta(U^+) - \eta(U^-)] \neq q(U^+) - q(U^-) = \frac{3}{4}.$$

In general, by (2.31), we can expect to find solutions for  $n \leq 2$  only. Furthermore, we now analyse how a convex entropy behaves when a small diffusion term is present. Consider  $\eta, q \in C^2$  with  $\eta$  is a convex. Multiplying  $\eta'(U^\varepsilon)$  both sides of (2.30) on the left and using (2.31), we obtain

$$\frac{\partial}{\partial t}[\eta(U^\varepsilon)] + \frac{\partial}{\partial x}[q(U^\varepsilon)] = \varepsilon\eta'(U^\varepsilon)\frac{\partial^2 U^\varepsilon}{\partial x^2} = \varepsilon\left\{\frac{\partial^2}{\partial x^2}[\eta(U^\varepsilon)] - \eta''(U^\varepsilon)\left(\frac{\partial^2 U^\varepsilon}{\partial x^2}\right)\right\}. \quad (2.34)$$

Since  $\eta$  is a convex function, the second term on the right-hand side of (2.34) satisfies

$$\eta''(U^\varepsilon)\left(\frac{\partial^2 U^\varepsilon}{\partial x^2}\right) = \frac{\partial^2 \eta(U^\varepsilon)}{\partial (U^\varepsilon)^2} \frac{\partial^2 U^\varepsilon}{\partial x^2} \geq 0,$$

hence, its second derivative at any point  $U^\varepsilon$  is a positive semidefinite quadratic form. Multiplying (2.34) by smooth function  $\phi \geq 0$  with compact support and integrating by parts yield the form

$$\iint \left\{ \eta(U^\varepsilon) \frac{\partial \phi}{\partial t} + q(U^\varepsilon) \frac{\partial \phi}{\partial x} \right\} dx dt \geq -\varepsilon \iint \eta(U^\varepsilon) \frac{\partial^2 \phi}{\partial x^2} dx dt. \quad (2.35)$$

If  $U^\varepsilon \rightarrow U$  in  $L^1$  as  $\varepsilon \rightarrow 0$ , the previous inequality yields

$$\iint \left\{ \eta(U) \frac{\partial \phi}{\partial t} + q(U) \frac{\partial \phi}{\partial x} \right\} dx dt \geq 0, \quad (2.36)$$

where  $\phi \in C_c^1, \phi \geq 0$ . The previous analysis implies the following definition

**Definition 2.2.6. (Entropy inequality).** A weak solution  $U$  of (2.1) is entropy admissible if

$$\frac{\partial \eta(U)}{\partial t} + \frac{\partial q(U)}{\partial x} \leq 0, \quad (2.37)$$

in the distributional sense, for every pair  $(\eta, q)$ , where  $\eta$  is a convex entropy for (2.1) and  $q$  is the corresponding entropy flux.

We conclude our discussion on the scalar case by presenting the existence and uniqueness of entropy solutions based on Kurzkov analysis [73].

**Definition 2.2.7. (Entropy solution).** A function  $U \in L^\infty(\mathbb{R}_+ \times \mathbb{R})$  is an entropy solution of (2.1) if satisfy the following

- $U$  is a weak solution of (2.1),
- For all entropy pair  $(\eta, q)$ ,  $U$  satisfy the condition

$$\iint \left\{ \eta(U) \frac{\partial \phi}{\partial t} + q(U) \frac{\partial \phi}{\partial x} \right\} dx dt + \int \eta(\bar{U}(x)) \phi(0, x) dx \geq 0, \quad (2.38)$$

for all  $\phi \in C_c^\infty(\mathbb{R}_+ \times \mathbb{R})$ ,  $\phi \geq 0$ . Where  $\eta(U, k) = |U - k|$  and  $q(U, k) = \text{sign}(U - k)(f(U) - f(k))$  for all  $k \in \mathbb{R}$ .

Furthermore, we denote  $BV(\mathbb{R})$  is the space of the functions of bounded variations on  $\mathbb{R}$  and a function  $u \in BV(\mathbb{R})$  iff  $u \in L^1(\mathbb{R})$  and  $TV(u) = \sup \{ \int u p' dx : p \in C_c^1(\mathbb{R}), |p| \leq 1 \text{ a.e. in } \mathbb{R} \}$ , where  $TV(u)$  is the total variation of  $u$  and *a.e.* stands for almost everywhere.

**Theorem 2.2.1.** Consider  $\bar{U} \in L^\infty(\mathbb{R})$  in (2.38), then, there exists a unique entropy solution  $U$  of (2.1) that satisfies

$$\|U(t, \cdot)\|_{L^\infty(\mathbb{R})} \leq \|\bar{U}\|_{L^\infty(\mathbb{R})}, \quad \forall t > 0.$$

Moreover, assume  $U_1$  and  $U_2$  are two entropy solutions corresponding to initial data  $\bar{U}_1, \bar{U}_2 \in L^\infty(\mathbb{R}) \cap L^1(\mathbb{R})$ , respectively, thus we have

$$\|U_1(t, \cdot) - U_2(t, \cdot)\|_{L^1(\mathbb{R})} \leq \|\bar{U}_1 - \bar{U}_2\|_{L^1(\mathbb{R})}, \quad \forall t > 0.$$

Finally, if  $\bar{U} \in BV(\mathbb{R})$ , then,  $TV(U(t, \cdot)) \leq TV(\bar{U})$  for all  $t > 0$ .

*Proof.* The proof of this theorem can be found in [73, 88] and the referenece therein. □



**Definition 2.3.1.** For every  $U$  contained in the open subset  $\Omega$  of  $\mathbb{R} \times \mathbb{R}$ , the pairs  $(\lambda_i, r_i)$  for  $i = 1, \dots, n$  are called the  $i$ -th characteristic field and the eigenvalues  $\lambda_i$  are called the  $i$ -th characteristic speed or  $i$ -wave.

(i) We say that  $(\lambda_i, r_i)$  is genuinely nonlinear if

$$\nabla_U \lambda_i(U) \cdot r_i(U) \neq 0, \quad \forall U \in \Omega.$$

(ii) We say that  $(\lambda_i, r_i)$  is linearly degenerate if

$$\nabla_U \lambda_i(U) \cdot r_i(U) = 0, \quad \forall U \in \Omega.$$

Here,  $\nabla \lambda_i(U), i = 1, \dots, n$  is the gradient vector obtained by differentiating the scalar  $\lambda_i(U)$  with respect to each component of the vector  $U$ .

**Example 2.3.1. (Euler equations of gas dynamics).** Consider the one-dimensional (1D) system of Euler equations of gas dynamics

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho u) &= 0, \\ \frac{\partial}{\partial t}(\rho u) + \frac{\partial}{\partial x}(\rho u^2 + P) &= 0, \\ \frac{\partial E}{\partial t} + \frac{\partial}{\partial x}(u(E + P)) &= 0, \end{aligned} \tag{2.43}$$

where  $\rho$  represents density,  $u$  is the velocity,  $E$  is the total energy per unit volume and  $P$  is the pressure. With  $\gamma$  is a constant of gas (the ratio of specific heats), for example, we could have the constitutive relationship (For calorically ideal gas)

$$E = \frac{P}{\gamma - 1} + \frac{1}{2}\rho u^2.$$

The pressure  $P$  is related to the internal energy  $e$  by the caloric equation of state  $P = P(\rho, e)$  so that

$$P = (\gamma - 1)\left[E - \frac{1}{2}\rho u^2\right].$$



We can rewrite the system (2.43) in the conservative form (2.1) where  $U = (U_1, U_2, U_3)$  and  $f(U) = (f_1(U), f_2(U), f_3(U))$  as

$$U = \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix} \equiv \begin{bmatrix} \rho \\ \rho u \\ E \end{bmatrix}, \quad f(U) = \begin{bmatrix} U_2 \\ \frac{1}{2}(3-\gamma)\frac{U_2^2}{U_1} + (\gamma-1)U_3 \\ \gamma\frac{U_2U_3}{U_1} - \frac{1}{2}(\gamma-1)\frac{U_2^3}{U_1^2} \end{bmatrix} \equiv \begin{bmatrix} \rho u \\ \rho u^2 + P \\ u(E+P) \end{bmatrix}. \quad (2.44)$$

Therefore, we can write (2.43) in the quasilinear form (2.41) with the  $3 \times 3$  Jacobian matrix  $A(U)$  computed as

$$A(U) = f'(U) \equiv \begin{bmatrix} 0 & 1 & 0 \\ -\frac{1}{2}(\gamma-3)\left(\frac{U_2}{U_1}\right)^2 & (3-\gamma)\frac{U_2}{U_1} & \gamma-1 \\ -\gamma\frac{U_2U_3}{U_1^2} + \frac{1}{2}(\gamma-1)\left(\frac{U_2}{U_1}\right)^3 & \gamma\frac{U_3}{U_1} - \frac{3}{2}(\gamma-1)\left(\frac{U_2}{U_1}\right)^2 & \gamma\frac{U_2}{U_1} \end{bmatrix} \quad (2.45)$$

$$= \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2}(\gamma-3)u^2 & (3-\gamma)u & \gamma-1 \\ (\gamma-1)u^2 - \frac{\gamma u E}{\rho} & \frac{\gamma E}{\rho} - \frac{3}{2}(\gamma-1)u^2 & \gamma u \end{bmatrix}.$$

The eigenvalues of the Jacobian matrix (2.45) are

$$\lambda_1(U) = u - \sqrt{\frac{\gamma P}{\rho}}, \quad \lambda_2(U) = u, \quad \lambda_3(U) = u + \sqrt{\frac{\gamma P}{\rho}}. \quad (2.46)$$

The eigenvalues (2.46) are real and distinct, therefore, the system (2.43) is a strictly hyperbolic system with eigenvectors

$$r_1(U) = \begin{bmatrix} 1 \\ u - a \\ H - ua \end{bmatrix}, \quad r_2(U) = \begin{bmatrix} 1 \\ u \\ \frac{1}{2}u^2 \end{bmatrix}, \quad r_3(U) = \begin{bmatrix} 1 \\ u + a \\ H + ua \end{bmatrix}, \quad (2.47)$$

where  $a = \sqrt{\frac{\gamma P}{\rho}}$  and  $H = \frac{a^2}{\gamma - 1} + \frac{1}{2}u^2 = \frac{E + P}{\rho}$  is the total enthalpy per unit volume.

Furthermore, the characteristic fields  $(\lambda_1(U), r_1(U))$  and  $(\lambda_3(U), r_3(U))$  are both genuinely non-linear, since for all  $U$ ,

$$\nabla \lambda_1(U) \cdot r_1(U) \neq 0, \quad \nabla \lambda_3(U) \cdot r_3(U) \neq 0,$$

while the characteristic field  $(\lambda_2(U), r_2(U))$  is linearly degenerate because

$$\nabla \lambda_2(U) = \nabla(u) \equiv \nabla \left( \frac{\rho u}{\rho} \right) = \begin{bmatrix} -\frac{u}{\rho} \\ \frac{1}{\rho} \\ 0 \end{bmatrix},$$

and

$$\nabla \lambda_2(U) \cdot r_2(U) = 0.$$

In the next section, we consider two primary cases of conservation laws: linear and non-linear systems and will discuss each one in-depth.

### 2.3.1 Linear systems of conservation laws

Consider a homogeneous system with constant coefficients in the form

$$\frac{\partial U}{\partial t} + A \frac{\partial U}{\partial x} = 0, \quad U(0, x) = \bar{U}(x), \quad (2.48)$$

where  $A$  represents an  $n \times n$  hyperbolic matrix with real eigenvalues  $\lambda_1 < \dots < \lambda_n$ , and  $r_i$  and  $l_i$  are the right and left eigenvectors, respectively, so that they satisfy

$$l_i \cdot r_j = 1 \quad \text{if } i = j \quad \text{and} \quad l_i \cdot r_j = 0 \quad \text{if } i \neq j.$$

Call the coordinates of a vector  $U \in \mathbb{R}^n$ , denoted by  $U_i := L_i \cdot U$ , with respect to the basis of right eigenvectors  $\{r_1, \dots, r_n\}$ . Multiplying (2.48) by  $l_1, \dots, l_n$ , we obtain

$$\begin{aligned} l_i \frac{\partial U}{\partial t} + l_i A \frac{\partial U}{\partial x} &= \frac{\partial}{\partial t}(l_i U) + \lambda_i \frac{\partial}{\partial x}(l_i U) \\ &= \frac{\partial U_i}{\partial t} + \lambda_i \frac{\partial U_i}{\partial x} \\ &= 0, \end{aligned} \tag{2.49}$$

$$U_i(0, x) = l_i \bar{U}(x) = \bar{U}_i(x).$$

As a result, (2.49) decouples (2.48) into  $n$  scalar Cauchy problems, each of which can be solved independently as (2.5). The function description

$$U(t, x) = \sum_{i=1}^n \bar{U}_i(x - \lambda_i t) r_i, \tag{2.50}$$

provides the explicit solution to the system (2.48) since

$$\frac{\partial U}{\partial t}(t, x) = \sum_{i=1}^n -\lambda_i \left( l_i \cdot \frac{\partial}{\partial x} \bar{U}(x - \lambda_i t) \right) r_i \tag{2.51}$$

$$= -A \frac{\partial U}{\partial x}(t, x). \tag{2.52}$$

Note that the initial profile is shifted with constant speed  $\lambda$  in the scalar case (2.5). The initial profile of the system (2.48) is decomposed into a sum of  $n$  waves, each travelling by one of the characteristic speeds  $\lambda_1, \dots, \lambda_n$ .

**Example 2.3.2.** Consider the initial value problem with the one dimension 1D linear system of conservation laws of the form

$$\begin{aligned} \frac{\partial U_1}{\partial t} + \frac{\partial}{\partial x}(U_1 + 4U_2) &= 0, \\ \frac{\partial U_2}{\partial t} + \frac{\partial}{\partial x}(U_1 + U_2) &= 0, \end{aligned} \tag{2.53}$$

with the initial data

$$U(0, x) = \begin{bmatrix} x^2 \\ \sin(x) \end{bmatrix}. \tag{2.54}$$

The system (2.53) can be written in the conservative form (2.48) with vector  $U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}$  and  $2 \times 2$

matrix  $A = \begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix}$ . Furthermore, the eigenvalues of  $A$  are  $\lambda_1 = 3$  and  $\lambda_2 = -1$  and corresponding eigenvectors are

$$r_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad \text{and} \quad r_2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}.$$

The general solution to the Cauchy problem (2.53) and (2.54) can be obtained from the equation (2.50) in the form

$$U(t, x) = \begin{bmatrix} 2(x - 3t)^2 - 2\sin(x + t) \\ (x - 3t)^2 + \sin(x + t) \end{bmatrix}. \quad (2.55)$$

As a **special case of Cauchy problems**, consider the Riemann initial data given as

$$U(0, x) = \bar{U}(x) = \begin{cases} U^- & \text{if } x < 0, \\ U^+ & \text{if } x > 0. \end{cases} \quad (2.56)$$

Then, the corresponding explicit solution in (2.50) can be obtained:

Decompose  $U^+ - U^-$  with the basis of right eigenvectors of  $A$  as

$$U^+ - U^- = \sum_{j=1}^n \alpha_j r_j. \quad (2.57)$$

The intermediate states are defined as

$$\omega_i = U^- + \sum_{j \leq i} \alpha_j r_j, \quad i = 0, \dots, n, \quad (2.58)$$

so that the difference  $\omega_i - \omega_{i-1}$  is an  $i$ -eigenvector of  $A$ , therefore, the solution takes the form

$$U(t, x) = \begin{cases} \omega_0 = U^- & \text{for } x/t < \lambda_1, \\ \dots & \dots \\ \omega_i & \text{for } \lambda_i < x/t < \lambda_{i+1}, \\ \dots & \dots \\ \omega_n = U^+ & \text{for } x/t > \lambda_n. \end{cases} \quad (2.59)$$

**Example 2.3.3.** Consider the Riemann problem for the linear system (2.53) with initial piecewise data

$$U(0,x) = \bar{U}(x) = \begin{cases} \begin{bmatrix} 2 \\ 3 \end{bmatrix} & \text{if } x < 0, \\ \begin{bmatrix} 6 \\ 4 \end{bmatrix} & \text{if } x > 0. \end{cases} \quad (2.60)$$

Then, the jump between the right and left states is

$$U^+ - U^- = \begin{bmatrix} 4 \\ 1 \end{bmatrix}. \quad (2.61)$$

We can write

$$\begin{bmatrix} 4 \\ 1 \end{bmatrix} = \alpha_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} -2 \\ 1 \end{bmatrix}. \quad (2.62)$$

Solving (2.62) gives the constants  $\alpha_1 = \frac{3}{2}$  and  $\alpha_2 = -\frac{1}{2}$ .

Now, we can find intermediate states in (2.58) as

$$\begin{aligned} \omega_0 &= U^- = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \\ \omega_1 &= \omega_0 + \alpha_1 r_1 = \begin{bmatrix} 5 \\ \frac{9}{2} \end{bmatrix}, \\ \omega_2 &= \omega_1 + \alpha_2 r_2 = \begin{bmatrix} 6 \\ 4 \end{bmatrix} \equiv U^+. \end{aligned} \quad (2.63)$$

Hence, from equation (2.59), the solution of the Riemann problem (2.53) and (2.60) is

$$U(t,x) = \begin{cases} \begin{bmatrix} 2 \\ 3 \end{bmatrix} & \text{for } x/t < -1, \\ \begin{bmatrix} 5 \\ \frac{9}{2} \end{bmatrix} & \text{for } -1 < x/t < 3, \\ \begin{bmatrix} 6 \\ 4 \end{bmatrix} & \text{for } x/t > 3. \end{cases} \quad (2.64)$$

### 2.3.2 Non-linear systems of conservation laws

Consider the system of conservation laws (2.1) to be non-linear. Here,  $U : \mathbb{R} \times [0, \infty) \mapsto \mathbb{R}^n$  and  $f : \mathbb{R}^n \mapsto \mathbb{R}^n$ . We can write this system in quasilinear form (2.41) where the eigenvalues and eigenvectors of the Jacobian matrix  $A(U) = Df(U)$  depend on  $U$ . Assume that the system is strictly hyperbolic, then, the eigenvectors are linearly independent. Thus, we can find a basis for these eigenvectors by normalising using the same approach as (2.42).

In general non-linear systems, the eigenvalues depend on  $U$  ( $\lambda_i = \lambda_i(U), i = 1, \dots, n$ ). Thus, the shape of components in the solution will vary over time and the waves can interact with each other, producing different waves. We will present some concepts that are needed to construct a weak solution to the Riemann problem for non-linear systems of conservation laws such as

**Integral curves.** An integral curve is a curve of the vector field  $r_i(U)$  with a tangent vector at each point  $U$ .

**Hugoniot locus.** Given a fixed state  $\bar{U} \in \mathbb{R}^n$ , determine all possible states  $U$  which can be connected to  $\bar{U}$  by a discontinuity satisfying the Rankine-Hugoniot condition (2.21) for some  $\lambda$ . This gives a system of  $n$  curves in  $(n+1)$  unknowns through any point  $\bar{U}$ . At each  $i = 1, \dots, n$ , we can expect one parameter (characteristic) families of solutions. For any scalar multiple  $\zeta$  of the jump

of  $r_i(\bar{U})$ , we can parametrise these curves by  $U_i(\zeta; \bar{U})$  with  $U_i(0; \bar{U}) = \bar{U}$  and we let  $\lambda_i = \lambda_i(\zeta; \bar{U})$  denotes the corresponding speed. Hence

$$U_i(\zeta; \bar{U}) = U_i(\zeta), \quad \lambda_i(\zeta; \bar{U}) = \lambda_i(\zeta).$$

The Rankine-Hugoniot condition gives

$$f(U_i(\zeta)) - f(\bar{U}) = \lambda_i(\zeta)(U_i(\zeta) - \bar{U}).$$

Differentiating the above expression with respect to  $\zeta$  and setting  $\zeta = 0$  gives

$$f'(\bar{U})U_i'(0) = \lambda_i(0)U_i'(0).$$

This suggests that  $U_i'(0)$  must be a scalar multiple of the eigenvector  $r_i(\bar{U})$  while  $\lambda_i(0) = \lambda_i(\bar{U})$ . Thus, at point  $\bar{U}$ , the curve  $U_i(\zeta)$  is a tangent vector to  $r_i(\bar{U})$ . These curves are called Hugoniot curves and the set of all points on these curves is called **Hugoniot locus**. Moreover, if  $\tilde{U}_i$  lies on  $i$ -the Hugoniot curve through  $\bar{U}$ , then, we say that  $\bar{U}$  and  $\tilde{U}_i$  are connected by  $i$ -shock.

### 2.3.3 Solution to the Riemann problem

Consider a non-linear system of conservation laws (2.1) with the piecewise constant initial data (2.56). In solving a problem of this kind, we can find an intermediary state  $\omega$  such that  $U^-$  and  $\omega$  are connected by a discontinuity satisfying the Rankine-Hugoniot conditions and so are  $\omega$  and  $U^+$ . Geometrically, this can be achieved by drawing the Hugoniot locus for each state  $U^-$  and  $U^+$  and looking for intersections.

The  $i$ -th eigenvalues  $\lambda_i(U)$  is strictly increasing along each integral curve of the corresponding field of eigenvectors  $r_i(U)$  in the genuinely non-linear case. Furthermore, if  $U = U(\zeta)$  is a parametrisation of an integral curve in the  $i$ -th family for any parameter  $\zeta \in \mathbb{R}$ , then, the tangent vector is proportional to  $r_i(U)$  at each point, i.e.,

$$\frac{dU}{d\zeta} = \gamma(\zeta)r_i(U(\zeta)),$$

where  $\gamma(\zeta)$  is some scalar factor. Since  $r_i(U(\zeta))$  is a smooth function of  $U$ , thus

$$\frac{d}{d\zeta}(\lambda_i(U)) = \nabla\lambda_i(U) \cdot \frac{dU}{d\zeta} = \nabla\lambda_i(U) \cdot r_i(U(\zeta)) > 0,$$

where equation (2.42) gives  $\gamma(\zeta) = 1$  in this case.

However, the eigenvalue  $\lambda_i$  is constant along each such curve in the linearly degenerate case. Consequently, a solution to the Riemann problem is either the simple wave (a rarefaction, shock, contact discontinuity) or a combination of these simple waves.

Given a fixed state  $U_0 \in \mathbb{R}^n$  and an  $i$ -th eigenvector  $r_i(U)$  of the Jacobian matrix  $A(U) = Df(U)$ . The integral curve of vector field  $r_i$  through the point  $U_0$  is a so-called  $i$ -rarefaction curve, which can be obtained by solving the Cauchy problem in state space

$$\frac{dU}{d\zeta} = r_i(U), \quad U(0) = U_0, \quad (2.65)$$

this corresponding to a curve given as

$$\zeta \mapsto R_i(\zeta; U_0). \quad (2.66)$$

Moreover, parametrisation depends on the choice of  $i$ -th eigenvectors  $r_i$ .

Next, consider the state  $U$ , which can be connected to the right of  $U_0$  by  $i$ -shock, satisfying Rankine-Hugoniot equations

$$\lambda(U - U_0) = f(U) - f(U_0). \quad (2.67)$$

Therefore, for a given convex matrix  $A$ , the equation (2.67) can be written in the form

$$\lambda(U - U_0) = A(U, U_0)(U - U_0). \quad (2.68)$$

Equation (2.68) is satisfied if  $U - U_0$  is orthogonal to each left  $j$ -eigenvector of  $A(U, U_0)$  for  $j \neq i$ ,  $j \in 1, \dots, n$ . Hence, conditions (2.68) can be written in the equivalent form as

$$l_j(U, U_0) \cdot (U - U_0) = 0, \quad \text{with } \lambda = \lambda_i(U, U_0), \quad \forall j \neq i. \quad (2.69)$$

By linearising around the point  $U = U_0$ , we have

$$l_j(U, U_0) \cdot (\omega - U_0) = 0, \quad \forall j \neq i.$$

The above equation is a linear system whose solutions are all points  $\omega = U_0 + cr_i(U_0)$ ,  $c \in \mathbb{R}$ . Furthermore, by the implicit function theorem, the set of solutions of (2.68) is a smooth curve



and tangent to the vector  $r_i$  at point  $U_0$ , and it is called the  $i$ -shock curve at point  $U_0$  and can be parametrised as

$$\zeta \longmapsto S_i(\zeta; U_0). \quad (2.70)$$

In constructing the solution to the Riemann problem in all the problems considered above, three cases in the solution process can be introduced.

**Case I.** Consider the Riemann problem with  $U^- < U^+$ . Given the  $i$ -th characteristic field is genuinely non-linear, and for some  $\zeta > 0$ ,  $U^+ = R_i(\zeta; U^-)$ . For each  $s \in [0, \zeta]$ , the characteristic speed is

$$\lambda_i(s) = \lambda_i(R_i(s; U^-)).$$

Therefore, there is a unique value  $s$ , for every  $\lambda \in [\lambda_i(U^-), \lambda_i(U^+)]$ . Such that  $\lambda = \lambda_i(s)$ , then, the piecewise smooth function

$$U(t, x) = \begin{cases} U^- & \text{if } \frac{x}{t} < \lambda_i(U^-), \\ R_i(s; U^-) & \text{if } \lambda_i(U^-) \leq \frac{x}{t} \leq \lambda_i(U^+), \forall t \geq 0, \\ U^+ & \text{if } \frac{x}{t} > \lambda_i(U^+), \end{cases} \quad (2.71)$$

is the weak solution of (2.1) and (2.56) and is called a centred rarefaction wave (a rarefaction wave) which a self-similar solution such that  $U(t, x) = U(x/t)$ .

**Case II.** Consider the Riemann problem with  $U^- > U^+$ . Assume that  $U^+ = S_i(\zeta; U^-)$ , for a given  $i$ -th characteristic field, which is genuinely non-linear and the Rankine-Hugoniot speed of the shock  $\lambda = \lambda_i(U^-, U^+)$ . Then, the piecewise function (2.20) is a weak solution known as a shock wave. If  $\zeta < 0$ , then, this solution is entropy admissible in the Lax sense. Indeed, the speed is monotonically increasing along the shock curve. Precisely,

$$\lambda_i(U^+) < \lambda_i(U^-, U^+) < \lambda_i(U^-). \quad (2.72)$$

**Case III.** When the  $i$ -th characteristic field is linearly degenerate and  $U^+ = R_i(\zeta; U^-)$  for some  $\zeta$ . Then, the  $i$ -th characteristic speed  $\lambda_i$  is constant along the curve. Therefore, the  $i$ -th shock and  $i$ -th rarefaction curves coincide, so that

$$S_i(\zeta; U_0) = R_i(\zeta; U_0), \quad \forall U_0 \text{ and } \zeta.$$

The resulting curve is called a contact discontinuity curve and the corresponding solution is called a contact discontinuity wave. Moreover, suppose the shock curves and the rarefaction curves satisfy the Lax entropy condition, then, the map

$$\Phi_i(\zeta, U^-) = \begin{cases} S_i(\zeta; U^-) & \text{if } \zeta < 0, \\ R_i(\zeta; U^-) & \text{if } \zeta \geq 0, \end{cases} \quad (2.73)$$

is the  $i$ -th Lax curve through  $U^-$ , which is smooth for  $\zeta \neq 0$  and twice continuously differentiable for  $\zeta = 0$ . Therefore, if  $U^+ = \Phi_i(\zeta, U^-)$  for some  $\zeta$ , the Riemann problem can be solved using an elementary wave: rarefaction, shock or contact discontinuity.

The shock admissibility conditions discussed above are inadequate to guarantee the uniqueness of the Riemann problem solution. Therefore, additional conditions can be reported here such as the Lax condition.

**Lemma 2.3.1. Lax admissibility condition.** *Given right and left states,  $U^+$  and  $U^-$  respectively and speed for the shock in the  $i$ -th characteristic field denoted by  $\lambda = \lambda_i(U^-, U^+)$  of the jump for all  $i \in \{1, \dots, n\}$ . A weak solution  $U = U(t, x)$  of (2.1) satisfies the Lax admissibility condition, if  $U$  at each point  $(\tau, \xi)$  of approximate jump satisfies*

$$\lambda_i(U^-) \geq \lambda \geq \lambda_i(U^+). \quad (2.74)$$

Geometrically, consider a piecewise smooth solution having a discontinuity along the line  $x = \gamma(t)$ , which jumps from a left state  $U^-$  to a right state  $U^+$ . Following equation (2.56), the discontinuity must travel with a speed  $\lambda = \dot{\gamma} = \lambda_i(U^-, U^+)$  equal to  $i$ -eigenvalues of the averaged matrix  $A(U^-, U^+)$ . Moreover, the Lax condition requires that  $i$ -th characteristics run into the shock from both sides.

Finally, by finding intermediate states  $U^- = \omega_0, \omega_1, \dots, \omega_n = U^+$ , the general solution to the Riemann problem can be obtained as a juxtaposition of the fixed states. Each pair  $(\omega_{i-1}, \omega_i)$ ,  $i = 1, \dots, n$  of states can be connected by an elementary wave, i. e.

$$\omega_i = \Phi_i(\zeta; \omega_{i-1}). \quad (2.75)$$

By piecing together the solutions to the  $n$  Riemann problems, the complete solution is obtained

$$U(t,x) = \begin{cases} \omega_{i-1} & \text{if } x < 0, \\ \omega_i & \text{if } x > 0, \end{cases} \quad (2.76)$$

on different sectors of the  $xt$  - plane. Furthermore, each of the problems has an entropy-admissible solution comprising a simple wave of the  $i$ -th characteristic field. We can assume that the intervals  $[\lambda_i^-, \lambda_i^+]$  are disjoint based on strict hyperbolicity and continuity where  $\lambda_i^- = \lambda_i(\omega_{i-1})$  and  $\lambda_i^+ = \lambda_i(\omega_i)$  meaning

$$\lambda_1^- \leq \lambda_1^+ < \lambda_2^- \leq \lambda_2^+ < \dots < \lambda_n^- \leq \lambda_n^+.$$

Hence, the piecewise solution  $U : [0, \infty[ \times \mathbb{R} \mapsto \mathbb{R}^n$  has the form

$$U(t,x) = \begin{cases} U^- = \omega_0 & \text{if } \frac{x}{t} < \lambda_1^-, \\ R_i(s; \omega_{i-1}) & \text{if } \lambda_i^- \leq \frac{x}{t} < \lambda_i^+, \\ \omega_i & \text{if } \lambda_i^+ \leq \frac{x}{t} < \lambda_{i+1}^-, \\ U^+ = \omega_n & \text{if } \frac{x}{t} \geq \lambda_n^+. \end{cases} \quad (2.77)$$

**Example 2.3.4. (Shallow water equations).** Consider one-dimensional system shallow water equations

$$\begin{aligned} \frac{\partial h}{\partial t} + \frac{\partial}{\partial x}(hu) &= 0, \\ \frac{\partial}{\partial t}(hu) + \frac{\partial}{\partial x}\left(hu^2 + \frac{1}{2}gh^2\right) &= 0, \end{aligned} \quad (2.78)$$

where  $h = h(t,x)$  is the water height (depth),  $g$  is the gravitational constant and  $u = u(t,x)$  is the water velocity (see Figure 2.2).

Equations (2.78) can be written in conservative form as

$$\frac{\partial}{\partial t} \begin{bmatrix} h \\ hu \end{bmatrix} + \frac{\partial}{\partial x} \begin{bmatrix} hu \\ hu^2 + \frac{1}{2}gh^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad (2.79)$$

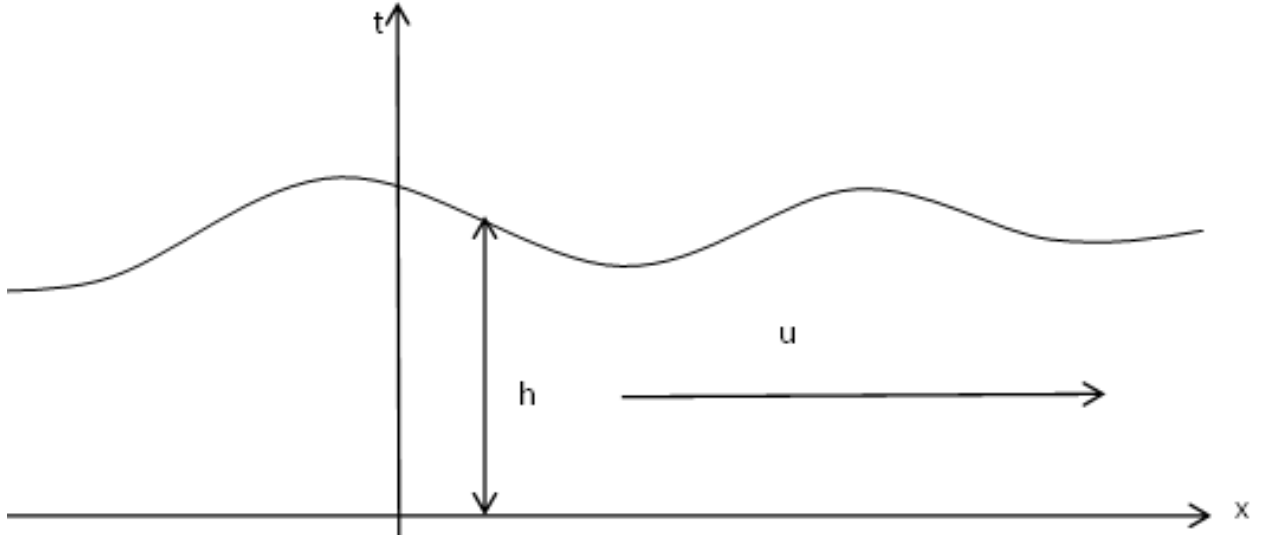


Figure 2.2: The height and velocity of water in a shallow channel.

with notations

$$U(t,x) = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \equiv \begin{bmatrix} h \\ hu \end{bmatrix}, \quad f(U) = \begin{bmatrix} U_2 \\ \frac{U_2^2}{U_1} + \frac{1}{2}gU_1^2 \end{bmatrix} \equiv \begin{bmatrix} hu \\ hu^2 + \frac{1}{2}gh^2 \end{bmatrix}.$$

This becomes

$$U_t + f(U)_x = 0.$$

We obtain the Jacobian matrix of the flux function and given as

$$A(U) = f'(U) \equiv \begin{bmatrix} 0 & 1 \\ -(\frac{U_2}{U_1})^2 + gU_1 & \frac{2U_2}{U_1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ gh - u^2 & 2u \end{bmatrix}. \quad (2.80)$$

Its eigenvalues are

$$\lambda_1(U) = u - \sqrt{gh}, \quad \lambda_2(U) = u + \sqrt{gh}. \quad (2.81)$$

If  $h > 0$ , then, the eigenvalues in (2.81) are real and distinct. Thus, the system (2.78) is strictly

hyperbolic and the corresponding eigenvectors are found as

$$r_1(U) = \begin{bmatrix} 1 \\ u - \sqrt{gh} \end{bmatrix}, \quad r_2(U) = \begin{bmatrix} 1 \\ u + \sqrt{gh} \end{bmatrix}. \quad (2.82)$$

Characteristic fields  $(\lambda_1(U), r_1(U))$  and  $(\lambda_2(U), r_2(U))$  are both genuinely nonlinear since

$$\nabla \lambda_1(U) \cdot r_1(U) = \frac{-3}{2} \sqrt{\frac{g}{h}} \neq 0, \quad (2.83)$$

and

$$\nabla \lambda_2(U) \cdot r_2(U) = \frac{3}{2} \sqrt{\frac{g}{h}} \neq 0. \quad (2.84)$$

To solve the Riemann problem, we must first construct the rarefaction and shock curves.

**Remark. Rarefaction curves.** For a given fixed point  $\bar{U} = (\bar{h}, \bar{q})^T$ , where  $\bar{q} = \bar{h}\bar{u}$ . Rarefaction curves are found as integral curves of the eigenvectors. Hence, we will reconstruct the eigenvectors by normalising  $\frac{d\tilde{U}}{d\xi} = \frac{r_i(U)}{\nabla \lambda_i(U) \cdot r_i(U)}, i = 1, 2$ , which give

$$\frac{d\tilde{U}}{d\xi} = \mp \frac{2}{3} \sqrt{\frac{h}{g}} \begin{bmatrix} 1 \\ \frac{q}{h} \mp \sqrt{gh} \end{bmatrix}. \quad (2.85)$$

We can choose (2.85) with the minus sign, for example, we have

$$\frac{dh}{d\xi} = -\frac{2}{3} \sqrt{\frac{h}{g}}, \quad (2.86a)$$

$$\frac{dq}{d\xi} = -\frac{2}{3} \sqrt{\frac{h}{g}} \left( \frac{q}{h} - \sqrt{gh} \right) = -\frac{2}{3} \frac{q}{\sqrt{gh}} + \frac{2}{3} h. \quad (2.86b)$$

Solving equation (2.86a) with  $h(0) = \bar{h}$  gives

$$h = \left( \sqrt{\bar{h}} - \frac{\xi}{3\sqrt{g}} \right)^2. \quad (2.87)$$

Substituting the value of  $h$  given by (2.87) in equation (2.86b) gives

$$\frac{dq}{d\xi} = -\frac{2}{3\sqrt{g}} \left( \frac{q}{\sqrt{\bar{h} - \frac{\xi}{3\sqrt{g}}}} \right) + \frac{2}{3} \left( \sqrt{\bar{h}} - \frac{\xi}{3\sqrt{g}} \right)^2. \quad (2.88)$$

Solving the differential equation (2.88) together with the initial condition  $q(0) = \bar{q}$  gives

$$q = \frac{\bar{q}}{h}\xi - 2\xi(\sqrt{g\bar{h}} - \sqrt{g\xi}). \quad (2.89)$$

Since  $\xi$  is representing the depth  $h$  in this curve so that gives the integral curve of  $r_1(U)$  corresponding to the eigenvalue  $\lambda_1(U)$  in terms of momentum, hence

$$q = \frac{\bar{q}}{h}h + 2h(\sqrt{g\bar{h}} - \sqrt{gh}). \quad (2.90)$$

We can rewrite equation (2.90) in terms of velocity by substituting  $q = hu$  and  $\bar{q} = \bar{h}\bar{u}$ , we get

$$u = \bar{u} + 2(\sqrt{g\bar{h}} - \sqrt{gh}). \quad (2.91)$$

Similarly, the integral curve of  $r_2(U)$  passing the point  $(\bar{h}, \bar{q})$  can be written as

$$q = \frac{\bar{q}}{h}h - 2h(\sqrt{g\bar{h}} - \sqrt{gh}), \quad (2.92)$$

and

$$u = \bar{u} - 2(\sqrt{g\bar{h}} - \sqrt{gh}). \quad (2.93)$$

For a centred rarefaction wave, a particular parameterisation of the integral curve is required since  $\xi = \frac{x}{t}$ . Rewriting  $x = \xi t$ , we can see that the value  $U(\xi)$  observed along the ray  $\frac{x}{t} = \xi$  is propagating at speed  $\xi$ , implying that  $\xi$  at every point on the integral curve must be equal to the characteristic speed  $\lambda_i(U(\xi))$ .

Furthermore, equation (2.91) describes an integral curve of  $r_1$  where  $(\bar{h}, \bar{u})$  is an arbitrary point on the curve. This can be expressed as

$$u + 2\sqrt{gh} = \bar{u} + 2\sqrt{g\bar{h}}.$$

Since  $(\bar{h}, \bar{u})$  and  $(h, u)$  can be any two points on the curve, the function

$$R_1(U) = u + 2\sqrt{gh},$$

has a similar value at all points on this curve. This function is called a Riemann invariant for the 1-family ( 1-Riemann invariant). It is a function of  $U$ , whose values are invariant along with every

integral curve of  $r_1(U)$  though it will take a different value on a different integral curve.

In the same way, equation (2.93) yields

$$R_2(U) = u - 2\sqrt{gh},$$

which is a 2-Riemann invariant, which is a function whose value is constant along any integral curve of  $r_2(U)$ .

**Remark. Shock curves.** The shock curve connecting the state  $U$  with the state  $\bar{U}$  satisfies the Rankine - Hugoniot conditions given by

$$\begin{aligned} s(\bar{h} - h) &= \bar{q} - q, \\ s(\bar{q} - q) &= \left(\frac{\bar{q}^2}{\bar{h}} + \frac{1}{2}g\bar{h}^2\right) - \left(\frac{q^2}{h} + \frac{1}{2}gh^2\right). \end{aligned} \quad (2.94)$$

We eliminate the shock speed  $s$  in the equations (2.94) to obtain

$$\frac{\bar{h} - h}{\bar{q} - q} = \frac{\bar{q} - q}{\left(\frac{\bar{q}^2}{\bar{h}} - \frac{q^2}{h}\right) + \frac{g}{2}(\bar{h}^2 - h^2)}, \quad (2.95)$$

this implies that

$$(\bar{q} - q)^2 = (\bar{h} - h) \left[ \left(\frac{\bar{q}^2}{\bar{h}} - \frac{q^2}{h}\right) + \frac{g}{2}(\bar{h}^2 - h^2) \right], \quad (2.96)$$

or

$$(\bar{h}\bar{u} - hu)^2 = (\bar{h} - h) \left[ (\bar{h}\bar{u}^2 - hu^2) + \frac{g}{2}(\bar{h}^2 - h^2) \right]. \quad (2.97)$$

Expanding and simplifying, we get a quadratic equation in  $u$  as

$$u^2 - 2\bar{u}u + \left[ \bar{u}^2 - \frac{g}{2}\left(\frac{1}{\bar{h}} + \frac{1}{h}\right)(\bar{h} - h)^2 \right] = 0, \quad (2.98)$$

which has the solution

$$u = \bar{u} \pm (\bar{h} - h) \sqrt{\frac{g}{2}\left(\frac{1}{\bar{h}} + \frac{1}{h}\right)}, \quad (2.99)$$

or alternatively,

$$q = hu = \frac{\bar{q}}{h}h \pm h(\bar{h} - h) \sqrt{\frac{g}{2}\left(\frac{1}{\bar{h}} + \frac{1}{h}\right)}. \quad (2.100)$$

Furthermore, substituting equation (2.100) into equation (2.94) to get a formula for the corresponding shock speeds:

$$s(\bar{h} - h) = \bar{q} - q,$$

or

$$s = \frac{\bar{q} - q}{\bar{h} - h},$$

which implies

$$s = \frac{\bar{h}\bar{u} - \bar{h}u + \bar{h}u - hu}{\bar{h} - h} = \frac{(\bar{u} - u)\bar{h} + u(\bar{h} - h)}{\bar{h} - h},$$

and then we get the shock speed

$$s = u \pm \sqrt{\frac{g}{2} \left( \frac{1}{\bar{h}} + \frac{1}{h} \right)}, \quad (2.101)$$

or

$$s = \bar{u} \pm h \sqrt{\frac{g}{2} \left( \frac{1}{\bar{h}} + \frac{1}{h} \right)}. \quad (2.102)$$

When  $h = \bar{h}$ , this implies that  $s = \lambda_{1,2}(U)$ . Thus, we can indicate that the wave families are invariant with speeds given as

$$s_1 \equiv s_1(h, \bar{u}) = \bar{u} - h \sqrt{\frac{g}{2} \left( \frac{1}{\bar{h}} + \frac{1}{h} \right)} = u - \bar{h} \sqrt{\frac{g}{2} \left( \frac{1}{\bar{h}} + \frac{1}{h} \right)}, \quad (2.103)$$

and

$$s_2 \equiv s_2(h, \bar{u}) = \bar{u} + h \sqrt{\frac{g}{2} \left( \frac{1}{\bar{h}} + \frac{1}{h} \right)} = u + \bar{h} \sqrt{\frac{g}{2} \left( \frac{1}{\bar{h}} + \frac{1}{h} \right)}. \quad (2.104)$$

Finally, the shock curves take on the following form

$$S_1(\bar{U}) := \left\{ \left[ \begin{array}{c} h \\ \left[ \frac{\bar{q}}{h} h + h(\bar{h} - h) \sqrt{\frac{g}{2} \left( \frac{1}{\bar{h}} + \frac{1}{h} \right)} \right] \end{array} \right] : h > 0 \right\}, \quad (2.105)$$

and

$$S_2(\bar{U}) := \left\{ \left[ \begin{array}{c} h \\ \left[ \frac{\bar{q}}{h} h - h(\bar{h} - h) \sqrt{\frac{g}{2} \left( \frac{1}{\bar{h}} + \frac{1}{h} \right)} \right] \end{array} \right] : h > 0 \right\}, \quad (2.106)$$



are the corresponding shocks as slow shocks (1- shocks) and fast shocks (2- shocks) waves, respectively. The 2-shock wave is admissible while the 1-shock wave failed to satisfy the entropy condition (2.72). Furthermore, The correct solution to this problem consists of 1-rarefaction and 2-shock waves. More precisely, we can consider the dam break problem to achieve this point.

**Example 2.3.5. Dam break problem.** Consider the shallow water equations (2.78) with piecewise-constant initial data

$$h(0,x) = \begin{cases} h^- & \text{if } x < 0, \\ h^+ & \text{if } x > 0, \end{cases} \quad (2.107)$$

where  $h^- > h^+ \geq 0$  and the velocity given by  $u(0,x) = 0$ . Here, a dam that separates two regions of water depths bursts at time  $t = 0$ . Since the initial water depth is given by equation (2.107), the solution should always consist of a one-rarefaction wave and one-shock wave in which the fluid is accelerated smoothly through the rarefaction wave and abruptly through the shock.

To construct a solution to the dam-break problem (2.78) and (2.107) that comprises a 1-rarefaction and a 2-shock, an intermediate state  $U^m$  can be determined, which is connected to  $U^-$  by a 1-rarefaction wave and simultaneously is connected to  $U^+$  by a 2-shock wave. The state  $U^m$  must lie on an integral curve of  $r_1$  passing through  $U^-$ , so by (2.91), we have

$$u^m = u^- + 2(\sqrt{gh^-} - \sqrt{gh^m}). \quad (2.108)$$

Moreover, according to (2.99),  $U^m$  must lie on the Hugoniot locus of 2-shocks moving through  $U^+$ , resulting in

$$u^m = u^+ + (h^m - h^+) \sqrt{\frac{g}{2} \left( \frac{1}{h^m} + \frac{1}{h^+} \right)}. \quad (2.109)$$

In equations (2.109) and (2.108), we can eliminate  $u^m$  to obtain a single nonlinear equation as

$$u^+ - u^- = 2(\sqrt{gh^-} - \sqrt{gh^m}) - (h^m - h^+) \sqrt{\frac{g}{2} \left( \frac{1}{h^m} + \frac{1}{h^+} \right)}, \quad (2.110)$$

thus, solving equation (2.110) to get  $h^m$ . Note that the structure of the rarefaction wave is connecting two points  $U^-$  and  $U^m$  on a single integral curve with  $\lambda_{1,2}(U^-) < \lambda_{1,2}(U^m)$ . This connection

is needed to spread out characteristics as time advances and then, rarefaction wave makes physical sense. However, given values of  $U^-$  and  $U^+$ , we might have any combination of shocks and rarefactions depending on the specific data. Thus, in general, to find  $U^m$ , the following functions  $\Phi_1^-, \Phi_1^+, \Phi_2^-$  and  $\Phi_2^+$ , respectively, can be defined as

$$\begin{aligned}
\Phi_1^-(h, U^-) &= \begin{cases} u^- - 2(\sqrt{gh^-} - \sqrt{gh}) & \text{if } h \geq h^-, \\ u^- - (h - h^-)\sqrt{\frac{g}{2}\left(\frac{1}{h} - \frac{1}{h^-}\right)} & \text{if } h < h^-; \end{cases} \\
\Phi_1^+(h, U^+) &= \begin{cases} u^+ - 2(\sqrt{gh^+} - \sqrt{gh}) & \text{if } h \leq h^+, \\ u^+ - (h - h^+)\sqrt{\frac{g}{2}\left(\frac{1}{h} - \frac{1}{h^+}\right)} & \text{if } h > h^+; \end{cases} \\
\Phi_2^-(h, U^-) &= \begin{cases} u^- + 2(\sqrt{gh^-} - \sqrt{gh}) & \text{if } h \leq h^-, \\ u^- + (h - h^-)\sqrt{\frac{g}{2}\left(\frac{1}{h} - \frac{1}{h^-}\right)} & \text{if } h > h^-; \end{cases} \\
\Phi_2^+(h, U^+) &= \begin{cases} u^+ + 2(\sqrt{gh^+} - \sqrt{gh}) & \text{if } h \geq h^+, \\ u^+ + (h - h^+)\sqrt{\frac{g}{2}\left(\frac{1}{h} - \frac{1}{h^+}\right)} & \text{if } h < h^+. \end{cases}
\end{aligned} \tag{2.111}$$

The function  $\Phi_{1,2}^-(h, U^-)$  returns the value of  $u$  that allows  $U$  to be connected to  $U^-$  by a physically correct 1-wave while  $\Phi_{1,2}^+(h, U^+)$  returns the value that allows  $U$  to be connected to  $U^+$  by a physically correct 2-wave. In addition, the system of shallow water has one intermediate state  $h^m$ , which can be found using the forward and backward Lax curves (2.111) where  $\Phi_{1,2}^-(h^m, U^-) = \Phi_{1,2}^+(h^m, U^+)$ ; to do so, a non-linear root finder can be used on function  $\Phi_{1,2}(h, U) = \Phi_{1,2}^-(h, U^-) - \Phi_{1,2}^+(h, U^+)$ .

We end our discussion on the system of conservation laws by introducing the existence and uniqueness of entropy solutions in the sense of Bressan [89]. In Definition 2.2.5, a function  $\eta : \Gamma \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  is an entropy for the system of conservation laws (2.1), with entropy  $q : \Gamma \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ , if satisfies

$$D\eta(U)Df(U) = Dq(U), \quad \forall U \in \Gamma, \tag{2.112}$$

where  $D$  is a derivatives operator.

**Theorem 2.3.1.** *Let the system (2.8) be strictly hyperbolic with smooth coefficients defined on an open set  $\Gamma \subseteq \mathbb{R}^n$ . Consider that the  $i$ -th characteristic fields are either genuinely non-linear or linearly degenerate, for each  $i \in \{1, \dots, n\}$ . Then, there exists  $\delta > 0$  sufficiently small such that for every initial data  $\bar{U} \in L^1$  with*

$$TV(\bar{U}) \leq \delta,$$

*the Cauchy problem (2.8) has a weak solution  $U(t, x)$ , defined for all times  $t \geq 0$ . In addition, if the system of conservation laws admits a convex entropy  $\eta$ , then, we can obtain a solution that is  $\eta$ -admissible.*

Proof of this theorem can be obtained in [78], which is achieved by constructing a sequence of approximate solutions, say,  $U^\varepsilon$  and showing that a subsequence of  $U^\varepsilon$  converges in  $L^1_{loc}$  to a weak solution of the Cauchy problem. The construction of an approximate solution has been done in the literature following two main approaches: the Glimm scheme [16] and the front tracking approximation [20, 90]. In general, the solutions are constructed as trajectories of a semi-group.

**Theorem 2.3.2.** *Under the assumption of Theorem 2.3.1, there exist positive constants  $\delta, L, L'$ , an open set  $\mathcal{D}$  and a map  $S : [0, +\infty[ \times \mathcal{D} \rightarrow \mathcal{D}$  with the following properties*

- $\mathcal{D} \supseteq \{U \in L^1(\mathbb{R}; \mathbb{R}^n) : U(x) \in \Gamma \text{ for } L^1\text{-a.e. } x \in \mathbb{R}, \quad TV(U) < \delta\},$
- for every  $U \in \mathcal{D}, t, s \geq 0$

$$S_0 U = U, \quad S_s(S_t U) = S_{s+t} U,$$

- for every  $U, V \in \mathcal{D}, t, s \geq 0$

$$\|S_t U - S_s V\|_{L^1} \leq L \|U - V\| + L' |t - s|,$$

- if  $U \in \mathcal{D}$  is piecewise constant, then for  $t > 0$  sufficiently small,  $S_t(U)$  coincides with the juxtaposition of the weak entropy solutions to the Riemann problem centred at the points of jump of  $U$ .

Furthermore, for every  $U \in \mathcal{D}$ , the map  $t \rightarrow S_t U$  is a weak solution to the Cauchy problem (2.8).  
 If the system of conservation laws admits a convex entropy  $\eta$ , then,  $S_t U$  is also  $\eta$ -admissible.

*Proof.* We can refer to [20, 78, 88, 90, 91] and the references therein for the proof. □

## 2.4 Finite volume method

This section is a brief description of the finite volume methods for the numerical approximations of systems of conservation laws formulated as

$$\frac{\partial U}{\partial t} + \frac{\partial f(U)}{\partial x} = 0, \quad x \in [a, b], \quad t \in [0, T], \quad U(0, x) = \bar{U}(x), \quad (2.113)$$

we use the space domain  $[a, b] \in \mathbb{R}$  for simplicity but the domain is considered infinite and no boundary conditions are required.

### 2.4.1 Derivation of the method

We discretise the space domain according to  $x_i = a + i\Delta x$ ,  $i = 0, 1, \dots, N$ , where  $N$  is an integer and  $\Delta x$  is the step length as in Figure 2.3. The midpoints of the grid are defined as  $x_{i+\frac{1}{2}} = (x_i + x_{i+1})/2$  and assume that the cells (control volume) are defined as  $C_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$ .

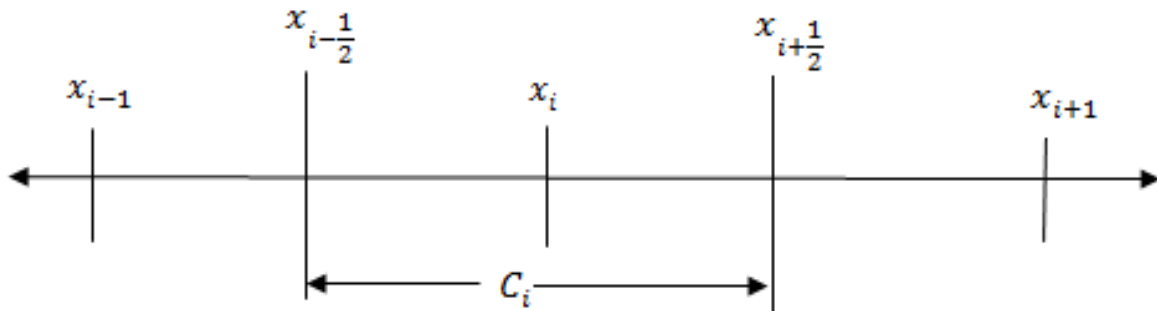


Figure 2.3: Space discretisation for the finite volume.

For any cell  $C_i$ , the method aims to approximate the volume average of the conserved variable  $U$  defined as

$$\tilde{U}(t, x_i) \equiv U_i^n = \frac{1}{\Delta x} \int_{C_i} U(t, x) dx. \quad (2.114)$$

Integrating the conservation law (2.113) over cell  $C_i$  gives

$$\int_{C_i} \left[ \frac{\partial U}{\partial t} + \frac{\partial}{\partial x} f(U) \right] dx = 0, \quad (2.115)$$

or

$$\frac{d}{dt} \int_{C_i} U(t, x) dx + \int_{C_i} \frac{\partial}{\partial x} f(U(t, x)) dx = 0.$$

Considering equation (2.114) and dividing the above equation by  $\Delta x$ , we get

$$\frac{d}{dt} \tilde{U}(t, x_i) = -\frac{1}{\Delta x} \left[ f(U(t, x_{i+\frac{1}{2}})) - f(U(t, x_{i-\frac{1}{2}})) \right]. \quad (2.116)$$

For the time discretisation, consider a temporal time domain  $[0, T]$ . We divide the domain into  $N$  points by introducing  $t_n = n\Delta t$  ( $n = 0, 1, \dots, N$ ) where  $\Delta t$  is the time step. Thus, integrating equation (2.116) over time  $t \in [t_n, t_{n+1}]$  with  $\Delta t = t_{n+1} - t_n$  gives

$$\int_{\tilde{U}_i^n}^{\tilde{U}_i^{n+1}} d\tilde{U}(t, x_i) = -\frac{1}{\Delta x} \int_{t_n}^{t_{n+1}} \left[ f(U(t, x_{i+\frac{1}{2}})) - f(U(t, x_{i-\frac{1}{2}})) \right] dt, \quad (2.117)$$

where  $\tilde{U}_i^n = \tilde{U}(t_n, x_i)$  and  $\tilde{U}_i^{n+1} = \tilde{U}(t_{n+1}, x_i)$ . This implies that

$$\tilde{U}_i^{n+1} - \tilde{U}_i^n = -\frac{1}{\Delta x} \left[ \int_{t_n}^{t_{n+1}} f(U(t, x_{i+\frac{1}{2}})) dt - \int_{t_n}^{t_{n+1}} f(U(t, x_{i-\frac{1}{2}})) dt \right]. \quad (2.118)$$

As seen in Figure 2.4, we can define  $F_{i+\frac{1}{2}}^n$  as approximations of the mean value of the flux in the  $xt$ -plane along the straight line  $x = x_{i+\frac{1}{2}}$ , with  $t$  varying between  $t_n$  and  $t_{n+1}$ .

$$F_{i+\frac{1}{2}}^n = \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} f(U(t, x_{i+\frac{1}{2}})) dt. \quad (2.119)$$

Equations (2.119) and (2.118) give

$$\tilde{U}(t_{n+1}, x_i) - \tilde{U}(t_n, x_i) = -\frac{\Delta t}{\Delta x} \left[ F_{i+\frac{1}{2}}^n - F_{i-\frac{1}{2}}^n \right]. \quad (2.120)$$

However, we cannot exactly evaluate the time integrals of the right-hand-side of (2.119), in general, since  $U(t, x_{i+\frac{1}{2}})$  will change with time along each side of the cell and do not have the exact solution to work with. Equations (2.114) and (2.120) give numerical methods of the conservative form

$$U_i^{n+1} = U_i^n - \frac{\Delta t}{\Delta x} \left[ F_{i+\frac{1}{2}}^n - F_{i-\frac{1}{2}}^n \right]. \quad (2.121)$$

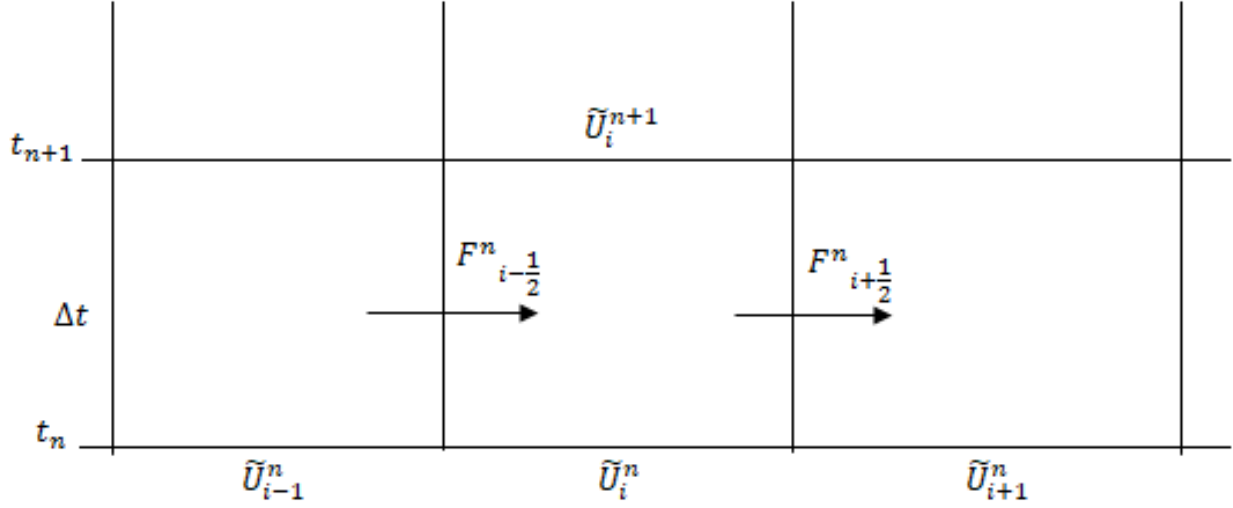


Figure 2.4: Numerical flux across cell boundaries.

Equation (2.121) is called a finite volume scheme. A complete discretisation of the Cauchy problem can be obtained if we can approximate the mean values of the flux at the cell boundaries, using the values of  $U^n(x)$  and conservation laws. For any  $U_i^n$  and  $U_{i+1}^n$ , we have the following expression

$$F_{i+\frac{1}{2}}^n = \phi(U_i^n, U_{i+1}^n). \quad (2.122)$$

**Definition 2.4.1.** We say that the scheme (2.121) is consistent with (2.113) if  $U(t, x) = \tilde{U}$  does not change over time and the numerical flux satisfies

$$\phi(\tilde{U}, \tilde{U}) = f(\tilde{U}), \quad (2.123)$$

for any value  $\tilde{U}$ .

Obviously, the above definition guarantees that if for all  $i$ ,  $U_i^n = \tilde{U}$  a constant, and therefore  $U_i^{n+1} = \tilde{U}$ . Using equation (2.122), the numerical scheme (2.121) becomes

$$U_i^{n+1} = U_i^n - \frac{\Delta t}{\Delta x} [\phi(U_i^n, U_{i+1}^n) - \phi(U_{i-1}^n, U_i^n)], \quad (2.124)$$

where  $\phi$  is the numerical flux.

The numerical method obtained above depends on the choice of  $\phi$ . However, the method is a three-point-stencil explicit method, meaning that the value of  $U_i^{n+1}$  will depend on the three values  $U_{i-1}^n$ ,  $U_i^n$  and  $U_{i+1}^n$  from the previous time step. Because data propagates at a finite velocity, the choice of time step used to be done very subtly using the notion of CFL condition.

*Remark. CFL condition.* For the finite volume method or other numerical methods to be stable and converge to conservation laws as the grid is refined, the CFL condition is required. In general, this condition is necessary but not sufficient to guarantee stability. The CFL number, also recognised as the courant number, is defined in terms of the eigenvalues  $\lambda_1, \dots, \lambda_n$  of the Jacobian matrix  $A(U) = Df(U)$  of the flux function as

$$v = \frac{\Delta t}{\Delta x} \max_i |\lambda_i|, i = 1, \dots, n. \quad (2.125)$$

For hyperbolic structures, we commonly use explicit three-point methods and grids in which the courant number is somewhat smaller than or equal to 1 ( $v \leq 1$ ). The numerical method can only be convergent if its numerical domain of dependence contains the exact domain of dependence of the conservation law, at least in the limit of  $\Delta x, \Delta t \rightarrow 0$ . Therefore, the CFL condition is only a necessary condition for stability and, hence, convergence.

Below are some of the most popular finite volume schemes for solving conservation laws.

## 2.4.2 Some numerical schemes

### 1. Upwind Scheme.

The first order Upwind conservative scheme has the form

$$U_i^{n+1} = U_i^n - \frac{k}{h} [f(U_i^n) - f(U_{i-1}^n)], \quad (2.126)$$

where  $k = \Delta t$  denotes the time step and  $h = \Delta x$  denotes the spatial domain length. Scheme (2.126) is known as the one-sided method and is stable with the characteristic speed  $f'(U) > 0$  for a scalar case or eigenvalues of the Jacobian matrix have the same sign in the system case.

When  $f'(U) < 0$ , the one-sided method read as

$$U_i^{n+1} = U_i^n - \frac{k}{h} [f(U_{i+1}^n) - f(U_i^n)], \quad (2.127)$$

## 2. Lax-Friedrichs Scheme.

This scheme takes the form

$$U_i^{n+1} = \frac{1}{2}(U_{i-1}^n + U_{i+1}^n) - \frac{k}{2h} [f(U_{i+1}^n) - f(U_{i-1}^n)], \quad (2.128)$$

where the numerical flux is given as

$$\phi(U_i^n, U_{i+1}^n) = \frac{1}{2}[f(U_i^n) + f(U_{i+1}^n) - \alpha(U_{i+1}^n - U_i^n)],$$

with  $\alpha = \Delta x / \Delta t$ .

## 3. Local Lax-Friedrichs Scheme.

The scheme of the form (2.128) is known as a local Lax-Friedrichs Scheme, where the numerical flux is defined as

$$\phi(U_i^n, U_{i+1}^n) = \frac{1}{2}[f(U_i^n) + f(U_{i+1}^n) - \alpha_{i+\frac{1}{2}}(U_{i+1}^n - U_i^n)]$$

with  $\alpha_{i+\frac{1}{2}} = \max |Df(U)|$  over all  $U$  between  $U_i^n$  and  $U_{i+1}^n$ . This is reduced to

$$\alpha_{i+\frac{1}{2}} = \max (|Df(U_i^n)|, |Df(U_{i+1}^n)|),$$

for a convex flux function. Notice that regardless of which end of the interval is larger,  $(U_i^n, U_{i+1}^n)$  is a non-empty interval.

## 4. Lax-Wendroff Scheme.

A Lax-Wendroff method is a second-order method that is given by the following scheme

$$U_i^{n+1} = U_i^n - \frac{k}{2h} [F_{i+1}^n - F_{i-1}^n] + \frac{k^2}{2h^2} \times \left[ \left( \frac{1}{2}(U_i^n + U_{i+1}^n) \right) (F_{i+1}^n - F_i^n) - \left( \frac{1}{2}(U_i^n + U_{i-1}^n) \right) (F_i^n - F_{i-1}^n) \right], \quad (2.129)$$

where the numerical flux is

$$\phi(U_i^n, U_{i+1}^n) = \frac{1}{2}[F_i^n + F_{i+1}^n - \frac{k}{h}((U_i^n + U_{i+1}^n)/2)(F_{i+1}^n - F_i^n)],$$

such that  $F_i^n = f(U_i^n)$ .



## 5. MacCormack Scheme.

The structure of the MacCormack scheme is of the predictor-corrector nature; it takes the following form

$$\begin{aligned} U_i^* &= U_i^n - \lambda[F_{i+1}^n - F_i^n], \\ U_i^{n+1} &= \frac{1}{2}(U_i^n + U_i^*) - \frac{\lambda}{2}[F_i^* - F_{i-1}^*], \end{aligned} \tag{2.130}$$

with the numerical flux given by

$$\phi(U_i^n, U_{i+1}^n) = \frac{1}{2}[F_i^n + F(U_i^n - \lambda(F_i^n - F_{i-1}^n))].$$

where  $F_i^* = f(U_i^*)$  and  $\lambda = \frac{\Delta t}{\Delta x}$ .

## 6. The Godunov Scheme.

A Godunov scheme can be written as in (2.121), which is a conservative scheme since the flux  $\phi(U_i^n, U_{i+1}^n)$  depends on the correct value (and Lipschitz continuity holds as well). The numerical flux of the Godunov scheme can be written as

$$\phi(U_i^n, U_{i+1}^n) = \begin{cases} \min_{U_i^n \leq U \leq U_{i+1}^n} f(U) & \text{if } U_i^n < U_{i+1}^n, \\ \max_{U_i^n \leq U \leq U_{i+1}^n} f(U) & \text{if } U_i^n \geq U_{i+1}^n. \end{cases}$$

This formula works on both convex and non-convex flux functions.

## 7. Murman - Roe Scheme.

A Murman - Roe Scheme is written as (2.121), with the numerical flux as

$$\phi(U_i^n, U_{i+1}^n) = \begin{cases} F_i^n & \text{if } a_{i+\frac{1}{2}}^n \geq 0, \\ F_{i+1}^n & \text{if } a_{i+\frac{1}{2}}^n < 0, \end{cases}$$

where  $F_i^n = f(U_i^n)$ , we will use a more sophisticated Roe average  $a_{i+\frac{1}{2}}^n = Df((U_i^n + U_{i+1}^n)/2)$

as

$$a_{i+\frac{1}{2}}^n = \begin{cases} \frac{F_{i+1}^n - F_i^n}{U_{i+1}^n - U_i^n} & \text{if } U_{i+1}^n \neq U_i^n, \\ Df(U_i^n) & \text{if } U_{i+1}^n = U_i^n. \end{cases}$$

## 8. Semi-discrete central Upwind scheme.

This scheme is based on an integral form of conservation laws that can be obtained by integrating (2.1) over control volumes. Furthermore, the numerical integration of the scheme takes the form

$$\frac{d}{dt} \bar{U}_i(t) = - \frac{H_{i+\frac{1}{2}}(t) - H_{i-\frac{1}{2}}(t)}{\Delta x}, \quad (2.131)$$

where  $H_{i+\frac{1}{2}}$  are numerical fluxes. Once the numerical fluxes in (2.131) have been computed, the scheme will be complete. However, we can define the numerical fluxes in (2.131) using the central-upwind numerical fluxes in the form

$$H_{i+\frac{1}{2}}(t) := \frac{a_{i+\frac{1}{2}}^+ f(U_{i+\frac{1}{2}}^-) - a_{i+\frac{1}{2}}^- f(U_{i+\frac{1}{2}}^+)}{a_{i+\frac{1}{2}}^+ - a_{i+\frac{1}{2}}^-} + a_{i+\frac{1}{2}}^+ a_{i+\frac{1}{2}}^- \left[ \frac{U_{i+\frac{1}{2}}^+ - U_{i+\frac{1}{2}}^-}{a_{i+\frac{1}{2}}^+ - a_{i+\frac{1}{2}}^-} \right]. \quad (2.132)$$

**Remark.** For convenience, we omitted the  $t$  notation from  $a_{i+\frac{1}{2}}^+$  and  $U_{i+\frac{1}{2}}^+$  for time dependency.

In equation (2.132),  $U_{i+\frac{1}{2}}^-$  and  $U_{i+\frac{1}{2}}^+$  are the left and right point values, respectively, of the piecewise linear reconstructions

$$\tilde{U}(x) = \sum_i \left[ \bar{U}_i + \left( \frac{\partial U}{\partial x} \right)_i (x - x_i) \right] \chi_{C_i}(x), \quad (2.133)$$

introduced at the cell interface  $x = x_{i+\frac{1}{2}}$  given by

$$U_{i+\frac{1}{2}}^+ = \bar{U}_{i+1} - \frac{\Delta x}{2} \left( \frac{\partial U}{\partial x} \right)_{i+1}, \quad U_{i+\frac{1}{2}}^- = \bar{U}_i + \frac{\Delta x}{2} \left( \frac{\partial U}{\partial x} \right)_i, \quad (2.134)$$

where  $\chi_{C_i}$  denotes the characteristic function of the interval  $C_i$  and  $(\partial U / \partial x)_i$  denotes the numerical derivatives, which can be calculated using a non-linear limiter to reduce oscillations. Here, we use the generalised minmod limiter in the form

$$\left( \frac{\partial U}{\partial x} \right)_i = \text{minmod} \left( \theta \frac{\bar{U}_i - \bar{U}_{i-1}}{\Delta x}, \frac{\bar{U}_{i+1}^n - \bar{U}_{i-1}^n}{2\Delta x}, \theta \frac{\bar{U}_{i+1} - \bar{U}_i}{\Delta x} \right), \quad \theta \in [1, 2], \quad (2.135)$$

with

$$\text{minmod}(z_1, z_2, \dots, z_m) = \begin{cases} \min_i \{z_i\} & \text{if } z_i > 0 \quad \forall i = 1, 2, \dots, m, \\ \max_i \{z_i\} & \text{if } z_i < 0 \quad \forall i = 1, 2, \dots, m, \\ 0 & \text{otherwise.} \end{cases} \quad (2.136)$$

The parameter  $\theta$  in equation (2.135) is used to control the amount of numerical viscosity in the resulting scheme. In general, large values of  $\theta$  result in less dissipative results.

Moreover,  $a_{i+\frac{1}{2}}^-$  and  $a_{i+\frac{1}{2}}^+$  are the left- and right-sided local speeds of propagation, respectively, that can be obtained using the smallest and largest eigenvalues of the Jacobian matrix  $A(U)$  defined as

$$\begin{aligned} a_{i+\frac{1}{2}}^- &= \min_{U \in C_i} \left\{ \lambda_1(A(U)), 0 \right\}, \\ a_{i+\frac{1}{2}}^+ &= \max_{U \in C_i} \left\{ \lambda_N(A(U)), 0 \right\}, \end{aligned} \quad (2.137)$$

where  $\lambda_1 < \dots < \lambda_N$  are the  $N$  eigenvalues of the corresponding Jacobian matrix.

The above scheme may involve some oscillations. For more accuracy, a correction term can be added to the numerical fluxes (6). The correction term, denoted by  $q_{i+\frac{1}{2}}$ , can be defined as follows

$$q_{i+\frac{1}{2}} = \alpha \cdot \text{minmod} \left( \frac{U_{i+\frac{1}{2}}^+ - W_{i+\frac{1}{2}}^{int}}{a_{i+\frac{1}{2}}^+ - a_{i+\frac{1}{2}}^-}, \frac{W_{i+\frac{1}{2}}^{int} - U_{i+\frac{1}{2}}^-}{a_{i+\frac{1}{2}}^+ - a_{i+\frac{1}{2}}^-} \right), \quad \alpha \in [0, 1], \quad (2.138)$$

where the intermediate values  $W_{i+\frac{1}{2}}^{int}$  are obtained as we pass to the (intermediate) cell averages at time  $t = t^{n+1}$ .

$$W_{i+\frac{1}{2}}^{int} = \frac{a_{i+\frac{1}{2}}^+ U_{i+\frac{1}{2}}^+ - a_{i+\frac{1}{2}}^- U_{i+\frac{1}{2}}^- - \left\{ f(U_{i+\frac{1}{2}}^+) - f(U_{i+\frac{1}{2}}^-) \right\}}{a_{i+\frac{1}{2}}^+ - a_{i+\frac{1}{2}}^-}. \quad (2.139)$$

Equations (2.131)-(2.132) and (2.138)-(2.139) are a one-parameter family of semi-discrete central upwind schemes. However, the amount of numerical dissipation decreasing as  $\alpha$  increases and once  $\alpha = 0$ , the scheme is equivalent to the original central-upwind system presented in (2.131)-(2.132).

The theoretical concepts of conservation laws and the numerical framework have been fully discussed above. In the subsequent section, we will show the application of these schemes by describing the numerical algorithm and the further implementation using Matlab for analysis.

## 2.5 Numerical results based on finite volume schemes

In this section, numerical results for the linear (Linear advection) and non-linear (Burger's equation) scalar conservation laws are presented, based on the schemes introduced above. Results on non-linear systems of conservation laws are also considered; in this case, we are interested in the experiment of two nonlinear systems namely the shallow water equations and the Euler equations of gas dynamics. Results from various schemes are also compared.

### 2.5.1 Linear advection equation

Consider the initial value problem for the linear advection equation

$$\frac{\partial U}{\partial t} + c \frac{\partial U}{\partial x} = 0, \quad (2.140)$$

with the initial data  $U(0, x) = x$ . Therefore, we can obtain the analytical solution to this Cauchy problem as

$$U(t, x) = x - ct. \quad (2.141)$$

However, the numerical and exact solutions presented in Figure 2.5 were obtained using the Upwind scheme. Results are good approximation under the computational domain  $[0, 3]$  and speed of propagation  $c = 1$  for convenience.

Following Liu et al. [92], we can assume a Riemann problem for the linear advection equation (2.140) with initial piecewise data given by

$$U(0, x) = \begin{cases} 2 & \text{if } x < 0.5, \\ 1 & \text{if } x > 0.5. \end{cases} \quad (2.142)$$

The weak solution of Riemann problem (2.140), (2.142) can be found as

$$U(t, x) = \begin{cases} 2 & \text{if } x < 0.5 + st, \\ 1 & \text{if } x > 0.5 + st, \end{cases} \quad (2.143)$$

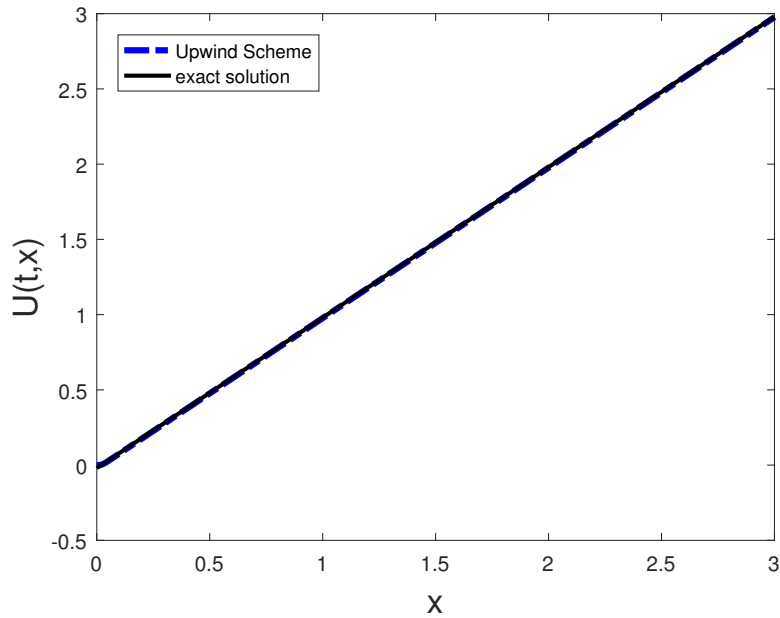


Figure 2.5: Numerical and exact solutions to the initial value problem (2.140) with  $U(0,x) = x$ .

where  $s = c$  is a wave speed given by the Rankine-Hugoniot equation. Numerical and exact solutions posted in Figure 2.6, using the upwind and semi-discrete central upwind schemes. It is observed that both results show very good approximations; however, the second-order schemes show more accuracy.

### 2.5.2 Burger's equation

A weak solution to the non-linear system may be losing uniqueness due to the velocity depending on the solution itself. One can start with Burger's equation [93] as a non-linear scalar system to demonstrate this behavior in practice. Consider the inviscid Burger's equation in the conservative form

$$\frac{\partial U}{\partial t} + \frac{\partial}{\partial x} \left( \frac{U^2}{2} \right) = 0, \quad (2.144)$$

with the smooth initial data  $U(0,x) = x$ . Therefore, The exact solution is given as

$$U(t,x) = \frac{x}{1+t}. \quad (2.145)$$

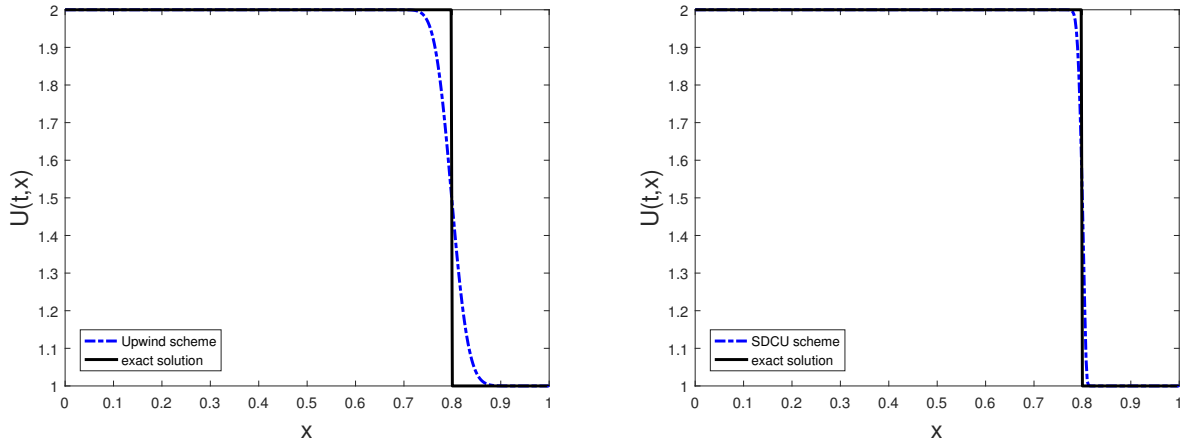


Figure 2.6: Solutions of linear advection equation (2.140) with Riemann initial data (2.142): Upwind scheme (Left) and Semi-discrete central Upwind scheme (Right).

The numerical solutions obtained using the Upwind method, Lax-Friedrich and Lax-Wendroff schemes are in excellent agreement. The error-norms for the Cauchy problem (2.144) with  $U(0, x) = x$  are presented in Table 2.1. However, the results show that the results with the Lax-Wendroff scheme are approximately slightly better than Upwind and Lax-Friedrich schemes.

Consider Riemann problem of Burger's equation (2.144) with the initial data as in [94]

$$U(0, x) = \begin{cases} 1 & \text{if } x < 0, \\ 0 & \text{if } x \geq 0. \end{cases} \quad (2.146)$$

Then, we obtain the exact solution as

$$U(t, x) = \begin{cases} 1 & \text{if } x < \frac{1}{2}t, \\ 0 & \text{if } x \geq \frac{1}{2}t. \end{cases} \quad (2.147)$$

Here, the solution is that shock wave propagates with speed  $\lambda = \frac{1}{2}$ , given by Rankine-Hugoniot jump condition. However, numerical and exact solutions of the Riemann problem (2.144) and (2.146) are depicted in Figure 2.7. Results presented represent approximations to the exact solution at time  $t = 0.5$ , using the Upwind, Lax-Friedrich and Lax-Wendroff schemes. On the other hand, suppose

N	Scheme	$L^1$ -error	$L^2$ -error	$L^\infty$ -error
100	Upwind	0.4435	0.0511	0.0087
	Lax-Friedrich	0.8436	0.2602	0.0977
	Lax-Wendroff	0.1136	0.0142	0.0053
200	Upwind	0.5112	0.0760	0.0092
	Lax-Friedrich	0.8678	0.2618	0.0727
	Lax-Wendroff	0.1222	0.0113	0.0042
300	Upwind	0.8323	0.0945	0.0094
	Lax-Friedrich	0.9630	0.2774	0.0615
	Lax-Wendroff	0.1317	0.0100	0.0039

Table 2.1: Error-norms for the Burger’s equation (2.144) with initial data  $U(0,x) = x$ , with grid points 100, 200 and 300.

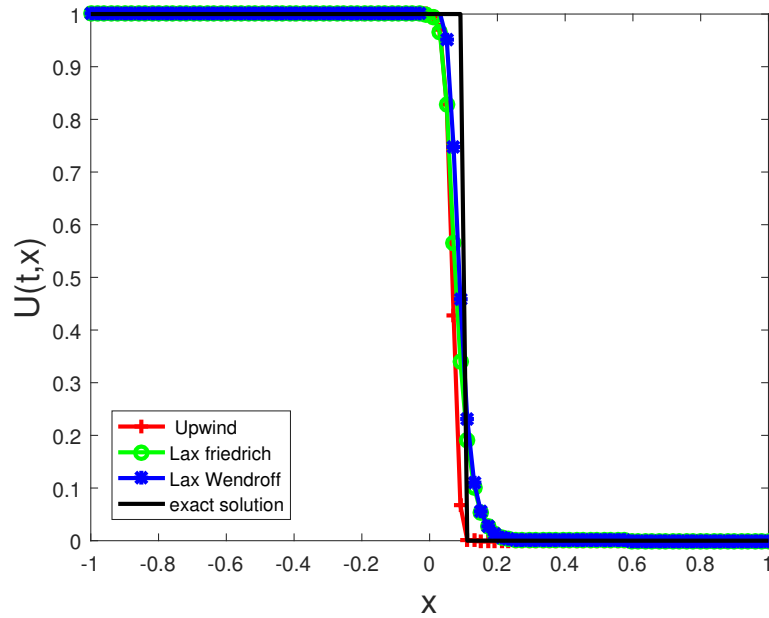


Figure 2.7: Numerical and exact solutions to the Riemann problem (2.144) and (2.146).

the Riemann problem with (2.144) and initial data are given as

$$U(0,x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 & \text{if } x \geq 0. \end{cases} \quad (2.148)$$

Thus, the solution to the Riemann problem (2.144) and (2.148) that includes discontinuity, which is a rarefaction wave, can be identified. The numerical solutions obtained by the Upwind, Lax-Friedrich and Lax-Wendroff methods are shown in Figure 2.8. All the presented results are approximately close to the exact solution. Furthermore, the error-norms for the Riemann problem (2.144) and (2.148) are presented in Table 2.2.

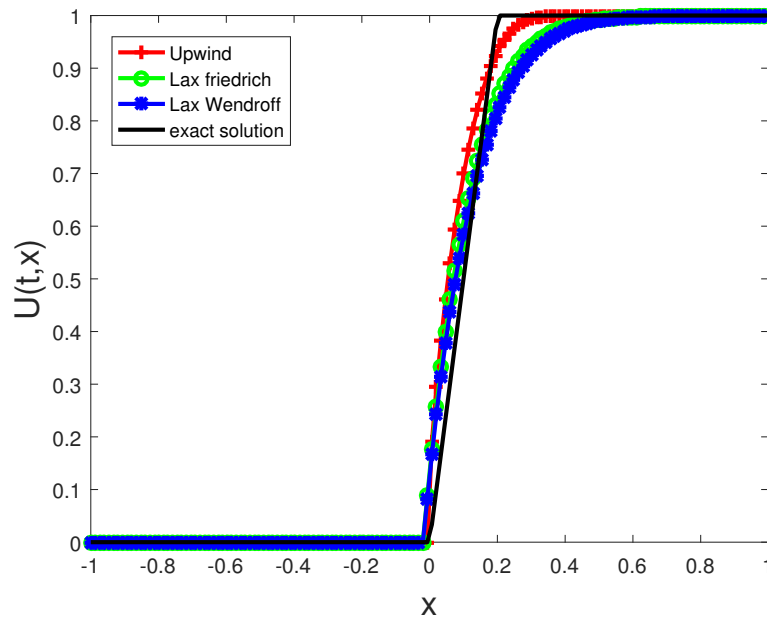


Figure 2.8: Numerical and exact solutions to the Riemann problem (2.144) and (2.148).



N	Scheme	$L^1$ -error	$L^2$ -error	$L^\infty$ -error
100	Upwind	0.1494	0.1301	0.1288
	Lax-Friedrich	0.6639	0.5566	0.4271
	Lax-Wendroff	0.6760	0.5575	0.4217
200	Upwind	0.3446	0.2878	0.2832
	Lax-Friedrich	0.9227	0.4677	0.3859
	Lax-Wendroff	0.9240	0.4643	0.3811
300	Upwind	0.4984	0.3920	0.3811
	Lax-Friedrich	0.8766	0.5161	0.3520
	Lax-Wendroff	0.8794	0.5128	0.3477
400	Upwind	0.5794	0.4253	0.4063
	Lax-Friedrich	0.9444	0.4917	0.3389
	Lax-Wendroff	0.9391	0.4866	0.3334
500	Upwind	0.7254	0.4712	0.4068
	Lax-Friedrich	0.9945	0.5157	0.3573
	Lax-Wendroff	0.9905	0.5103	0.3521

Table 2.2: Error-norms for the Burger's equation (2.144) with initial data (2.148), with grid points 100, 200, 300, 400 and 500.

### 2.5.3 Shallow water equations

Consider the one-dimensional system of shallow water equations in a rectangular channel with the zero-flat bottom as

$$\begin{aligned}
\frac{\partial h}{\partial t} + \frac{\partial}{\partial x}(hu) &= 0, \\
\frac{\partial}{\partial t}(hu) + \frac{\partial}{\partial x}\left(hu^2 + \frac{1}{2}gh^2\right) &= 0,
\end{aligned}
\tag{2.149}$$

and the initial piecewise data [93] given by

$$h(0,x) = \begin{cases} 10 & \text{if } x < 0, \\ 5 & \text{if } x > 0, \end{cases} \quad u(0,x) = 0. \quad (2.150)$$

The Riemann problem (2.149) and (2.150) is called a **Dam break problem**. In addition, the left state must be higher than the right state to be consistent with the physical phenomenon of a dam-break problem. Therefore, at  $t = 0$ , the dam collapses and the flow problem consists of a shock wave travelling downstream and a rarefaction wave travelling upstream. Results of the dam break problem are introduced in Figure 2.9 using the Semi-discrete central upwind schemes, with computational domains  $[-1000, 1000]$  and without correction term. However, the implementation with a large number of grid points is slightly efficient.

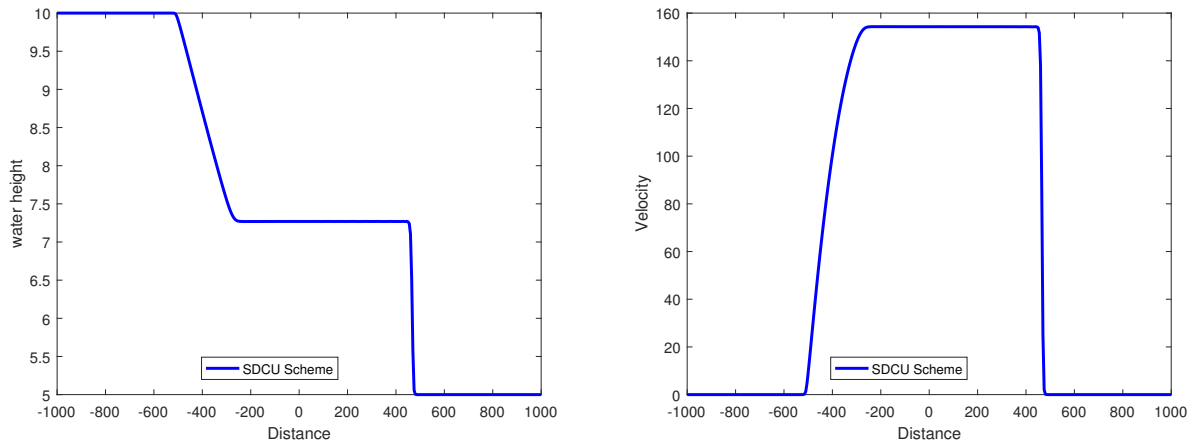


Figure 2.9: Numerical results for the dam break problem (2.149) and (2.150).

## 2.5.4 System of Euler equations

In this section, numerical results of the Riemann problem for Euler equations with initial data will be reported in the following lines. Consider the 1D system of Euler equations of gas dynamics in

the form

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho u) &= 0, \\ \frac{\partial}{\partial t}(\rho u) + \frac{\partial}{\partial x}(\rho u^2 + P) &= 0, \\ \frac{\partial E}{\partial t} + \frac{\partial}{\partial x}(u(E + P)) &= 0, \end{aligned} \tag{2.151}$$

with initial data [95] given as

$$(\rho, u, P)(0, x) = \begin{cases} (1.0, 0.0, 1.0) & \text{if } x < 0.5, \\ (0.125, 0.0, 0.1) & \text{if } x > 0.5. \end{cases} \tag{2.152}$$

The problem (2.151) with (2.152) is called a **shock tube problem**. The numerical results for the density, velocity, pressure and energy are presented in Figure 2.10 by implementing the semi-discrete central Upwind scheme with the correction term to drop the oscillations. The results obtained are in excellent agreement. However, second-order schemes produce more accurate computations.

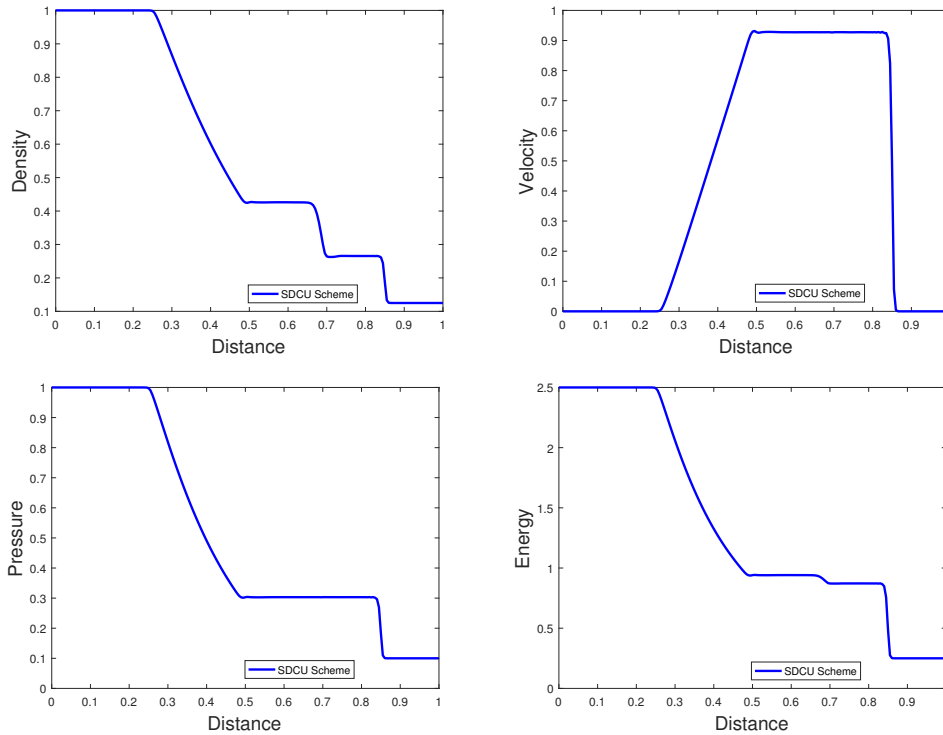


Figure 2.10: Numerical results for the Shock tube problem (2.151) and (2.152).

Up to this point, systems of conservation laws and their solutions, which are nonlinear, have been presented. Explicit solutions may not be possible to obtain due to the nonlinearity and discontinuities may exist that pose challenges in both analytical solutions and numerical simulations. Thus, the following section will introduce a semi-linear approximation of systems of conservation laws namely the Relaxation approach.

## 2.6 Relaxation systems

The relaxation system is used to convert the non-linear system of conservation laws into a linear transport equation system with a non-linear source term following the approach proposed by Jin and Xin [9]. Given the initial value problem (2.8), we will construct a linear hyperbolic system with a stiff source term that approximates the original system with a small dissipative correction. Therefore, the relaxation system has the form

$$\begin{aligned}\frac{\partial}{\partial t}U + \frac{\partial}{\partial x}V &= 0, \\ \frac{\partial}{\partial t}V + a^2 \frac{\partial}{\partial x}U &= -\frac{1}{\varepsilon}(V - f(U)), \\ U(0,x) &= U_0(x), \quad V(0,x) = f(U_0(x)),\end{aligned}\tag{2.153}$$

where  $V \in \mathbb{R}$  is the artificial relaxation variable,  $\varepsilon$  is the relaxation rate and  $a$  is a given positive constant (characteristic speed) of the relaxation system. The relaxation system (2.153) has a typical semi-linear structure with the characteristic variables of the transport part given by

$$V + aU \quad \text{and} \quad V - aU.\tag{2.154}$$

Using the small relaxation limit  $\varepsilon \rightarrow 0$ , the solution of the system (2.153) approximates the solution of the conservation law (2.1), that is, the relaxation system can be approximated to have local equilibrium and original conservation law, respectively, as

$$\begin{aligned}V &= f(U), \\ \frac{\partial}{\partial t}U + \frac{\partial}{\partial x}f(U) &= 0.\end{aligned}\tag{2.155}$$

Moreover, for small  $\varepsilon$ , the stability criterion can be (formally) derived by using the Chapman-Enskog expansion [57]. Namely, we can write the first-order approximation for  $V$  as

$$V = f(U) + \varepsilon V_1, \quad (2.156)$$

thus, we have

$$\frac{\partial}{\partial x} V = \frac{\partial}{\partial x} f(U) + \varepsilon \frac{\partial}{\partial x} V_1. \quad (2.157)$$

Inserting (2.157) in the first equation of the relaxation system (2.153), we get

$$\frac{\partial}{\partial t} U + \frac{\partial}{\partial x} f(U) = -\varepsilon \frac{\partial}{\partial x} V_1, \quad (2.158)$$

and substituting (2.156) in the second equation of (2.153), we obtain

$$\frac{\partial}{\partial t} f(U) + a^2 \frac{\partial}{\partial x} U + V_1 = O(\varepsilon), \quad (2.159)$$

where  $\frac{\partial}{\partial t} f(U)$  is the derivative of  $f$  with respect to  $t$ . Then, (2.159) becomes

$$f'(U) \frac{\partial}{\partial t} U + a^2 \frac{\partial}{\partial x} U + V_1 = O(\varepsilon), \quad (2.160)$$

where  $f'(U)$  is the Jacobian matrix of the flux function  $f(U)$ . Using the notation that  $\frac{\partial U}{\partial t} = -\frac{\partial V}{\partial x} = -\left[\frac{\partial}{\partial x} f(U) + \varepsilon \frac{\partial}{\partial x} V_1\right]$  in equation (2.160), we have

$$-f'(U)^2 \frac{\partial}{\partial x} U + a^2 \frac{\partial}{\partial x} U = O(\varepsilon) - V_1. \quad (2.161)$$

Therefore, dropping the  $O(\varepsilon)$  term, we have the first-order approximation for  $\varepsilon \ll 1$  gives as

$$(a^2 - f'(U)^2) \frac{\partial}{\partial x} U = -V_1. \quad (2.162)$$

Substituting (2.162) into (2.158), we have the second-order approximation for  $U$ , that is,

$$\frac{\partial}{\partial t} U + \frac{\partial}{\partial x} f(U) = \varepsilon \frac{\partial}{\partial x} \left[ (a^2 - f'(U)^2) \frac{\partial}{\partial x} U \right]. \quad (2.163)$$

The system (2.163) is dissipative if and only if (iff) the sub-characteristic condition

$$\max_U |(f'(U))| \leq a, \quad (2.164)$$

holds. Then, the relaxation system (2.153) converges to the system of conservation laws (2.1) if and only if the sub-characteristic condition (2.164) is satisfied.

For the relaxation system defined above, the discretisation can be done without using Riemann solvers and can be numerically solved using a first-order upwind scheme [9, 81]. A second-order MUSCL scheme for the space discretisation together with a second-order TVD implicit-explicit (IMEX) Runge–Kutta scheme for the time integration [9, 82] has been introduced.

## 2.7 Discretisation of the relaxation system

In this section, we discuss the numerical discretisation of the system of conservation laws based on the Jin–Xin discretisation of the relaxation system (2.153) that was proposed by Jin and Xin [9]. The method of lines [96] will be applied where we consider the spatial discretisation of a system while maintaining it continuously in time. Then, the TVD Runge–Kutta time discretisations method for time discretisation will be presented.

### 2.7.1 Spatial discretisation

To discretise the relaxation system (2.153), for simplicity, we introduce the spatial grid points  $x_{i+\frac{1}{2}}$  with grid spacing  $\Delta x_i = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$ , the uniform discrete time step  $\Delta t = t^{n+1} - t^n$ , where  $n = 0, 1, \dots, N$ . The approximation notation is given as  $U_{i+\frac{1}{2}}^n = U(t^n, x_{i+\frac{1}{2}})$  and define

$$D_x U_i = \frac{U_{i+\frac{1}{2}} - U_{i-\frac{1}{2}}}{\Delta x_i}. \quad (2.165)$$

The semi-discrete approximation for the relaxation system (2.153) in the conservation form can be written as

$$\begin{aligned} \frac{\partial U_i}{\partial t} + D_x V_i &= 0, \\ \frac{\partial V_i}{\partial t} + a^2 D_x V_i &= -\frac{1}{\varepsilon} (V_i - f_i), \end{aligned} \quad (2.166)$$

where the average quantities are given by

$$\begin{aligned} f_i &= \frac{1}{\Delta x_i} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} f(U) dx = f \left( \frac{1}{\Delta x_i} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} U dx \right) + O(\overline{\Delta x}^2) \\ &= f(U_i) + O(\overline{\Delta x}^2), \end{aligned} \quad (2.167)$$

with an accuracy of  $O(\overline{\Delta x}^2)$  such that  $\Delta x = \max_i \Delta x_i$ . For sufficiently accurate spatial discretisations, system (2.166) can be written as

$$\begin{aligned} \frac{\partial U_i}{\partial t} + D_x V_i &= 0, \\ \frac{\partial V_i}{\partial t} + a^2 D_x V_i &= -\frac{1}{\varepsilon} (V_i - f(U_i)). \end{aligned} \quad (2.168)$$

### 1. First-order discretisation.

Applying the first-order upwind scheme to the characteristic variables (2.154) gives the point value quantities  $U_{i+\frac{1}{2}}$  and  $V_{i+\frac{1}{2}}$  as

$$\begin{aligned} (V + aU)_{i+\frac{1}{2}} &= (V + aU)_i = V_i + aU_i, \\ (V - aU)_{i+\frac{1}{2}} &= (V - aU)_{i+1} = V_{i+1} - aU_{i+1}. \end{aligned} \quad (2.169)$$

Solving (2.169), we obtain

$$\begin{aligned} U_{i+\frac{1}{2}} &= \frac{1}{2} (U_i + U_{i+1}) - \frac{1}{2a} (V_{i+1} - V_i), \\ V_{i+\frac{1}{2}} &= \frac{1}{2} (V_i + V_{i+1}) - \frac{1}{2} a (U_{i+1} - U_i). \end{aligned} \quad (2.170)$$

Substituting (2.170) into (2.168) and using the notation (2.165), we obtain the first-order semi-discrete upwind approximation to the relaxation system (2.153) as

$$\begin{aligned} \frac{\partial U_i}{\partial t} + \frac{1}{2\Delta x_i} (V_{i+1} - V_{i-1}) - \frac{a}{2\Delta x_i} (U_{i+1} - 2U_i + U_{i-1}) &= 0, \\ \frac{\partial V_i}{\partial t} + \frac{a^2}{2\Delta x_i} (U_{i+1} - U_{i-1}) - \frac{a}{2\Delta x_i} (V_{i+1} - 2V_i + V_{i-1}) &= -\frac{1}{\varepsilon} (V_i - f(U_i)). \end{aligned} \quad (2.171)$$

### 2. Second-order discretisation.

In this section, we consider a second-order MUSCL scheme for the approximations of the

system (2.153). Briefly, we introduce the construct of a slope limiter type scheme with enough diffusion to avoid oscillations and has been presented in [65, 82]. Given a piecewise linear interpolation, when applied to the  $r$ -th components (Denoted as  $v \pm a_r u$ ) of  $V \pm aU$  gives

$$\begin{aligned}(v + a_r u)_{i+\frac{1}{2}} &= (v + a_r u)_i + \frac{1}{2} \Delta x_i S_i^+, \\ (v - a_r u)_{i+\frac{1}{2}} &= (v - a_r u)_{i+1} - \frac{1}{2} \Delta x_{i+1} S_{i+1}^-, \end{aligned} \quad (2.172)$$

where  $u$  and  $v$  are the  $r$ -th ( $r = 1, 2, \dots, n$ ) components of  $U$  and  $V$  respectively and  $S_i^\pm$  is the slope of  $v \pm a_r u$  on the  $i$ -th cell. The slopes are given by

$$S_i^\pm = \frac{1}{\Delta x_i} (v_{i+1} \pm a_r u_{i+1} - v_i \mp a_r u_i) \phi(\theta_i^\pm), \quad (2.173)$$

where

$$\theta_i^\pm = \frac{v_i \pm a_r u_i - v_{i-1} \mp a_r u_{i-1}}{v_{i+1} \pm u_{i+1} - v_i \mp a_r u_i}. \quad (2.174)$$

The function  $\phi$  is a slope limiter function and it satisfies the more general condition

$$0 \leq \frac{\phi(\theta)}{\theta} \leq 2 \quad \text{and} \quad 0 \leq \phi(\theta) \leq 2. \quad (2.175)$$

The simplest choice of a slope limiter is the so-called minmod limiter given by

$$\phi(\theta) = \max(0, \min(1, \theta)), \quad (2.176)$$

where

$$\text{minmod}(A, B) = \frac{\text{sgn}(A) + \text{sgn}(B)}{2} \min(|A|, |B|),$$

and a sharper van Leer limiter [97], which is given by

$$\phi(\theta) = \frac{|\theta| + \theta}{1 + |\theta|}. \quad (2.177)$$

Solving (2.172) for  $u_{i+\frac{1}{2}}$  and  $v_{i+\frac{1}{2}}$  gives

$$\begin{aligned}u_{i+\frac{1}{2}} &= \frac{1}{2} (u_i + u_{i+1}) - \frac{1}{2a_r} (v_{i+1} - v_i) + \frac{1}{4a_r} (\Delta x_i S_i^+ + \Delta x_{i+1} S_{i+1}^-), \\ v_{i+\frac{1}{2}} &= \frac{1}{2} (v_i + v_{i+1}) - \frac{a_r}{2} (u_{i+1} - u_i) + \frac{1}{4} (\Delta x_i S_i^+ - \Delta x_{i+1} S_{i+1}^-). \end{aligned} \quad (2.178)$$



In (2.178) and (2.168), we obtain the MUSCL for the relaxation system (2.153) component-wise as

$$\begin{aligned}
\frac{\partial u_i}{\partial t} + \frac{1}{2\Delta x_i} (v_{i+1} - v_{i-1}) - \frac{a_r}{2\Delta x_i} (u_{i+1} - 2u_i + u_{i-1}) \\
- \frac{1}{4\Delta x_i} (\Delta x_{i+1} S_{i+1}^- - \Delta x_i (S_i^+ + S_i^-) + \Delta x_{i-1} S_{i-1}^+) = 0, \\
\frac{\partial v_i}{\partial t} + \frac{a_r}{2\Delta x_i} (u_{i+1} - u_{i-1}) - \frac{a_r}{2\Delta x_i} (v_{i+1} - 2v_i + v_{i-1}) \\
+ \frac{a_r}{4\Delta x_i} (\Delta x_{i+1} S_{i+1}^- + \Delta x_i (S_i^+ - S_i^-) - \Delta x_{i-1} S_{i-1}^+) \\
= -\frac{1}{\varepsilon} (v_i - f^{(r)}(u_i)).
\end{aligned} \tag{2.179}$$

Equations (2.179) are known as the second-order relaxing scheme where  $\varepsilon > 0$ .

## 2.7.2 Time integration

In this subsection, we consider an Implicit-Explicit (IMEX) algorithm, which has been presented by Banda and Seaid [82]. Following a similar Runge–Kutta time discretisation scheme of the relaxation system (2.153), the scheme takes two steps: an implicit step for a stiff ordinary differential equation (ODE) and an explicit step for the system of transport equations. Before discretising the relaxation system (2.153), we split the system into two parts namely a system of Stiff ODE

$$\begin{aligned}
\frac{\partial U}{\partial t} &= 0, \\
\frac{\partial V}{\partial t} &= -\frac{1}{\varepsilon} (V - f(U)),
\end{aligned} \tag{2.180}$$

and the non-stiff transport system

$$\begin{aligned}
\frac{\partial U}{\partial t} + \frac{\partial V}{\partial x} &= 0, \\
\frac{\partial V}{\partial t} + a^2 \frac{\partial U}{\partial x} &= 0.
\end{aligned} \tag{2.181}$$

The fully discrete relaxation scheme for given starting initial data  $U_i^n$  and  $V_i^n = f(U_i^n)$  as the following:

### 1. First-order discretisation.

The implementation of the first-order relaxation algorithm to solve (2.153) is carried out in simple steps as follows:

$$\begin{aligned}
 U_i^* &= U_i^n, \\
 V_i^* &= V_i^n - \frac{\Delta t}{\varepsilon} (V_i^* - f(U_i^*)); \\
 U_i^{(1)} &= U_i^* - \Delta t D_x V_i^*, \\
 V_i^{(1)} &= V_i^* - \Delta t a^2 D_x U_i^*; \\
 U_i^{n+1} &= U_i^{(1)}, \\
 V_i^{n+1} &= V_i^{(1)}.
 \end{aligned} \tag{2.182}$$

The scheme (2.182) can be written explicitly using equations (2.170) as

$$\begin{aligned}
 U_i^* &= U_i^n, \\
 V_i^* &= \left( \frac{\varepsilon}{\varepsilon + \Delta t} \right) \left( V_i^n + \frac{\Delta t}{\varepsilon} f(U_i^*) \right); \\
 U_i^{(1)} &= U_i^* - \frac{\Delta t}{2\Delta x_i} [(V_{i+1}^* - V_{i-1}^*) - a(U_{i+1}^* - 2U_i^* + U_{i-1}^*)], \\
 V_i^{(1)} &= V_i^* - \frac{\Delta t a^2}{2\Delta x_i} \left[ (U_{i+1}^* - U_{i-1}^*) - \frac{1}{a}(V_{i+1}^* - 2V_i^* + V_{i-1}^*) \right]; \\
 U_i^{n+1} &= U_i^{(1)}, \\
 V_i^{n+1} &= V_i^{(1)}.
 \end{aligned} \tag{2.183}$$

When  $\varepsilon \rightarrow 0$ , equations (2.182) reduce to the so-called relaxed scheme

$$\begin{aligned}
 U_i^{(1)} &= U_i^n - \Delta t D_x V_i^n |_{V_i^n = f(U_i^n)}, \\
 U_i^{n+1} &= U_i^{(1)},
 \end{aligned} \tag{2.184}$$

which is the first-order explicit scheme.

### 2. Second-order discretisation.

The second-order relaxation scheme to solve equations (2.153) as follows

$$\begin{aligned}
U_i^* &= U_i^n, \\
V_i^* &= V_i^n + \frac{\Delta t}{\varepsilon} (V_i^* - f(U_i^*)); \\
U_i^{(1)} &= U_i^* - \Delta t D_x V_i^*, \\
V_i^{(1)} &= V_i^* - \Delta t a^2 D_x U_i^*; \\
U_i^{**} &= U_i^{(1)}, \\
V_i^{**} &= V_i^{(1)} - \frac{\Delta t}{\varepsilon} (V_i^{**} - f(U_i^{**})) - \frac{2\Delta t}{\varepsilon} (V_i^* - f(U_i^*)); \\
U_i^{(2)} &= U_i^{**} - \Delta t D_x V_i^{**}, \\
V_i^{(2)} &= V_i^{**} - \Delta t a^2 D_x U_i^{**}; \\
U_i^{n+1} &= \frac{1}{2} (U_i^n + U_i^{(2)}), \\
V_i^{n+1} &= \frac{1}{2} (V_i^n + V_i^{(2)}).
\end{aligned} \tag{2.185}$$

Using equations (2.178) with the initial data  $U^n = u_i^n$  and  $V^n = f(U^n) = f(u_i^n)$ , we can rewrite the scheme (2.185) explicitly as

$$\begin{aligned}
U_i^* &= U_i^n, \\
V_i^* &= \left( \frac{\varepsilon}{\varepsilon - \Delta t} \right) \left( V_i^n - \frac{\Delta t}{\varepsilon} f(U_i^*) \right); \\
U_i^{(1)} &= U_i^* - \frac{\Delta t}{2\Delta x_i} (V_{i+1}^* - V_{i-1}^*) + \frac{\Delta t a_r}{2\Delta x_i} (U_{i+1}^* - 2U_i^* + U_{i-1}^*) \\
&\quad + \frac{\Delta t}{4\Delta x_i} (\Delta x_{i+1} S_{i+1}^{*-} - \Delta x_i (S_i^{+*} + S_i^{-*}) + \Delta x_{i-1} S_{i-1}^{+*}), \\
V_i^{(1)} &= V_i^* - \frac{\Delta t a^2}{2\Delta x_i} (U_{i+1}^* - U_{i-1}^*) + \frac{\Delta t a^2}{2a_r \Delta x_i} (V_{i+1}^* - 2V_i^* + V_{i-1}^*) \\
&\quad - \frac{\Delta t a^2}{4a_r \Delta x_i} (\Delta x_{i+1} S_{i+1}^{*-} + \Delta x_i (S_i^{+*} - S_i^{-*}) - \Delta x_{i-1} S_{i-1}^{+*});
\end{aligned}$$

$$\begin{aligned}
U_i^{**} &= U_i^{(1)}, \\
V_i^{**} &= \left( \frac{\varepsilon}{\varepsilon + \Delta t} \right) \left( V_i^{(1)} + \frac{\Delta t}{\varepsilon} f(U_i^{**}) \right) - \left( \frac{2\Delta t}{\varepsilon + \Delta t} \right) (V_i^* - f(U_i^*)); \\
U_i^{(2)} &= U_i^{**} - \frac{\Delta t}{2\Delta x_i} (V_{i+1}^{**} - V_{i-1}^{**}) + \frac{\Delta t a_r}{2\Delta x_i} (U_{i+1}^{**} - 2U_i^{**} + U_{i-1}^{**}) \\
&\quad + \frac{\Delta t}{4\Delta x_i} (\Delta x_{i+1} S_{i+1}^{-**} - \Delta x_i (S_i^{+**} + S_i^{-**}) + \Delta x_{i-1} S_{i-1}^{+**}), \\
V_i^{(2)} &= V_i^{**} - \frac{\Delta t a^2}{2\Delta x_i} (U_{i+1}^{**} - U_{i-1}^{**}) + \frac{\Delta t a^2}{2a_r \Delta x_i} (V_{i+1}^{**} - 2V_i^{**} + V_{i-1}^{**}) \\
&\quad - \frac{\Delta t a^2}{4a_r \Delta x_i} (\Delta x_{i+1} S_{i+1}^{-**} + \Delta x_i (S_i^{+**} - S_i^{-**}) - \Delta x_{i-1} S_{i-1}^{+**}); \\
U_i^{n+1} &= \frac{1}{2} (U_i^n + U_i^{(2)}), \\
V_i^{n+1} &= \frac{1}{2} (V_i^n + V_i^{(2)}).
\end{aligned}$$

When  $\varepsilon \rightarrow 0$ , the variables  $V_i^*$  and  $V_i^{**}$  in equations (2.185) approximate the local equilibrium  $f(U_i^*)$  and  $f(U_i^{**})$ , respectively. Therefore, a second-order relaxed scheme is obtained as

$$\begin{aligned}
U_i^{(1)} &= U_i^n - \Delta t D_x V_i^n |_{V_i^n = f(U_i^n)}, \\
U_i^{(2)} &= U_i^{(1)} - \Delta t D_x V_i^{(1)} |_{V_i^{(1)} = f(U_i^{(1)})}, \\
U_i^{n+1} &= \frac{1}{2} (U_i^n + U_i^{(2)}).
\end{aligned} \tag{2.186}$$

In the following section, we will demonstrate numerical results based on these relaxation methods.

## 2.8 Numerical results based on relaxation schemes

Herein, we will present numerical results of problems that have been introduced in Section 2.5, applying the relaxation schemes.

### 2.8.1 Linear advection equation

Consider the Cauchy problem (2.140) and  $U(0, x) = x$ . The numerical and exact solutions presented in Figure 2.11 were obtained using the first- and second-order relaxation schemes. Results

are good approximation compared with finite volume results under the computational domain  $[0,3]$  and speed of propagation  $c = 1$ . Furthermore, the error-norms for the Cauchy problem (2.140) and  $U(0,x) = x$  are presented in Table 2.3.

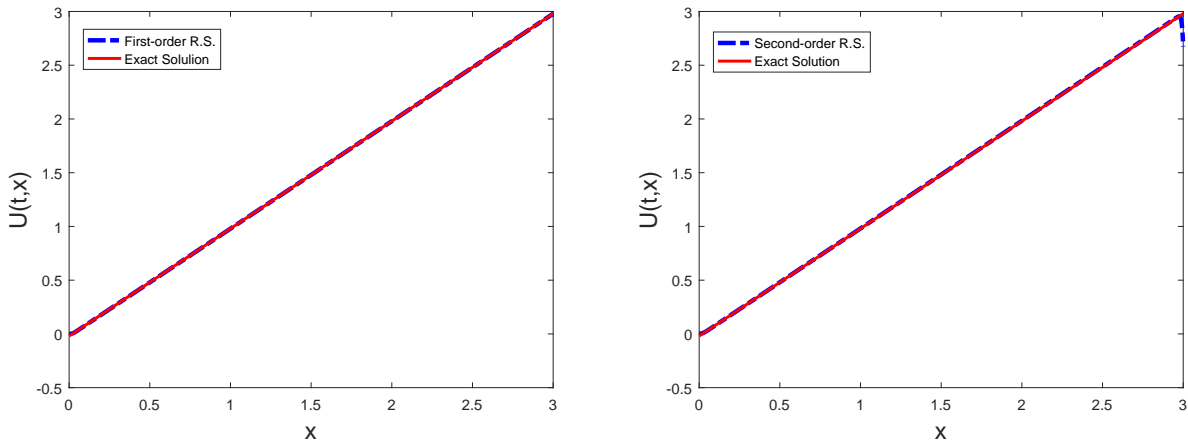


Figure 2.11: Numerical and exact solutions to the Cauchy problem (2.140) and  $U(0,x) = x$  using relaxation schemes: First-order (Left) and second-order (Right).

We assume that the initial piecewise data were given as in (2.142). Then, numerical and exact solutions to the Riemann problem (2.140) and (2.142) are posted in Figure 2.12 using the first- and second-order relaxation schemes. It is observed that both results show excellent approximations.

## 2.8.2 Burger's equation

To be more confident that the above schemes work appropriately, we solve the Burger's equation (2.144) with the smooth initial data  $U(0,x) = x$ . Figure 2.13 shows the comparison between the numerical solution found by first- and second-order relaxation schemes and the exact solution. Results have shown approximation significantly.

Considering the Riemann problem of Burger's equation (2.144), with the initial data as in (2.146), numerical and exact solutions are depicted in Figure 2.14. Results are in good agreement with the

N	Scheme	$L^1$ -error	$L^2$ -error	$L^\infty$ -error
100	1 <sup>st</sup> order	0.2254	0.0349	0.0269
	2 <sup>nd</sup> order	0.6900	0.7856	0.6605
200	1 <sup>st</sup> order	0.2728	0.0335	0.0249
	2 <sup>nd</sup> order	0.6671	0.4222	0.2577
300	1 <sup>st</sup> order	0.3809	0.0356	0.0239
	2 <sup>nd</sup> order	0.6512	0.1630	0.9834
400	1 <sup>st</sup> order	0.4284	0.0362	0.0232
	2 <sup>nd</sup> order	0.6456	0.9636	0.7842
500	1 <sup>st</sup> order	0.4760	0.0371	0.0227
	2 <sup>nd</sup> order	0.6740	0.8137	0.6428
600	1 <sup>st</sup> order	0.5844	0.0394	0.0223
	2 <sup>nd</sup> order	0.7130	0.6972	0.5380

Table 2.3: Error-norms for the Linear advection equation (2.144) with smooth initial data  $U(0, x) = x$  using relaxation schemes of different gridpoints 100, 200, 300, 400, 500 and 600.

exact solution, using the first- and second-order relaxation schemes. Furthermore, the error-norms for the Riemann problem (2.144) and (2.146) are presented in Table 2.4.

Moreover, consider the Riemann problem (2.144) and (2.148), the numerical solutions obtained by the first- and second-order relaxation schemes are shown in Figure 2.15, which are approximately close to the exact solution compared with finite volume methods.

### 2.8.3 Shallow water equations

Here, we consider the dam-break problem of the shallow water equations (2.149) and initial data (2.150). The water height and velocity results are demonstrated in Figure 2.16 using the first- and second-order relaxation schemes, with the computational domain  $[-10, 10]$ . However, results are in excellent agreement compared with one implemented by the finite volume method.

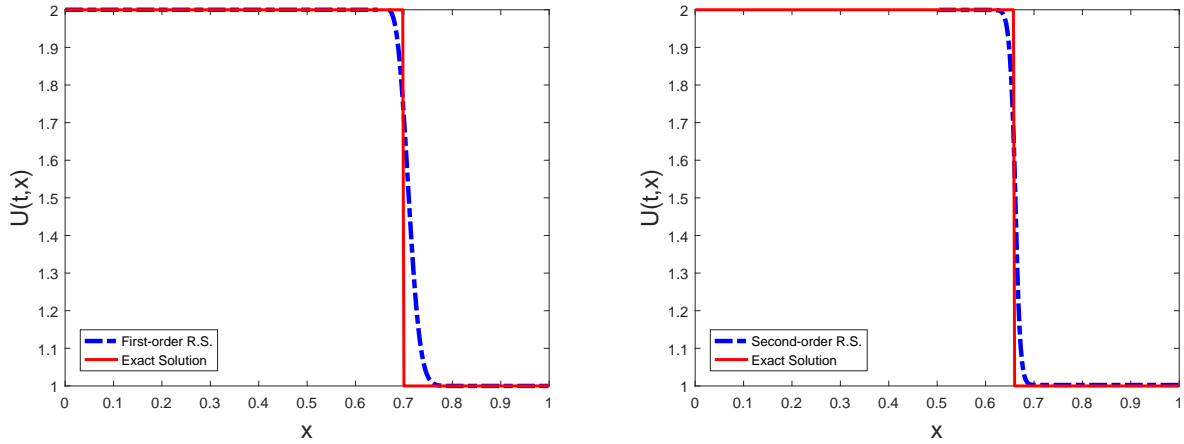


Figure 2.12: Solutions to the Riemann problem (2.140) and (2.142) using relaxation schemes: First-order (Left) and second-order (Right).

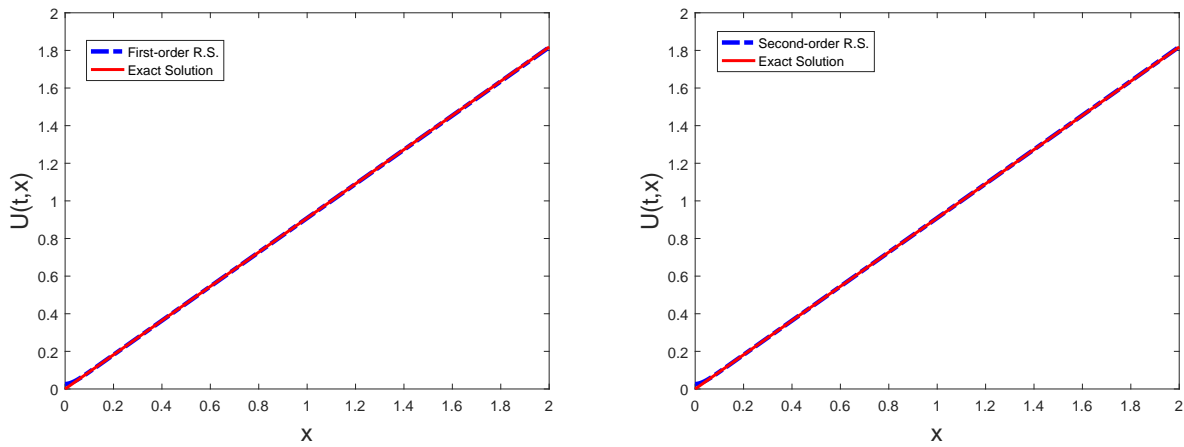


Figure 2.13: Numerical and exact solutions to the Cauchy problem (2.144) and  $U(0,x) = x$  using relaxation schemes: First-order (Left) and second-order (Right).

## 2.8.4 System of Euler equations

We assume the shock tube problem of the Euler equations (2.151) and initial data (2.152). Numerical results of the shock tube problem are presented in Figure 2.17 using the first- and second-order

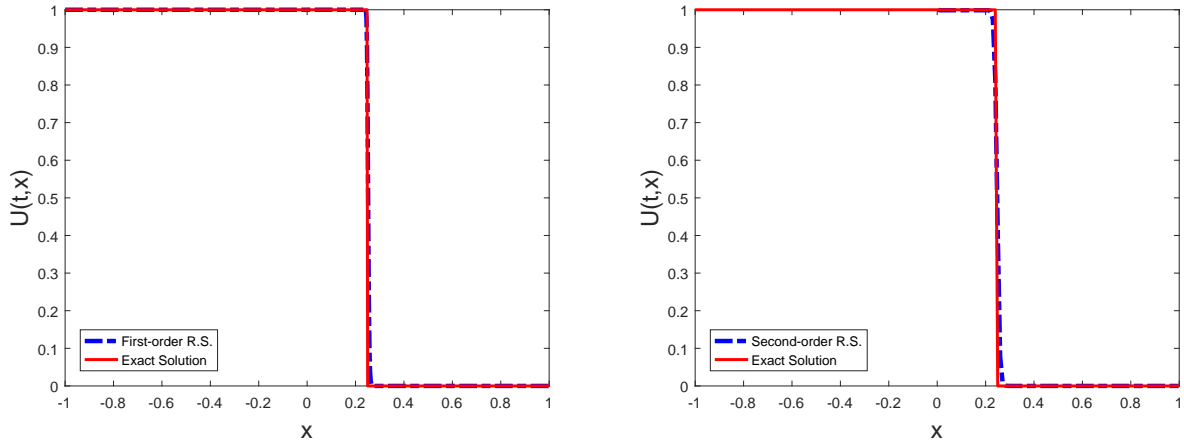


Figure 2.14: Numerical and exact solutions to the Riemann problem (2.144) and (2.146) using relaxation schemes: First-order (Left) and second-order (Right).

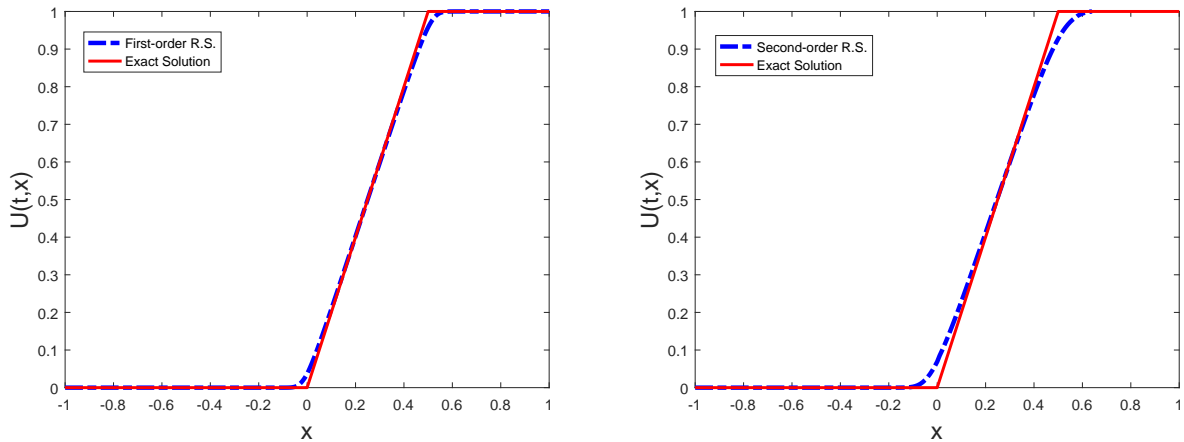


Figure 2.15: Numerical and exact solutions to the Riemann problem (2.144) and (2.148) using relaxation schemes: First-order (Left) and second-order (Right).

relaxation schemes. The results obtained using both orders are in excellent agreement and well approximations compared with the finite volume scheme results.



N	Scheme	$L^1$ -error	$L^2$ -error	$L^\infty$ -error
100	1 <sup>st</sup> order	0.9460	0.5772	0.5352
	2 <sup>nd</sup> order	0.7175	0.3504	0.2805
200	1 <sup>st</sup> order	0.7725	0.4384	0.3090
	2 <sup>nd</sup> order	0.8960	0.4065	0.3737
300	1 <sup>st</sup> order	0.8815	0.5236	0.4753
	2 <sup>nd</sup> order	0.8909	0.4678	0.4315
400	1 <sup>st</sup> order	0.8059	0.4462	0.3595
	2 <sup>nd</sup> order	0.9087	0.3323	0.2997
500	1 <sup>st</sup> order	0.8276	0.4792	0.4132
	2 <sup>nd</sup> order	0.9269	0.4124	0.3655
600	1 <sup>st</sup> order	0.8453	0.4731	0.4126
	2 <sup>nd</sup> order	0.9990	0.3743	0.3413

Table 2.4: Error-norms for the Burger’s equation (2.144) with piecewise initial data (2.146) using relaxation schemes of different gridpoints 100, 200, 300, 400, 500 and 600.

## 2.9 Discontinuous Galerkin method

The discontinuous Galerkin method (DGM) has also been introduced recently to approximate solutions to systems of conservation laws due to stability issues of classical finite element methods for hyperbolic conservation laws, see for example [98, 99]. The method involves entirely discontinuous basis functions across each element, and it can be considered as a combination of finite volume and finite element methods. DGM is a particular class of finite element methods, and they usually consist of piecewise polynomials defined locally. We can now briefly present this method for a Cauchy problem of the hyperbolic conservation law of the form

$$\frac{\partial U}{\partial t} + \frac{\partial f(U)}{\partial x} = 0, \quad x \in [a, b], \quad t \geq 0, \quad U(0, x) = \bar{U}(x), \quad (2.187)$$

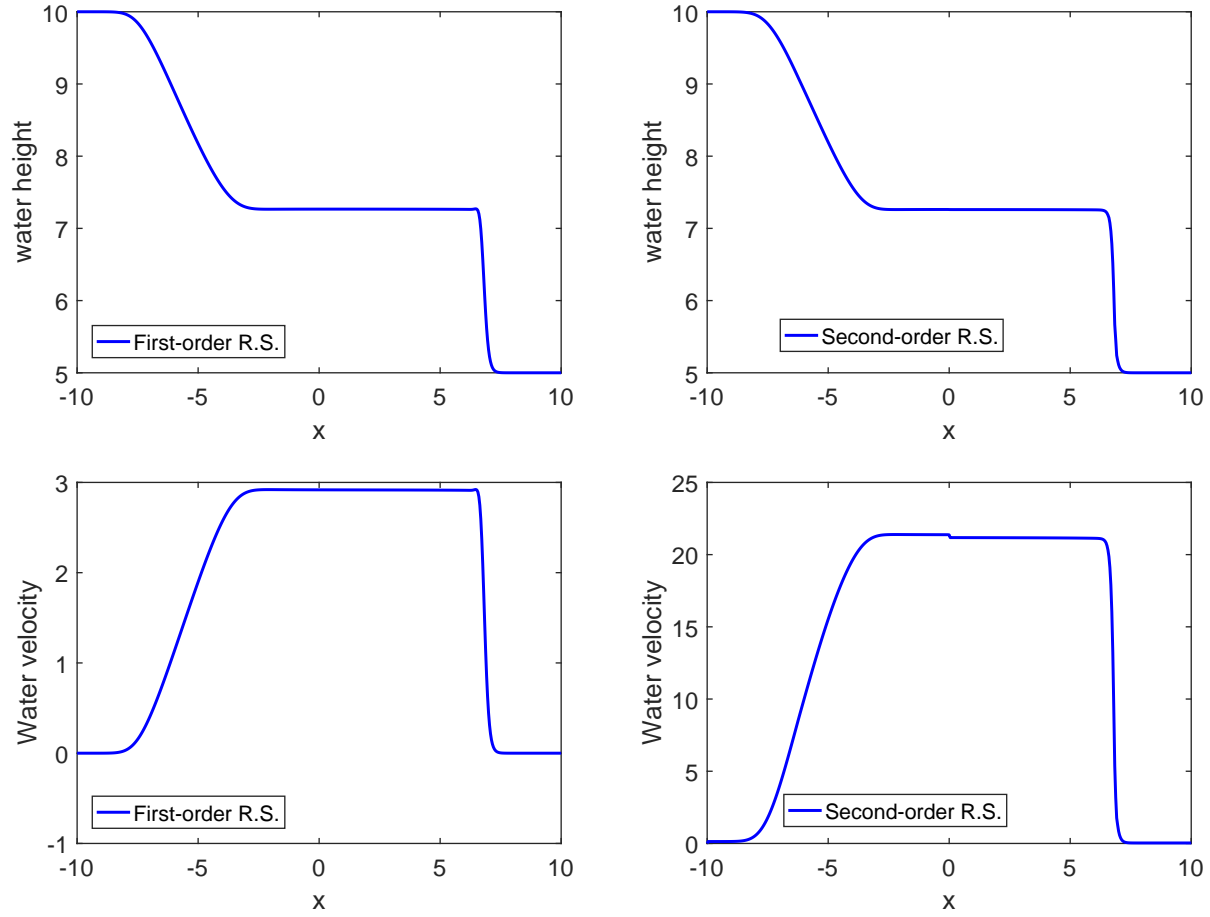


Figure 2.16: Numerical results for the dam break problem (2.149) and (2.150) using relaxation schemes: First-order (Left) and second-order (Right).

Motivated by the finite element framework, we present a space discretisation consisting of  $N$  cells  $I_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$  of length  $\Delta x_i = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$ . We consider that the mesh is regular, i.e there exists a constant  $c > 0$  independent of  $\Delta x$  such that

$$c\Delta x \leq \Delta x_i, \quad (2.188)$$

where  $\Delta x = \max_i \Delta x_i$ ,  $i = 1, \dots, N$ . Thus, we first assume a variational formulation of (2.187) in order to define the discontinuous Galerkin method. Let  $\phi$  be a smooth test function. Then, we

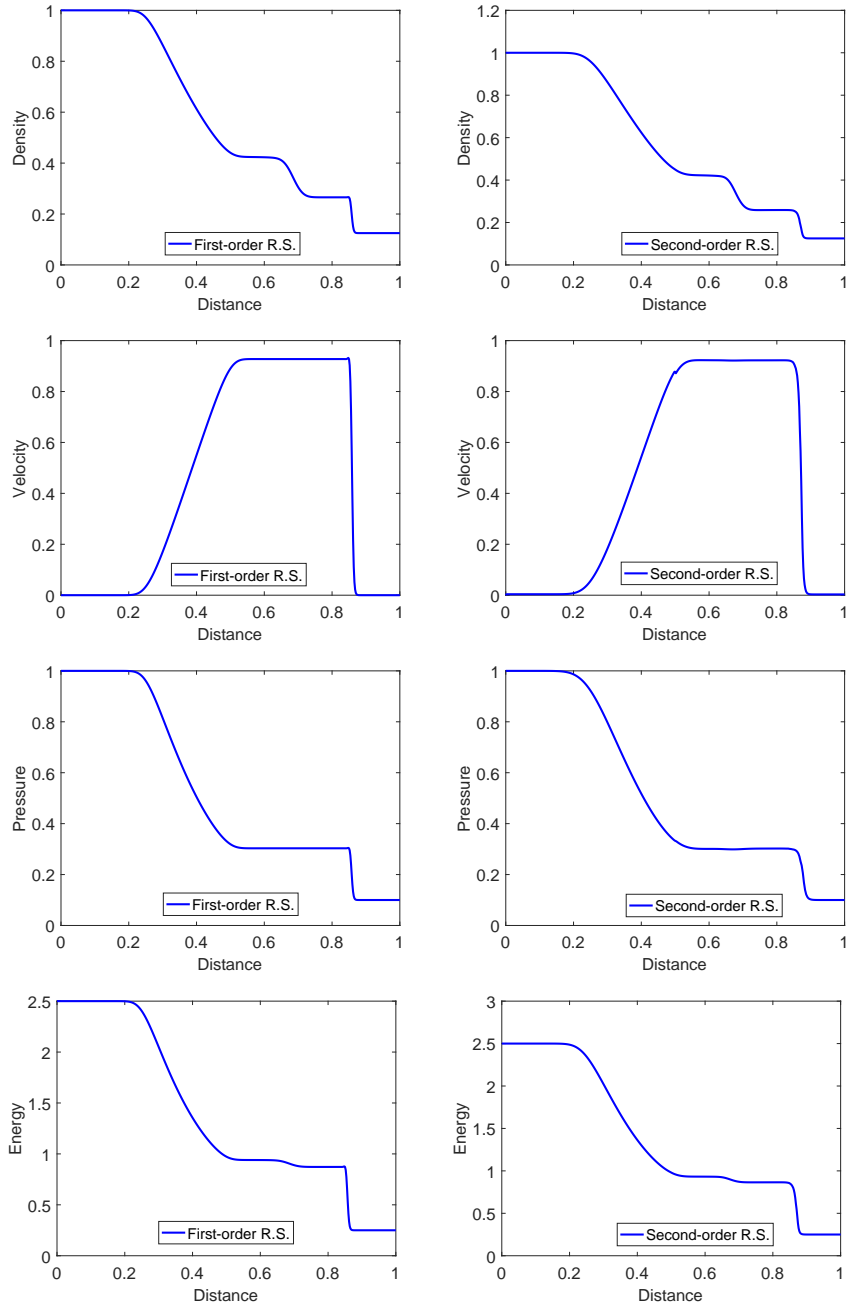


Figure 2.17: Numerical results for the Shock tube problem (2.151) and (2.152) using relaxation schemes: First-order (Left) and second-order (Right).

multiply both sides of (2.187) by  $\phi$  and integrate by parts over the cell  $I_i$ , we have

$$\begin{aligned}
& \int_{I_i} \left( \frac{\partial U}{\partial t} + \frac{\partial f(U)}{\partial x} \right) \phi dx \\
&= \int_{I_i} \frac{\partial U}{\partial t} \phi dx + \int_{I_i} \frac{\partial f(U)}{\partial x} \phi dx \\
&= \int_{I_i} \frac{\partial U}{\partial t} \phi dx - \int_{I_i} f(U) \frac{\partial \phi}{\partial x} dx + f(U_{i+\frac{1}{2}}) \phi_{i+\frac{1}{2}} - f(U_{i-\frac{1}{2}}) \phi_{i-\frac{1}{2}} = 0
\end{aligned} \tag{2.189}$$

Now assume both the solution  $U$  and the test function  $\phi$  come from a finite dimensional approximation space  $V_{\Delta x}$ , which is usually taken as the space of piecewise polynomials of degree up to  $r$ :

$$V_{\Delta x} = \{ \phi \in L^2([a, b]) : \phi|_{I_i} \in \mathbb{P}^r(I_i) \}, \tag{2.190}$$

where where  $\mathbb{P}^r$  represents the set of polynomials of degree up to  $r$  on  $I_i$ . However, the boundary terms like  $f(U_{i+\frac{1}{2}})$  and  $\phi_{i+\frac{1}{2}}$  are not well defined when  $U$  and  $\phi$  are in this space, as they are discontinuous at the cell interfaces.

Thus, we seek  $U \in V_{\Delta x}$  an approximate solution of (2.187), such that for all test functions  $\phi \in V_{\Delta x}$  satisfies the following, the semi-discrete discontinuous Galerkin method is defined as follows: find  $U \in V_{\Delta x}$  such that for all elements we have

$$\int_{I_i} \frac{\partial U}{\partial t} \phi dx - \int_{I_i} f(U) \frac{\partial \phi}{\partial x} dx + \hat{f}_{i+\frac{1}{2}} \phi_{i+\frac{1}{2}}^- - \hat{f}_{i-\frac{1}{2}} \phi_{i-\frac{1}{2}}^+ = 0, \quad \forall \phi \in V_{\Delta x}, \tag{2.191}$$

where  $\phi_{i-\frac{1}{2}}^+$  and  $\phi_{i+\frac{1}{2}}^-$  are Values from inside  $I_i$  for the test function  $\phi$ , and  $\hat{f}_{i+\frac{1}{2}}$  is the numerical flux that approximates the boundary terms of the system (2.187). The way we choose the numerical flux can be assumed as the key idea of DGM since it merge the classical finite element and the finite volume methods. In general, the numerical flux  $\hat{f}_{i+\frac{1}{2}}$  is defined as a two variable function which depends on the value of approximate solution  $U$  from both sides of the interface  $x_{i+\frac{1}{2}}$ , i.e.

$$\hat{f}_{i+\frac{1}{2}} = \hat{f}(U_{i+\frac{1}{2}}^-, U_{i+\frac{1}{2}}^+),$$

and it satisfy the properties of consistency  $\hat{f}(U, U) = f(U)$ , monotonicity  $\hat{f}(\uparrow, \downarrow)$  and  $\hat{f}(\cdot, \cdot)$  is Lipschitz continuous with respect to both arguments.

The consistency and the Lipschitz continuity are required to obtain a conservative scheme, while monotonicity ensures that the numerical schemes satisfy all entropy conditions and has the total variation diminishing property, as we pointed out in the previous sections for the finite volume methods.

For the time discretisation, we could done by applied the total variation diminishing (TVD) Runge-Kutta method, we can refer the reader to [100–102] and the references therein for more details.

## 2.10 Concluding remarks

In this chapter, we discussed the general theory and the numerical solutions of conservation laws. Because of the loss of regularity of classical solutions, we defined the notion of weak solutions, which are required to satisfy the equations in an integral form. In the framework of weak solutions, the uniqueness of the solution is lost and we used the concept of entropy to single out the unique physically relevant solution. The key property of these solutions is that at their point of the jump, the Rankine-Hugoniot condition has to be satisfied. Also, we reviewed different numerical methods for the approximations of the solutions of conservation laws. One of them is the finite volume methods that ensures the consistency of the numerical flux with the flux function of the original equation. The other is the relaxation method, which starts from a linear approximation of the equation with a stiff source term. The stability of the schemes is determined by the choice of the time step, according to the CFL conditions. Numerical examples are presented and are related to different Cauchy problems associated with the Burger's equation, the shallow water equation and the Euler equations. Also, we briefly presented the discontinuous Galerkin methods related to the finite element framework as another numerical approach that can be used to solve the conservation laws. In the next chapter, we consider optimal control problems constrained by scalar conservation laws.

# Chapter 3

## Optimal control of 1D Scalar conservation laws using the relaxation method

This chapter deals with optimal control problems constrained by one-dimensional hyperbolic systems of conservation laws. Because the flow generated by the nonlinear system of conservation laws is non-differentiable, we replace the scalar nonlinear conservation laws with the corresponding relaxation system in the problem formulation. In an adjoint approach, we derive the adjoint equations and the optimality condition at the continuous level. These are then discretised and we propose a numerical method for the solution to the optimal control problem. In brief, the methods iteratively solve the flow equations forward in time, the adjoint equation backwards in time and we use the optimality condition to update the control variable until convergence is achieved. The notion of the generalised tangent vectors approach is introduced. We implemented our numerical method for some flow matching problems related to the advection and Burger equations.

### 3.1 Introduction

In this chapter, we consider optimal control problems constrained by hyperbolic conservation laws. The problem is formulated as

$$\min_{U_0} J(U(T, \cdot); U_d), \quad (3.1)$$

where  $J(\cdot)$  is a given cost functional defined, for example, in the case of flow matching as

$$J(U(T, \cdot); U_d) = \frac{1}{2} \int_{\mathbb{R}} \|U(T, x) - U_d(x)\|^2 dx, \quad x \in \mathbb{R}. \quad (3.2)$$

Here,  $U_d$  is the desired state of the system, which depend on the space variable and  $U$  is the unique entropy solution of the conservation law

$$\begin{aligned} \frac{\partial U}{\partial t} + \frac{\partial f(U)}{\partial x} &= 0, \quad (t, x) \in [0, T] \times \mathbb{R}, \\ U(0, x) &= U_0(x), \quad x \in \mathbb{R}, \end{aligned} \quad (3.3)$$

where  $U : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$  represents the conserved quantity,  $f : \mathbb{R} \rightarrow \mathbb{R}$  is the corresponding flux function that is in general, nonlinear and the initial data  $U_0(x)$  is an arbitrary bounded measurable function defined on  $\mathbb{R}$ .

This chapter aims to investigate the solution of problem (3.1) using an adjoint-based approach. We use a relaxation method to derive optimality systems instead of the non-linear hyperbolic equations (3.3), which is replaces the non-linear systems with semi-linear equations with linear flux functions and stiff non-linear source terms. Relaxation approximations, commonly known as Jin and Xin relaxation approximations, were first discussed in [9]. Numerous discussions on such approximations have been emerging [81, 82, 103–105]. The idea of the relaxation approximation is to present a relaxation scheme to a system of conservation laws that can generate an entropy solution in the zero relaxation limit while preserving the hyperbolic structure at the expense of additional source terms. The convergence analysis for the relaxation approximation can be found in [56–58].

It is important to note that solutions of the system (3.3) may not be possible to obtain explicitly and develop discontinuities in finite time even for smooth initial data due to the nonlinearity, see the previous chapter, Chapter 2 and [4–8]. In general, the semi-group generated by a conservation law is non-differentiable in  $L^1$ , even in the scalar one-dimensional case. A differential structure on general  $BV$ – solutions for hyperbolic conservation laws in one space dimension has been discussed in [13–15, 106, 107].

The discontinuities (shock waves) that occur in the solution of hyperbolic conservation laws pose challenges in both analytical and numerical solutions of the optimisation problems [32, 46, 108–110]. Pfaff and Ulbrich in [111] proposed the optimal control of scalar conservation laws based



on the switching times of an on/off control. They presented the differentiability properties of the control-to-state mapping for entropy solutions and investigated the differentiability of the reduced cost functional in the presence of shocks. In [109], Lecaros and Zuazua presented the sensitivity analysis of shocks that appear in the solution to the optimisation problem. They compared the performance between the discrete approach and the alternating descent methods for the 1D scalar case. Analysis of the existence and uniqueness of the solution for the optimal control problems have been described by Ulbrich in [106, 112], Aguilar et al. in [38] and Hajian et al. [41] for one-dimensional space.

The optimisation problems based on relaxation approximations to the hyperbolic system of conservation laws have been extensively studied in the literature [65–67, 70, 72, 113]. They used both continuous and discrete approaches. In [65], Yohana and Banda investigated the optimisation problems constrained by hyperbolic conservation laws with high-order relaxation approaches. Albi et al. [72] studied the optimal control problems with Linear multistep methods for the hyperbolic relaxation systems. The accuracy for Adams-Moulton and Adams-Bashford methods for optimal control problems for ordinary differential equations and the extension to the semi-lagrangian discretisations of hyperbolic relaxation systems were also considered. In [66], Steffensen et al. are interested in the optimisation problem based on the continuous and discretised relaxation schemes for the 1D scalar conservation laws. Numerical results on tracking type problems with non-smooth desired states and convergence were presented. A numerical algorithm of solving optimal control problems governed by hyperbolic conservation laws based on the Tangent vectors approach has been presented by Herty and Piccoli in [71]. The derivation was also combined with the relaxation method to resolve the evolution of the tangent vectors. In this chapter, we consider the optimal control problems constrained with the relaxation approximation to the conservation laws. The structure allows us to derive a simple and efficient optimisation algorithm for solving the optimal control problems numerically, without using Riemann solvers or non-linear systems of algebraic equations.

This chapter is organised as follows: Formulation of the optimal control problem constrained with relaxation systems of hyperbolic conservation laws is presented in Section 3.2. In Section 3.3, the

derivation of optimality conditions of an unconstrained problem based on the relaxation method is introduced. The optimisation algorithm for solving optimal control problems is stated in Section 3.4. The discretisation of the adjoint equations and relaxation schemes are displayed in Section 3.5. Numerical results and discussion are demonstrated in Section 3.6 related to the linear advection and Burger's equations. The notion of the tangent vectors approach is discussed briefly in Section 3.7 to improve the gradient descent method and numerical analysis with the algorithm to solve the problem numerically. Finally, the concluding remarks of this chapter are reported in Section 3.8.

## 3.2 Problem formulation

We are interested in optimal control problems governed by the relaxed systems of conservation laws of the form

$$\begin{aligned} \frac{\partial}{\partial t}U + \frac{\partial}{\partial x}V &= 0, \\ \frac{\partial}{\partial t}V + a^2 \frac{\partial}{\partial x}U &= -\frac{1}{\varepsilon}(V - f(U)), \\ U(0, x) &= U_0(x), \quad V(0, x) = f(U_0(x)). \end{aligned} \tag{3.4}$$

A full description of the system (3.4) has been given in Section 2.6 of Chapter 2. Moreover, the relaxation system (3.4) converges to the scalar conservation laws (3.3) if and only if the sub-characteristic condition given by

$$\max_U |(f'(U))| \leq a, \tag{3.5}$$

is satisfied [9]. As a prototype, the problem is formulated as

$$\min_{U_0} J(U(T, \cdot); U_d) \quad \text{subject to} \quad \begin{cases} \frac{\partial}{\partial t}U + \frac{\partial}{\partial x}V = 0, \\ \frac{\partial}{\partial t}V + a^2 \frac{\partial}{\partial x}U = -\frac{1}{\varepsilon}(V - f(U)), \\ U(0, x) = U_0(x), \quad V(0, x) = f(U_0(x)), \end{cases} \tag{3.6}$$

where  $J(\cdot)$ , the cost functional of tracking type has given in (3.2). The initial data  $U_0$  acts as a control function. Updating  $U_0$  produces optimal solutions,  $U(T, \cdot)$ , that matches a given desired state,  $U_d$  at the final time  $T$ .

### 3.3 Derivation of an optimality system

Given a subset  $\Omega$  of  $\mathbb{R}$ , to solve the optimisation problem (3.6) by using the method known as the adjoint-based approach, we introduce the Lagrangian

$$\begin{aligned}
L(\cdot) &= L(U(T, \cdot), p, q; U_d) = J(U(T, \cdot); U_d) \\
&+ \int_0^T \int_{\Omega} p \left( \frac{\partial}{\partial t} U + \frac{\partial}{\partial x} V \right) dx dt \\
&+ \int_0^T \int_{\Omega} q \left( \frac{\partial}{\partial t} V + a^2 \frac{\partial}{\partial x} U + \frac{1}{\varepsilon} (V - f(U)) \right) dx dt,
\end{aligned} \tag{3.7}$$

where  $p, q \in \Omega$  are co-state (adjoint) variables, which are assumed to be smooth functions with compact support in  $\Omega$  and  $p$  and  $q$  vanish on the boundaries  $\partial\Omega$  of  $\Omega$ . To derive the formal first-order optimality system, we set the first variations of  $L(\cdot)$  with respect to each of the functions  $p, q, U, V$  and  $U_0$  equal to zero. Setting the first partial derivative of  $L(\cdot)$  in (3.7) with respect to  $p$  and  $q$  equal to zero, we have the relaxation system (3.4). Furthermore, since  $p$  and  $q$  are smooth functions and by using integration by parts, we have

$$\begin{aligned}
L(\cdot) &= J(U(T, \cdot); U_d) + \int_{\Omega} (pU)|_0^T dx - \int_{\Omega} \int_0^T U \frac{\partial}{\partial t} p dt dx \\
&+ \int_0^T (pV)|_{\partial\Omega} dt - \int_0^T \int_{\Omega} V \frac{\partial}{\partial x} p dx dt + \int_{\Omega} (qV)|_0^T dx \\
&- \int_{\Omega} \int_0^T V \frac{\partial}{\partial t} q dt dx + a^2 \int_0^T (qU)|_{\partial\Omega} dt - a^2 \int_0^T \int_{\Omega} U \frac{\partial}{\partial x} q dx dt \\
&+ \frac{1}{\varepsilon} \int_0^T \int_{\Omega} qV dx dt - \frac{1}{\varepsilon} \int_0^T \int_{\Omega} qf(U) dx dt.
\end{aligned} \tag{3.8}$$

Also, since  $U$  and  $V$  vanish at the boundaries of  $\Omega$ , then

$$\begin{aligned}
L(\cdot) &= J(U(T, \cdot); U_d) + \int_{\Omega} [p(T, x)U(T, x) - p(0, x)U(0, x)] dx \\
&+ \int_0^T \int_{\Omega} U \left[ -\frac{\partial p}{\partial t} - a^2 \frac{\partial q}{\partial x} - \frac{1}{\varepsilon} qf'(U) \right] dx dt \\
&+ \int_{\Omega} [q(T, x)V(T, x) - q(0, x)V(0, x)] dx \\
&+ \int_0^T \int_{\Omega} V \left[ -\frac{\partial p}{\partial x} - a^2 \frac{\partial q}{\partial t} + \frac{1}{\varepsilon} q \right] dx dt
\end{aligned} \tag{3.9}$$

Setting the first partial derivatives of  $L(\cdot)$  in (3.9) with respect to  $U$  and  $V$  equal to zero, respectively, gives

$$\frac{\partial L(\cdot)}{\partial U} = \int_0^T \int_{\Omega} \left[ -\frac{\partial}{\partial t} p - a^2 \frac{\partial}{\partial x} q - \frac{1}{\varepsilon} q f'(U) \right] dx dt = 0, \quad (3.10)$$

$$\frac{\partial L(\cdot)}{\partial V} = \int_0^T \int_{\Omega} \left[ -\frac{\partial}{\partial x} p - \frac{\partial}{\partial t} q + \frac{1}{\varepsilon} q \right] dx dt = 0, \quad (3.11)$$

where  $f'(U)$  is the derivative of the flux function  $f(U)$ . Then, we have the adjoint system in the form

$$\begin{aligned} -\frac{\partial}{\partial t} p - a^2 \frac{\partial}{\partial x} q &= \frac{1}{\varepsilon} f'(U) q, & p(t=T, x) &= p_T(x), \\ -\frac{\partial}{\partial t} q - \frac{\partial}{\partial x} p &= -\frac{1}{\varepsilon} q, & q(t=T, x) &= q_T(x). \end{aligned} \quad (3.12)$$

The terminal conditions  $p_T(x)$  and  $q_T(x)$  can be obtained by setting the partial derivatives of  $L(\cdot)$  in (3.9) with respect to  $U(T, x) = U_T$  and  $V(T, x) = V_T$  equal to zero, which gives

$$\begin{aligned} p_T(x) &= U(T, x) - U_d(x), \\ q_T(x) &= 0. \end{aligned} \quad (3.13)$$

An expansion in terms of  $\varepsilon$  results in the solution  $(p, q)$  that solves the second-order equation [67, 70] of the form

$$-\frac{\partial}{\partial t} p - f'(U) \frac{\partial}{\partial x} p = \varepsilon a^2 \frac{\partial^2}{\partial x^2} p. \quad (3.14)$$

Therefore, setting the partial derivative of  $L(\cdot)$  in (3.9) with respect to  $U_0$  equal to zero gives the optimality condition

$$\frac{\partial L(\cdot)}{\partial U_0} = \frac{\tilde{J}(\cdot)}{\partial U_0} + \int_{\Omega} [-p(0, x) - (q(0, x) f'(U(0, x)))] dx = 0, \quad (3.15)$$

where  $\tilde{J}(\cdot) = \tilde{J}(U_0; U_d)$  is the reduced cost functional. Therefore, the gradient of the reduced cost functional can be given as

$$\nabla_{U_0} \tilde{J}(\cdot) = \int_{\Omega} p(0, x) + f'(U(0, x)) q(0, x) dx. \quad (3.16)$$

## 3.4 Numerical algorithm

The first-order optimality conditions presented in the previous Section 3.3 serve as a basis for a numerical algorithm for the solution of optimal control problems governed by relaxation systems of conservation laws. These conditions include the relaxation system (3.4) with initial data, the adjoint system (3.12) with terminal data (3.13) and the gradient (3.16) where (3.4) should be solved forward in time (from  $t = 0$  to  $t = T$ ) and the adjoint system (3.12) should be solved backwards in time (from  $t = T$  to  $t = 0$ ). Therefore, the algorithm can be seen as a generalisation of the so-called forward-backwards sweep algorithm [32, 46, 65] of optimal control problems. The steps of the algorithm are:

1. Generate an initial guess for the initial data  $U_0$ .
2. Numerically solve the hyperbolic relaxation system (3.4) with the initial data  $U_0$  and  $V_0 = f(U_0)$  forward in time for the state variable  $U$ .
3. Solve the adjoint system (3.12) subject to the terminal data (3.13) backwards in time for adjoint variables  $p$  and  $q$ .
4. Use the adjoint variable and the control variable  $U_0$  to evaluate the gradient of the cost functional (3.2),  $\nabla_{U_0} \tilde{J}(\cdot)$  as in (3.16).

5. Update the control  $U_0$  using

$$U_0^{m+1} = U_0^m - \alpha \nabla_{U_0} \tilde{J}(\cdot),$$

for some stepsize  $\alpha$ ,  $m = 0, 1, \dots$ ; See, for example, [114] for more details.

6. Repeat Steps 2 to 5 until convergence is achieved.

## 3.5 Discretisation techniques

In this section, we present the discretisation of the optimal control problem (3.6). In Step 2 of the numerical algorithm 3.4, the relaxation system (3.4) of conservation law with an initial data can be

solved using the numerical schemes that have been presented in Section 2.7 of Chapter 2. Following Step 3 of the numerical algorithm, a similar discretisation approach of the adjoint system (3.12) will be considered. Note that the adjoint system must be solved backwards in time. The discretisation processes for both spatial and time are achieved separately using the method of lines [96]. For convenience, we briefly introduce spatial grid points  $x_{i+\frac{1}{2}}$  with grid spacing  $\Delta x_i = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$  and the uniform discrete time step  $\Delta t = t^{n+1} - t^n$ , where  $n = 0, 1, \dots, N$ . Following the approach introduced by Banda and Herty in [67], we can rewrite the adjoint system (3.12) as the linear transport equations by the form

$$\begin{aligned}\frac{\partial}{\partial t}p + a^2 \frac{\partial}{\partial x}q &= 0, \\ \frac{\partial}{\partial t}q + \frac{\partial}{\partial x}p &= 0,\end{aligned}\tag{3.17}$$

and the stiff ordinary differential equations (ODE)

$$\begin{aligned}\frac{\partial}{\partial t}p &= -\frac{1}{\varepsilon}f'(U)q, \\ \frac{\partial}{\partial t}q &= \frac{1}{\varepsilon}q.\end{aligned}\tag{3.18}$$

Therefore, the characteristic variables  $p \pm aq$  of an adjoint system as it is shown in [65, 67] satisfy

$$\begin{aligned}\frac{\partial}{\partial t}(p \pm aq) \pm a^2 \frac{\partial}{\partial x}(p \pm aq) &= 0, \\ -\frac{\partial}{\partial t}(-p \mp a(-q)) \pm a^2 \frac{\partial}{\partial x}(-p \pm a(-q)) &= 0.\end{aligned}\tag{3.19}$$

### 3.5.1 Spatial discretisation

#### 1. First-order discretisation.

Since the adjoint system (3.12) is solved backwards in time, hence an upwind discretisation for the transport system (3.17) advects  $p + aq$  and  $-p + a(-q)$ , respectively with velocity  $\pm a$ . Moreover, we have

$$\begin{aligned}(p + aq)_{i+\frac{1}{2}} &= p_{i+1} + aq_{i+1}, \\ (p - aq)_{i+\frac{1}{2}} &= p_i - aq_i.\end{aligned}\tag{3.20}$$

We can solve (3.20) for  $p_{i+\frac{1}{2}}$  and  $q_{i+\frac{1}{2}}$ , to obtain

$$\begin{aligned} p_{i+\frac{1}{2}} &= \frac{1}{2}(p_{i+1} + p_i) + \frac{a}{2}(q_{i+1} - q_i), \\ q_{i+\frac{1}{2}} &= \frac{1}{2}(q_{i+1} + q_i) + \frac{1}{2a}(p_{i+1} - p_i). \end{aligned} \quad (3.21)$$

Substituting  $p$  by  $-p$  and  $q$  by  $-q$  in equations (3.21), we obtain the expression

$$\begin{aligned} p_{i+\frac{1}{2}} &= -\frac{1}{2}(p_{i+1} + p_i) - \frac{a}{2}(q_{i+1} - q_i), \\ q_{i+\frac{1}{2}} &= -\frac{1}{2}(q_{i+1} + q_i) - \frac{1}{2a}(p_{i+1} - p_i). \end{aligned} \quad (3.22)$$

Furthermore, we can introduce the discrete derivative of  $p_i$  as

$$D_x p_i = \frac{1}{\Delta x}(p_{i+\frac{1}{2}} - p_{i-\frac{1}{2}}). \quad (3.23)$$

Thus, the discrete version of the adjoint system (3.12) in the conservative formula can be written as

$$\begin{aligned} \frac{\partial p_i}{\partial t} + a^2 D_x q_i &= -\frac{1}{\varepsilon} f'(U_i) q_i, \\ \frac{\partial q_i}{\partial t} + D_x p_i &= \frac{1}{\varepsilon} q_i. \end{aligned} \quad (3.24)$$

The approximation (3.24) has an order of accuracy of  $O(\overline{\Delta x}^2)$  [65]. Equations (3.22) - (3.24) give the first-order discrete scheme of the adjoint system (3.12) as

$$\begin{aligned} \frac{\partial p_i}{\partial t} - \frac{a^2}{2\Delta x}(q_{i+1} - q_{i-1}) - \frac{a}{2\Delta x}(p_{i+1} - 2p_i + p_{i-1}) &= -\frac{1}{\varepsilon} f'(U_i) q_i, \\ \frac{\partial q_i}{\partial t} - \frac{1}{2\Delta x}(p_{i+1} - p_{i-1}) - \frac{a}{2\Delta x}(q_{i+1} - 2q_i + q_{i-1}) &= \frac{1}{\varepsilon} q_i. \end{aligned} \quad (3.25)$$

## 2. Second-order discretisation.

The second-order in the spatial discretisation of the adjoint equations (3.12) can be presented similarly to the first-order scheme. Consider the characteristics variables  $p \pm aq$  of the adjoint system and recall the approximation (3.21). Also, using the polynomial  $\Upsilon$  of the form

$$\Upsilon_i(x; \varphi) = \varphi_i + \sigma(\varphi_i)(x - x_i), \quad (3.26)$$

where  $\varphi_i^- = p_i - aq_i$ ,  $\varphi_i^+ = p_i + aq_i$  and  $\sigma$  as introduced below.

Equation (3.26) allows defining characteristic variables in terms of a slope limiter for the second-order schemes as

$$\begin{aligned}(p + aq)_{i+\frac{1}{2}} &= (p + aq)_{i+\frac{1}{2}}^+ = \Upsilon_{i+1}(x_{i+\frac{1}{2}}; \varphi^+), \\ (p - aq)_{i+\frac{1}{2}} &= (p - aq)_{i+\frac{1}{2}}^- = \Upsilon_i(x_{i+\frac{1}{2}}; \varphi^-).\end{aligned}\tag{3.27}$$

Solving (3.27) for  $p_{i+\frac{1}{2}}$  and  $q_{i+\frac{1}{2}}$ , we obtain

$$\begin{aligned}p_{i+\frac{1}{2}} &= \frac{1}{2} \left[ \Upsilon_{i+1}(x_{i+\frac{1}{2}}; \varphi^+) + \Upsilon_i(x_{i+\frac{1}{2}}; \varphi^-) \right], \\ q_{i+\frac{1}{2}} &= \frac{1}{2a} \left[ \Upsilon_{i+1}(x_{i+\frac{1}{2}}; \varphi^+) - \Upsilon_i(x_{i+\frac{1}{2}}; \varphi^-) \right],\end{aligned}\tag{3.28}$$

where the superscripts + and - correspond to the right and left cell of a cell boundary at  $x_{i+\frac{1}{2}}$  respectively and

$$\varphi_i^+ = p_i + aq_i, \quad \varphi_i^- = p_i - aq_i.$$

Thus, the second-order terms (3.28) can be re-formulated as follows

$$\begin{aligned}p_{i+\frac{1}{2}} &= \frac{1}{2} \left[ \Upsilon_{i+1}(x_{i+\frac{1}{2}}; \varphi^+) + \Upsilon_i(x_{i+\frac{1}{2}}; \varphi^-) \right] \\ &= \frac{1}{2} \left[ \varphi_{i+1}^+ + \sigma(\varphi_{i+1}^+)(x_{i+\frac{1}{2}} - x_{i+1}) + \varphi_i^- + \sigma(\varphi_i^-)(x_{i+\frac{1}{2}} - x_i) \right] \\ &= \frac{1}{2} \left[ \varphi_{i+1}^+ - \frac{1}{2}\sigma(\varphi_{i+1}^+) + \varphi_i^- + \frac{1}{2}\sigma(\varphi_i^-) \right] \\ &= \frac{1}{2} \left[ p_{i+1} + aq_{i+1} - \frac{1}{2}\sigma_{i+1}^+ + p_i - aq_i + \frac{1}{2}\sigma_i^- \right] \\ &= \frac{1}{2} \left[ p_{i+1} + p_i + a(q_{i+1} - q_i) + \frac{1}{2}(\sigma_i^- - \sigma_{i+1}^+) \right].\end{aligned}\tag{3.29}$$

Similarly,

$$\begin{aligned}q_{i+\frac{1}{2}} &= \frac{1}{2a} \left[ \Upsilon_{i+1}(x_{i+\frac{1}{2}}; \varphi^+) - \Upsilon_i(x_{i+\frac{1}{2}}; \varphi^-) \right] \\ &= \frac{1}{2a} \left[ \varphi_{i+1}^+ + \sigma(\varphi_{i+1}^+)(x_{i+\frac{1}{2}} - x_{i+1}) - \varphi_i^- - \sigma(\varphi_i^-)(x_{i+\frac{1}{2}} - x_i) \right] \\ &= \frac{1}{2a} \left[ \varphi_{i+1}^+ - \frac{1}{2}\sigma(\varphi_{i+1}^+) - \varphi_i^- - \frac{1}{2}\sigma(\varphi_i^-) \right] \\ &= \frac{1}{2a} \left[ p_{i+1} + aq_{i+1} - \frac{1}{2}\sigma_{i+1}^+ - p_i + aq_i - \frac{1}{2}\sigma_i^- \right] \\ &= \frac{1}{2a} \left[ p_{i+1} - p_i + a(q_{i+1} + q_i) - \frac{1}{2}(\sigma_{i+1}^+ + \sigma_i^-) \right].\end{aligned}\tag{3.30}$$



Substituting  $-p$  for  $p$  and  $-q$  for  $q$  in equations (3.29) and (3.30), we obtain

$$\begin{aligned} p_{i+\frac{1}{2}} &= -\frac{1}{2}(p_{i+1} + p_i) - \frac{a}{2}(q_{i+1} - q_i) + \frac{1}{4}(\sigma_i^- - \sigma_{i+1}^+), \\ q_{i+\frac{1}{2}} &= -\frac{1}{2}(q_{i+1} + q_i) - \frac{1}{2a}(p_{i+1} - p_i) - \frac{1}{4a}(\sigma_{i+1}^+ + \sigma_i^-). \end{aligned} \quad (3.31)$$

Where

$$\begin{aligned} \sigma_{i+1}^+ &= (\varphi_{i+2}^+ - \varphi_{i+1}^+) \Psi(\vartheta_{i+1}^+) = (p_{i+2} + aq_{i+2} - p_{i+1} - aq_{i+1}) \Psi(\vartheta_{i+1}^+), \\ \sigma_i^- &= (\varphi_{i+1}^- - \varphi_i^-) \Psi(\vartheta_i^-) = (p_{i+1} - aq_{i+1} - p_i + aq_i) \Psi(\vartheta_i^-), \end{aligned} \quad (3.32)$$

and

$$\Psi(\vartheta_i^\pm) = \Psi\left(\frac{\varphi_i^\pm - \varphi_{i-1}^\pm}{\varphi_{i+1}^\pm - \varphi_i^\pm}\right) = \Psi\left(\frac{p_i \pm aq_i - p_{i-1} \mp aq_{i-1}}{p_{i+1} \pm aq_{i+1} - p_i \mp aq_i}\right). \quad (3.33)$$

Furthermore, from equations (3.32), (3.33) and expressions of  $\varphi_i^+$  and  $\varphi_i^-$ , we have

$$\begin{aligned} \sigma_{i+1}^+ &= (-p_{i+2} - aq_{i+2} + p_{i+1} + aq_{i+1}) \Psi\left(\frac{-p_{i+1} - aq_{i+1} + p_i + aq_i}{-p_{i+2} - aq_{i+2} + p_{i+1} + aq_{i+1}}\right) \\ &= [-(p_{i+2} + aq_{i+2}) - (-(p_{i+1} + aq_{i+1}))] \\ &\quad \times \Psi\left(\frac{-(p_{i+1} + aq_{i+1}) - (-(p_i + aq_i))}{-(p_{i+2} + aq_{i+2}) - (-(p_{i+1} + aq_{i+1}))}\right), \\ \sigma_i^- &= (-p_{i+1} + aq_{i+1} + p_i - aq_i) \Psi\left(\frac{-p_i + aq_i + p_{i-1} - aq_{i-1}}{-p_{i+1} + aq_{i+1} + p_i - aq_i}\right) \\ &= [-(p_{i+1} - aq_{i+1}) - (-(p_i - aq_i))] \\ &\quad \times \Psi\left(\frac{-(p_i - aq_i) - (-(p_{i-1} - aq_{i-1}))}{-(p_{i+1} - aq_{i+1}) - (-(p_i - aq_i))}\right). \end{aligned} \quad (3.34)$$

We can rewrite equations (3.31) to conform with the format of the adjoint system as

$$\begin{aligned} p_{i+\frac{1}{2}} &= -\left[\frac{1}{2}(p_{i+1} + p_i) + \frac{a}{2}(q_{i+1} - q_i) - \frac{1}{4}(\sigma_i^- - \sigma_{i+1}^+)\right], \\ q_{i+\frac{1}{2}} &= -\left[\frac{1}{2}(q_{i+1} + q_i) + \frac{1}{2a}(p_{i+1} - p_i) + \frac{1}{4a}(\sigma_{i+1}^+ + \sigma_i^-)\right]. \end{aligned} \quad (3.35)$$

Finally, equations (3.23), (3.24) and (3.35) give the second-order discrete scheme of the adjoint system (3.12) as

$$\begin{aligned} \frac{\partial p_i}{\partial t} - \frac{a^2}{2\Delta x}(q_{i+1} - q_{i-1}) - \frac{a}{2\Delta x}(p_{i+1} - 2p_i + p_{i-1}) - \frac{a}{4}(\sigma_{i+1}^+ - (\sigma_i^+ - \sigma_i^-) - \sigma_{i-1}^-) &= -\frac{1}{\varepsilon}f'(U_i)q_i, \\ \frac{\partial q_i}{\partial t} - \frac{1}{2\Delta x}(p_{i+1} - p_{i-1}) - \frac{a}{2\Delta x}(q_{i+1} - 2q_i + q_{i-1}) - \frac{1}{4}(\sigma_{i+1}^+ - (\sigma_i^- + \sigma_i^+) + \sigma_{i-1}^-) &= \frac{1}{\varepsilon}q_i. \end{aligned} \quad (3.36)$$

### 3.5.2 Time integration

In this subsection, we proposed a time discretisation of (3.36) following closely the method proposed in [67]. Finite-difference approximations (3.22) and (3.35) are incorporated in this discretisation process to obtain a full first- and second-order discrete scheme, respectively.

#### 1. First-order scheme.

The first-order TVD Runge-Kutta time discretisation of the adjoint equations (3.12) takes the form

$$\begin{aligned}
 p_i^* &= p_i^{n+1} - \frac{\Delta t}{\varepsilon} f'(U_i^{n+1}) q_i^{n+1}, \\
 q_i^* &= q_i^{n+1} + \frac{\Delta t}{\varepsilon} q_i^{n+1}; \\
 p_i^{(1)} &= p_i^* - \Delta t a^2 D_x q_i^*, \\
 q_i^{(1)} &= q_i^* - \Delta t D_x p_i^*; \\
 p_i^n &= p_i^{(1)}, \\
 q_i^n &= q_i^{(1)}.
 \end{aligned} \tag{3.37}$$

Moreover, we can use the approximations (3.22) and (3.23) to rewrite the first-order scheme (3.37) explicitly as

$$\begin{aligned}
 p_i^* &= p_i^{n+1} - \frac{\Delta t}{\varepsilon} f'(U_i^{n+1}) q_i^{n+1}, \\
 q_i^* &= \frac{\varepsilon + \Delta t}{\varepsilon} q_i^{n+1}; \\
 p_i^{(1)} &= p_i^* + \frac{\Delta t a^2}{2\Delta x} \left[ q_{i+1}^* - q_{i-1}^* + \frac{1}{a} (p_{i+1}^* - 2p_i^* + p_{i-1}^*) \right], \\
 q_i^{(1)} &= q_i^* + \frac{\Delta t}{2\Delta x} [p_{i+1}^* - p_{i-1}^* + a(q_{i+1}^* - 2q_i^* + q_{i-1}^*)]; \\
 p_i^n &= p_i^{(1)}, \\
 q_i^n &= q_i^{(1)}.
 \end{aligned} \tag{3.38}$$

## 2. Second-order scheme.

An explicit step for the transport equations and an implicit step for the stiff ODE part are applied. Therefore, the second-order TVD Runge–Kutta time discretisation of the adjoint equations (3.12) takes the form

$$\begin{aligned}
p_i^* &= p_i^{n+1} + \frac{\Delta t}{\varepsilon} f'(U_i^{n+1}) q_i^{n+1}, \\
q_i^* &= q_i^{n+1} - \frac{\Delta t}{\varepsilon} q_i^{n+1}; \\
p_i^{(1)} &= p_i^* - \Delta t a^2 D_x q_i^*, \\
q_i^{(1)} &= q_i^* - \Delta t D_x p_i^*; \\
p_i^{**} &= p_i^{(1)} - \frac{\Delta t}{\varepsilon} f'(U_i^{**}) q_i^{(1)} - \frac{2\Delta t}{\varepsilon} f'(U_i^*) q_i^*, \\
q_i^{**} &= q_i^{(1)} + \frac{\Delta t}{\varepsilon} q_i^{(1)} + \frac{2\Delta t}{\varepsilon} q_i^*; \\
p_i^{(2)} &= p_i^{**} - \Delta t a^2 D_x q_i^{**}, \quad p_i^n = \frac{1}{2}(p_i^* + p_i^{(2)}); \\
q_i^{(2)} &= q_i^{**} - \Delta t D_x p_i^{**}, \quad q_i^n = \frac{1}{2}(q_i^* + q_i^{(2)}).
\end{aligned} \tag{3.39}$$

Furthermore, we can use the approximations (3.23) and (3.35) to reformulate the second-order scheme (3.39) explicitly as

$$\begin{aligned}
p_i^* &= p_i^{n+1} + \frac{\Delta t}{\varepsilon} f'(U_i^{n+1}) q_i^{n+1}, \\
q_i^* &= \frac{\varepsilon - \Delta t}{\varepsilon} q_i^{n+1}; \\
p_i^{(1)} &= p_i^* + \frac{\Delta t a^2}{2\Delta x} \left[ q_{i+1}^* - q_{i-1}^* + \frac{1}{a} (p_{i+1}^* - 2p_i^* + p_{i-1}^*) + \frac{1}{2a} (\sigma_{i+1}^{+*} - \sigma_i^{+*}) + \frac{1}{2a} (\sigma_i^{-*} - \sigma_{i-1}^{-*}) \right], \\
q_i^{(1)} &= q_i^* + \frac{\Delta t}{2\Delta x} \left[ p_{i+1}^* - p_{i-1}^* - a(q_{i+1}^* - 2q_i^* + q_{i-1}^*) - \frac{1}{2} (\sigma_i^{-*} - \sigma_{i-1}^{-*}) + \frac{1}{2} (\sigma_{i+1}^{+*} - \sigma_i^{+*}) \right];
\end{aligned}$$

$$\begin{aligned}
p_i^{**} &= p_i^{(1)} - \frac{\Delta t}{\varepsilon} f'(U_i^{**}) q_i^{(1)} - \frac{2\Delta t}{\varepsilon} f'(U_i^*) q_i^*, \\
q_i^{**} &= \frac{\varepsilon + \Delta t}{\varepsilon} q_i^{(1)} + \frac{2\Delta t}{\varepsilon} q_i^*; \\
p_i^{(2)} &= p_i^{**} + \frac{\Delta t a^2}{2\Delta x} \left[ q_{i+1}^{**} - q_{i-1}^{**} + \frac{1}{a} (p_{i+1}^{**} - 2p_i^{**} + p_{i-1}^{**}) + \frac{1}{2a} (\sigma_{i+1}^{+**} - \sigma_i^{+**}) + \frac{1}{2a} (\sigma_i^{-**} - \sigma_{i-1}^{-**}) \right], \\
q_i^{(2)} &= q_i^{**} + \frac{\Delta t}{2\Delta x} \left[ p_{i+1}^{**} - p_{i-1}^{**} - a(q_{i+1}^{**} - 2q_i^{**} + q_{i-1}^{**}) - \frac{1}{2} (\sigma_i^{-**} - \sigma_{i-1}^{-**}) + \frac{1}{2} (\sigma_{i+1}^{+**} - \sigma_i^{+**}) \right]; \\
p_i^n &= \frac{1}{2} (p_i^* + p_i^{(2)}), \\
q_i^n &= \frac{1}{2} (q_i^* + q_i^{(2)}),
\end{aligned}$$

which satisfies the TVD property and is of second-order accurate. We can summarise the above computations to approximate adjoint system in the following proposition

**Proposition 3.5.1.** (*[67, Proposition 2]*). *Let  $\varepsilon > 0$ . The discrete adjoint equations (3.37) and (3.39) are first- and second-order discretisations in time and spatial of the continuous adjoint equations (3.12), respectively. The time discretisation is a (backwards) explicit Euler scheme in the transport and (backwards) implicit Euler step in the stiff ODE part. We obtain the first- and second-order Upwind schemes for the variables  $p \pm aq$ .*

Finally, to update the control, the discrete gradient of the reduced cost functional (3.16) with respect to variations in the initial data  $U_0$  on the interval  $I_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$  is given by

$$\nabla_{U_{0,i}} \tilde{J}(\cdot) = \Delta x \sum_{i=0}^{N+1} (p_{0,i} + f'(U_{0,i}) q_{0,i}). \quad (3.40)$$

## 3.6 Numerical results and discussion

In this section, the numerical results that were obtained from the first- and second-order relaxation approach previously discussed will be presented. All numerical results were performed by Matlab software, which was run on a 64-bit operating system with an Intel Core (TM) i5-7500 CPU

running at 3.40 GHz. Herein, we present the smooth and non-smooth examples related to the linear advection equation and the inviscid Burgers equation. For a given initial data  $U_0(x)$ , we can solve the flow equations forward in time for the state variable  $U(T, \cdot)$  using the relaxation schemes that have been presented in Chapter 2, Section 2.7. The optimal control problem (3.6) can then be rewritten as an unconstrained minimisation problem for the reduced cost functional (3.2). At each grid point, the gradient of the reduced cost functional (3.16) can be computed using the adjoint-based approach as discussed in the optimisation algorithm 3.4. It is important to take into account the forward solutions of the flow equations that were demonstrated in Section 2.8 of Chapter 2 by the use of first- and second-order relaxation schemes. The results for different grid sizes show that the number of iterations of the optimisation algorithm is independent of the grid size used in examples.

### 3.6.1 Optimal control of advection equation

In the following, we consider the optimal control problem (3.6) constrained with the linear advection equation

$$\frac{\partial U}{\partial t} + c \frac{\partial U}{\partial x} = 0, \quad (3.41)$$

where  $c$  is the constant propagation speed. First, we present optimal control results with smooth initial data for the desired (target) state

$$U_{d,0}(x) = \frac{1}{2} + x. \quad (3.42)$$

The initial control for the optimal solution is started with an initial guess  $U_0(x) = x$ , for  $x \in [0, 3]$  and the speed  $c = 1$ . Optimal control results for both first- and second-order relaxation schemes are presented in Figure 3.1, which shows that the method converges to the desired state and optimality is achieved.

Next, we assume an optimal control problem with Riemann data of the advection equation (3.41)

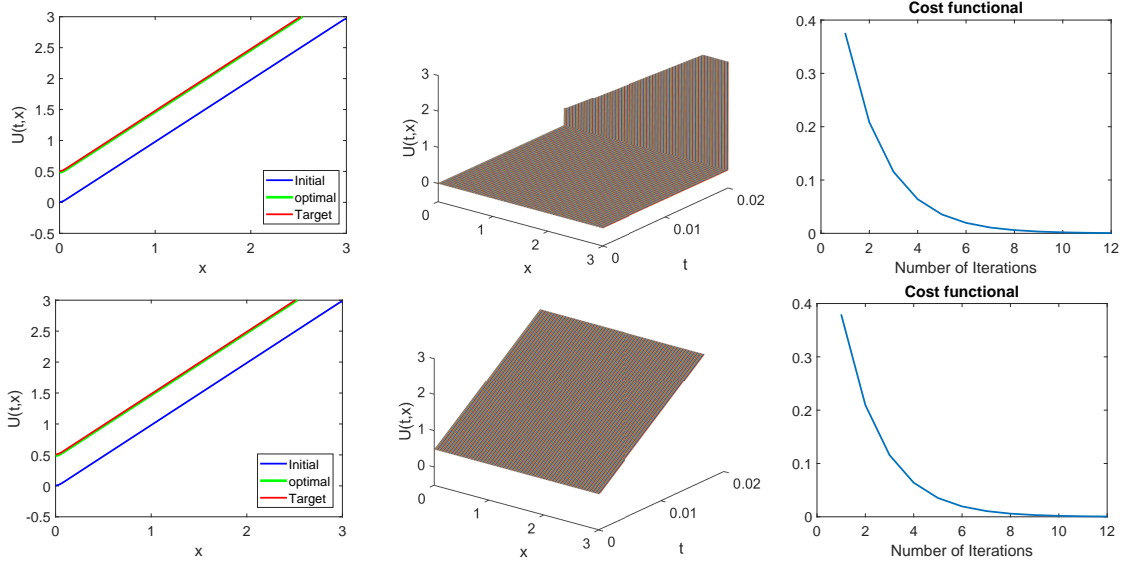


Figure 3.1: Initial, optimised and target values (Left), surface solution (Middle) and convergence history (Right) of the optimal control of the linear advection equation (3.41) with initial data (3.42) obtained at  $T = 0.02$  with the relaxation scheme: First-order (Top) and Second-order (Bottom).

and the initial data for the target state

$$U_{d,0}(x) = \begin{cases} 2.2 & \text{if } x < 0.5, \\ 1.2 & \text{if } x > 0.5, \end{cases} \quad (3.43)$$

and the initial data is taken from [92] that we start the optimisation process, which is

$$U_0(x) = \begin{cases} 2.0 & \text{if } x < 0.5, \\ 1.0 & \text{if } x > 0.5. \end{cases} \quad (3.44)$$

The numerical results of this example are displayed in Figure 3.2. The results obtained are a good match between optimal and the desired solutions. Furthermore, the computation time of the optimisation algorithm with tolerance  $tol = 10^{-4}$  for different grid points  $N = 100, 200, 300, 400, 500$  is reported. In Table 3.2, the number of iterations obtained with the first- and second-order scheme for the relaxation method and the computation times until convergence is introduced. It is shown that the number of optimisation iterations is independent of the grid points  $N$ .

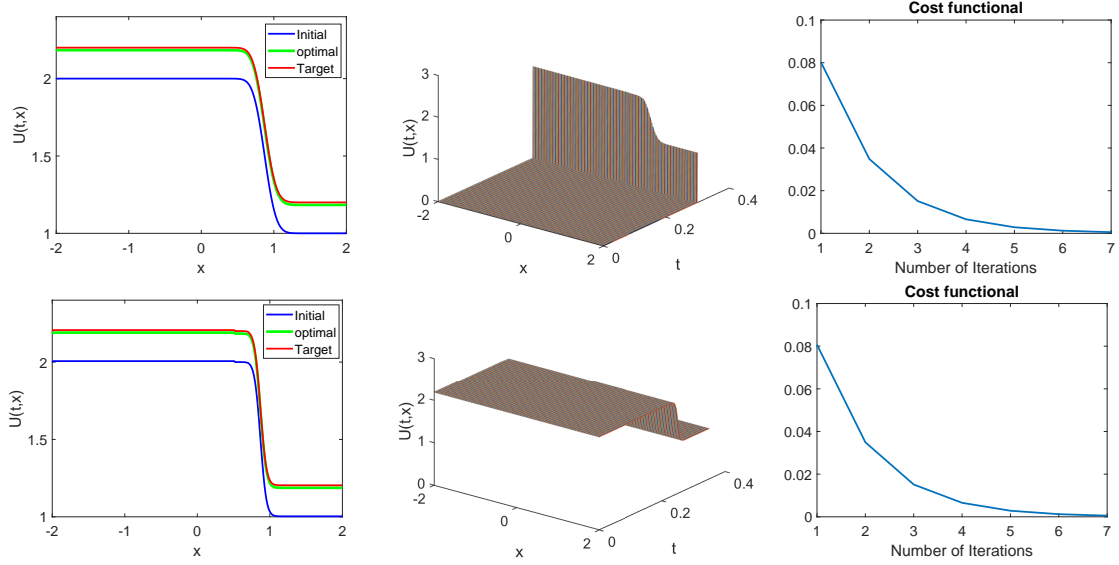


Figure 3.2: Initial, optimised and target values (Left), surface solution (Middle) and convergence history (Right) of the optimal control of the linear advection equation (3.41) with initial data (3.43) obtained at  $T = 0.4$  with the relaxation scheme: First-order (Top) and Second-order (Bottom).

### 3.6.2 Optimal control of Burger's equation

In this example, we consider the optimal control problem (3.6) governed by the inviscid Burger's equation

$$\frac{\partial U}{\partial t} + \frac{\partial}{\partial x} \left( \frac{U^2}{2} \right) = 0. \quad (3.45)$$

First, we begin with the optimal smooth initial data such that the flow solutions at final time  $T$  match the desired flow properties (at final time  $T$ ) given by the initial data

$$U_{d,0}(x) = \frac{1}{2} + \sin(x), \quad (3.46)$$

on  $x \in [0, 2\pi]$  and the design initial guess for optimal solution is  $U_0(x) = \sin(x)$ ; this example is taken from [67]. Figure 3.3 demonstrates the numerical results obtained from the adjoint approach combined with the first- and second-order relaxation methods, which are in excellent agreement and achieve optimality. Next, we consider the desired state, having a shock wave is the solution of

N	No. It.	First-order	Second-order
100	7	6.406250e-01	2.718750e+00
200	7	7.343750e-01	5.921875e+00
300	7	8.593750e-01	9.906250e+00
400	7	1.093750e+00	1.531250e+01
500	7	1.171875e+00	2.290625e+01

Table 3.1: Computational time (in second) and the number of iterations (No. It.) for the inverse design in the advection equation (3.41) and Riemann data (3.43) obtained with the relaxation approach.

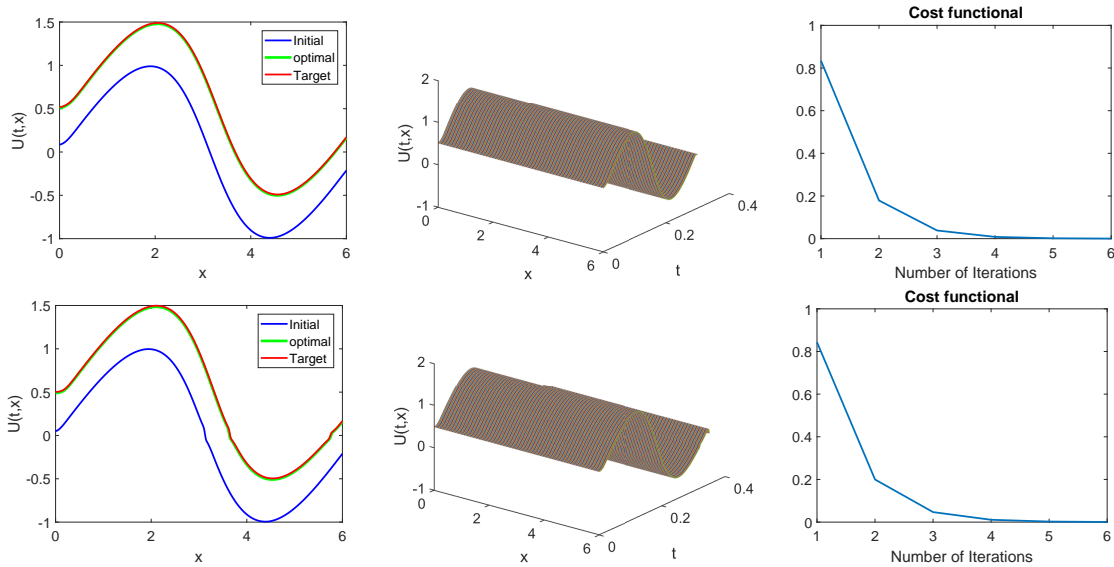


Figure 3.3: Initial, optimised and target values (Left), surface solution (Middle) and convergence history (Right) of the optimal control of the Cauchy problem (3.45) with initial data (3.46), obtained with the relaxation scheme: First-order (Top) and Second-order (Bottom).

the Riemann problem with the initial data

$$U_{d,0}(x) = \begin{cases} 2.2 & \text{if } x < 0.0, \\ 0.7 & \text{if } x > 0.0. \end{cases} \quad (3.47)$$



The initial control for the optimal solution is started with an initial guess

$$U_0(x) = \begin{cases} 2.0 & \text{if } x < 0.0, \\ 0.5 & \text{if } x > 0.0. \end{cases} \quad (3.48)$$

The numerical results are reported in Figure 3.4. The results reveal that the approach discussed above, using the relaxation method in both orders, can recover solutions with discontinuities and then optimal control is converged. The computation time for simulation and the number of iterations of optimal control results for the first- and second-order relaxation schemes is stated in Table 3.2, with final time  $t = 0.3$  and the grid points  $N = 100, 200, 300, 400, 500$ . The time taken for the optimisation algorithm to converge for both orders increase with the number of discretisation points,  $N$ . However, the time needed for the first-order to converge is smaller than that needed for the second-order. Finally, we present optimal control results for Burger's equation (3.45) with

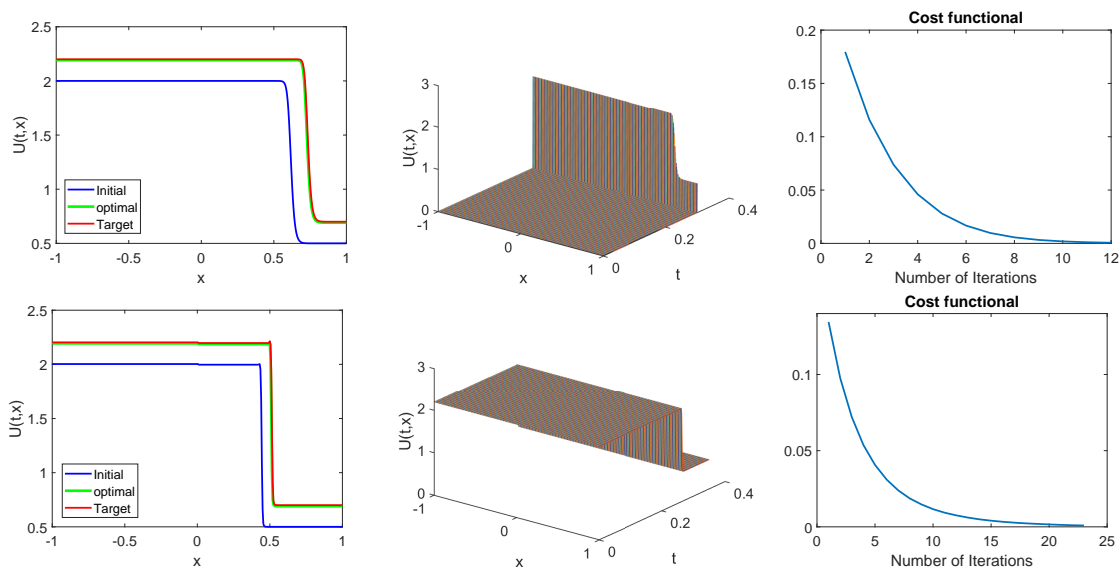


Figure 3.4: Initial, optimised and target values (Left), surface solution (Middle) and convergence history (Right) of the optimal control of the Riemann problem (3.45) with initial data (3.47) that comprises a shock wave, obtained with the relaxation scheme: First-order (Top) and Second-order (Bottom).

non-smooth data. The desired state consists of the rarefaction wave, which is the solution of the

N	First-order		Second-order	
	No. It.	CPU time	No. It.	CPU time
100	12	7.500000e-01	23	4.140625e+00
200	12	9.218750e-01	23	2.710938e+01
300	12	1.109375e+00	23	6.712500e+01
400	12	1.390625e+00	23	1.327969e+02
500	12	1.578125e+00	23	2.320156e+02

Table 3.2: Computational time (CPU time in second) and the number of iterations (No. It.) for the inverse design in the Burgers equation (3.45) and Riemann data (3.47) obtained with the relaxation approach.

Riemann problem with the initial data

$$U_{d,0}(x) = \begin{cases} 0.2 & \text{if } x < 0.0, \\ 1.2 & \text{if } x > 0.0, \end{cases} \quad (3.49)$$

is computed at final time  $T = 0.5$ . The initial guess for the iterative optimisation problem is chosen from [94] as

$$U_0(x) = \begin{cases} 0.0 & \text{if } x < 0.0, \\ 1.0 & \text{if } x > 0.0. \end{cases} \quad (3.50)$$

We obtained good results for this example, with the match between the optimal and target solutions. The optimised flow results at the final time  $T = 0.5$  are depicted in Figure 3.5 and it can be observed that the discontinuity is well recovered.

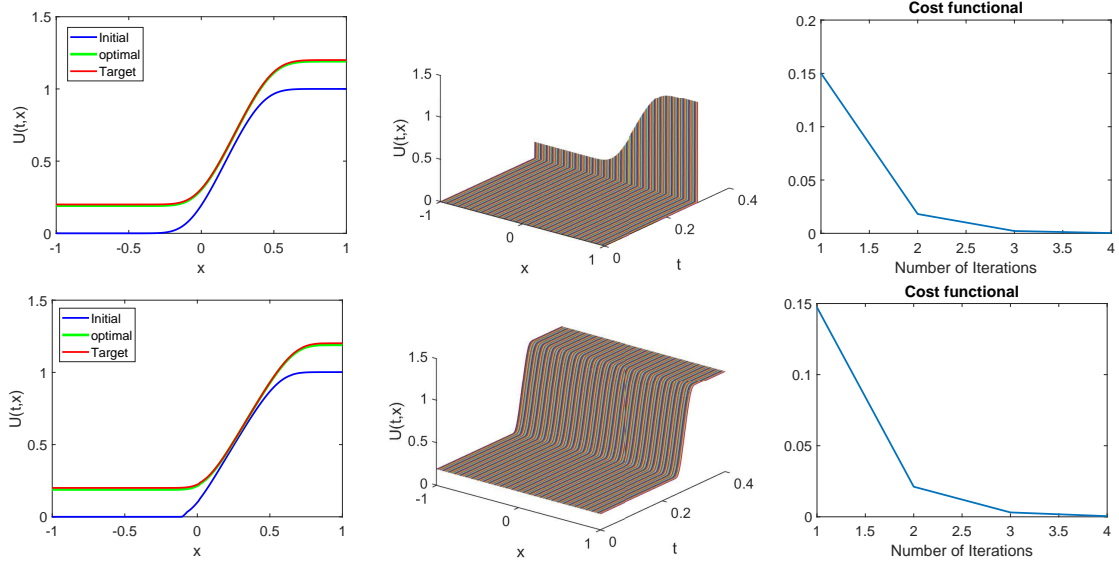


Figure 3.5: Initial, optimised and target values (Left), surface solution (Middle) and convergence history (Right) of the optimal control of the Riemann problem (3.45) with initial data (3.49) that comprises a rarefaction wave, obtained with the relaxation scheme: First-order (Top) and Second-order (Bottom).

Since discontinuities may occur in solutions of conservation laws, we might need an approximation to the generalised tangent vectors to improve the gradient descent method. Below we briefly discuss the tangent vectors approach that has been introduced by Herty and Piccoli [71].

### 3.7 Tangent vectors approach

We start by rewriting the hyperbolic relaxation approximation (3.4) to the conservation laws (3.3) as

$$\frac{\partial}{\partial t} \mathbf{Y} + \begin{pmatrix} 0 & 1 \\ a^2 & 0 \end{pmatrix} \frac{\partial}{\partial x} \mathbf{Y} = \begin{pmatrix} 0 \\ -\frac{1}{\varepsilon}(V - f(U)) \end{pmatrix}, \quad (3.51)$$

where  $\mathbf{Y} = (U, V)^T$ ,  $\mathbf{T}$  denotes the transpose, with the initial data given by

$$U(0, x) = U_0(x), \quad V(0, x) = f(U_0(x)). \quad (3.52)$$

Assume that all properties of the relaxation approximation (3.51), as stated in previous sections, are satisfied, we will discuss the optimal control problem (3.1) subject to (3.51) with initial data  $U_0$  acting as the control variable. The notations  $TV(\cdot)$  and  $\mathcal{U}$  denote the total variation and the set of admissible controls, respectively. For  $U_d \in L^1(\mathbb{R})$  and bounded interval  $\Omega \subseteq \mathbb{R}$ , we consider an unregularised cost functional of tracking type as

$$J(U(T, \cdot); U_d) = \frac{1}{2} \int \chi_\Omega(x) \|U(T, x) - U_d(x)\|^2 dx. \quad (3.53)$$

**Definition 3.7.1.** Given  $C > 0$ . We indicate by

$$\mathcal{U} := \{U : \Omega \longrightarrow \mathbb{R} \mid U \text{ is measurable and piecewise constant, } TV(U) \leq C\}$$

the set of admissible controls. For every  $U \in \mathcal{U}$ , we indicate by  $x_i = x_i(U), i = 1, 2, \dots, N$  with  $x_1 < x_2 < \dots < x_N$  the points of discontinuity of  $U$ .

**Definition 3.7.2.** We say that a function  $\gamma : \Omega \longrightarrow \mathbb{R}$  is a continuous path, if  $\gamma$  is continuous on the interval  $\Omega = [a, b], a < b$  with respect to  $L^1$ -norm, i.e., for all  $x \in \Omega$  and  $\alpha \neq 0$  such that

$$\lim_{\alpha \rightarrow 0} \|\gamma(x + \alpha) - \gamma(x)\|_{L^1} = 0. \quad (3.54)$$

**Definition 3.7.3.** Let  $U \in \mathcal{U}$  be a given function. A generalised tangent vector consists of two components: the  $L^1$  infinitesimal displacement  $w \in L^1(\mathbb{R})$  and the infinitesimal displacement of  $N$  discontinuities  $\xi \in \mathbb{R}$ . This vector has the form  $(w, \xi) \in L^1(\mathbb{R}) \times \mathbb{R}$ , with the pointwise limit

$$w(t, x) = \lim_{\delta \rightarrow 0} \frac{U_\delta(t, x) - U(t, x)}{\delta}, \quad \xi_i(t) = \lim_{\delta \rightarrow 0} \frac{x_i^\delta(t) - x_i(t)}{\delta}, \quad (3.55)$$

where  $U_\delta$  is the small variation on  $U$  and  $x_i^\delta$  is the shift of  $x_i$ .

The space of tangent vectors  $T_U = L^1(\mathbb{R}; \mathbb{R}) \times \mathbb{R}$  has a norm that depends on  $U$  through the number of points of discontinuity given by

$$\|(w, \xi)\| = \|w\|_{L^1} + \sum_{i=1}^N |\Delta_i U| |\xi_i|, \quad (3.56)$$

where  $\Delta_i U = U(t, x_i+) - U(t, x_i-)$  is the jumps in  $U$  at  $x_i$ .

Further, generalised tangent vectors may be used to describe variations of  $U$ . For  $\delta > 0$ , we have

$$U_\delta = U + \delta w + \sum_{i=1}^N \Delta_i U \chi_{[x_i - \delta \xi_i, x_i]} - \sum_{i=1}^N \Delta_i U \chi_{[x_i, x_i + \delta \xi_i]}, \quad (3.57)$$

means that we can obtain  $U_\delta$  starting with  $U$ , adding  $\delta w$  and then, shifting the  $i$ th discontinuity by  $\delta \xi_i$  where  $U$  has a jump. If  $\xi$  does not equal to zero, then, the function  $\delta \rightarrow U_\delta$  is non-differentiable in  $L^1$  as the first limit in (3.55) does not converge to any limit in  $L^1$  when  $\delta$  approaches zero. However, the limit (3.55) remains meaningful if we interpreted it as a weak limit in the space of measures with a singular point mass located at  $x_i$  and having magnitude  $|\Delta_i U| \xi_i$ . Therefore, the generalised tangent vector  $(w, \xi)$ , if it exists, is necessarily unique and can describe up to first-order variations [13].

Now, consider the system (3.51). Let  $U \in L^1(\mathbb{R}; \mathbb{R})$  be a piecewise Lipschitz continuous function with  $N$  jumps. We denote that  $\Sigma_U$  as the class of all continuous paths  $\gamma: [0, \delta_0] \rightarrow L^1_{loc}$  with  $\gamma(0) = U$ . With  $\delta_0 > 0$  might depend on  $\gamma$ , (cf. for example [13, 14] for more information). Therefore, we have the following definition:

**Definition 3.7.4.** ([71, Definition 2.2]) Let  $T_U = L^1(\mathbb{R}; \mathbb{R}) \times \mathbb{R}$  be a space of generalised tangent vectors to a piecewise Lipschitz function  $U$  with jumps at the points  $x_1 < x_2 < \dots < x_N$ . We say that a continuous path  $\gamma \in \Sigma_U$  generates a tangent vector  $(w, \xi) \in T_U$  if

$$\lim_{\delta \rightarrow 0} \frac{1}{\delta} \|\gamma(\delta) - \tilde{\gamma}(\delta)\|_{L^1} = 0,$$

with

$$\tilde{\gamma} = U + \delta w + \sum_{i=1}^N \Delta_i U \chi_{[x_i - \delta \xi_i, x_i]} - \sum_{i=1}^N \Delta_i U \chi_{[x_i, x_i + \delta \xi_i]}.$$

Assuming that  $U$  is a piecewise Lipschitz continuous function having simple discontinuities, therefore, we have the following result

**Theorem 3.7.1.** ([13, Lemma 2.1]) Let  $\gamma \in \Sigma_U$  be a regular variation for  $U$ . Moreover,  $\gamma(\delta) = U_\delta$  are piecewise Lipschitz continuous functions having simple discontinuities and the jumps  $x_i^\delta$  depend

continuously on  $\delta$ . We say that  $\gamma$  generates a tangent vector  $(w, \xi) \in T_U$  if and only if

$$\begin{aligned} \xi_i &= \lim_{\delta \rightarrow 0} \frac{x_i^\delta - x_i}{\delta}, \quad i = 1, 2, \dots, N, \\ \lim_{\delta \rightarrow 0} \int_a^b \left\| \frac{U_\delta(x_i^\delta + y) - U(x_i + y)}{\delta} - w(x_i + y) - \xi_i U_x(x_i + y) \right\| dy &= 0, \end{aligned} \quad (3.58)$$

whenever  $[x_i + a, x_i + b]$  contains only the point of discontinuity  $x_i$ .

Proof of the Theorem 3.7.1 can be found in [13]. In addition, the length of a regular path  $\gamma$  can be computed using the formula (3.56). Considering the basic assumptions on the system (3.51) outlined in [13, Assumptions (H1 - H3)] are true and the regular variations are locally preserved. Considering the definition of broad solution [71, DEFINITION A.2], we can derive a linearised system for the time evolution of the tangent vector  $(w, \xi)$ ; therefore, we have the following result

**Theorem 3.7.2.** ([71, Lemma 2.1]) *Let  $Y(t, x)$  be a piecewise Lipschitz continuous solution to (3.51) with initial data (3.52)  $Y(0, x) = \tilde{Y}$  in the class of piecewise Lipschitz functions with  $N$  simple discontinuities. Let  $(\tilde{w}, \tilde{\xi}) \in T_{\tilde{Y}}$  be a tangent vector to  $\tilde{Y}$  generated by the regular variation  $\gamma$  with  $\gamma(\delta) = \tilde{Y}_\delta$ . Let  $Y_\delta(t, x)$  be the solution to (3.51) with initial data  $Y_\delta(0, x) = \tilde{Y}_\delta(\cdot)$ . Then, there exists a time  $\bar{t} > 0$  such that for all  $t \in [0, \bar{t}]$ , the path  $\tilde{\gamma}$  with  $\tilde{\gamma}(\delta) = Y_\delta(t, x)$  is a regular variation for  $Y(t, x)$  generating the tangent vector  $(w, \xi) \in T_Y$ . Moreover, the vector  $(w, \xi)$  is the unique broad solution of the system*

$$\frac{\partial w}{\partial t} + \begin{pmatrix} 0 & 1 \\ a^2 & 0 \end{pmatrix} \frac{\partial w}{\partial x} = \begin{pmatrix} 0 \\ -\frac{1}{\varepsilon}(w_2 - f'(U)w_1) \end{pmatrix}, \quad (3.59)$$

with the initial data

$$\xi(0) = \tilde{\xi}, \quad w(0, x) = \tilde{w}(x), \quad (3.60)$$

where  $w = (w_1, w_2)^T$  and outside of the discontinuities of  $Y$ . For  $i = 1, 2, \dots, N$ , we have

$$\xi_i(t) = \tilde{\xi}_i, \quad \text{and} \quad L_j \cdot \left( \Delta_i w + \xi_i \Delta_i \frac{\partial Y}{\partial x} \right) = 0, \quad j \neq i, \quad (3.61)$$

along each line of discontinuity  $x_i = x_i(t)$  where  $Y$  has a discontinuity in the  $i$ th characteristic family where  $\Delta_i w = w(t, x_i+) - w(t, x_i-)$  and  $L_j$  is the  $j$ th left eigenvectors of the matrix  $\begin{pmatrix} 0 & 1 \\ a^2 & 0 \end{pmatrix}$ .

The proof of Theorem 3.7.2 follows from [13, Theorem 2.2], which has been also restated in [71, Appendix A]. By [13, Definition 2],  $\tilde{U} \in \mathcal{U}$  and  $\tilde{\mathbf{Y}} = (\tilde{U}, \tilde{V})^{\mathbf{T}}$  are piecewise Lipschitz with simple discontinuities. Additionally, because of the linear transport in the hyperbolic relaxation, systems (3.59) and (3.61) are particularly simple. Therefore, the equation for  $\xi_i$  can be solved effortlessly. However, in the previous result, it is assumed that all variations,  $\mathbf{Y}_\delta$  possess the same number of discontinuities which in the case of problem (3.1) is unknown a priori.

We can now use tangent vectors and their property (3.58) to compute the variations of the cost functional (3.53) as

**Theorem 3.7.3.** *Consider the cost functional given by (3.53), and let the linearised system of the tangent vector  $(w, \xi)$  be given by Theorem 3.7.2. A gradient of the cost functional  $J(\cdot)$  with respect to  $(w, \xi)$  for initial data  $U(0, x) = (U_0(x))$  is given as*

$$\begin{aligned} \nabla_{(w, \xi)} J(\cdot) &= \int \chi_\Omega (U(T, x) - U_d(x)) w_1(T, x) dx + \sum_{i=1}^N (U(T, x_i+) - U_d(T, x_i+)) \Delta_i U(T, x_i) \xi_i(T) \\ &\quad + \sum_{i=1}^N (U(T, x_i-) - U_d(T, x_i-)) \Delta_i U(T, x_i) \xi_i(T). \end{aligned} \quad (3.62)$$

The proof of Theorem 3.7.3 is like the proof of the proposition presented in [115, Proposition 3.1]. For a given initial data  $U_0$  and a stepsize  $\alpha > 0$ , an update initial data  $\bar{U}_0$  corresponding to a smaller value of the cost functional  $J(\cdot)$  is given by

$$\bar{U}_0(x) = U_0(x) - \left( \alpha w(0, x) + \sum_{i=1}^N \Delta_i U_0 \chi_{[x_i - \alpha \xi_i(0), x_i]} - \sum_{i=1}^N \Delta_i U_0 \chi_{[x_i, x_i + \alpha \xi_i(0)]} \right), \quad (3.63)$$

where  $w(0, x)$  is the solution at  $t = 0$  of (3.59) with transversality data

$$\begin{aligned} w_1(T, x) &= U(T, x) - U_d(x), \\ w_2(T, x) &= 0, \end{aligned} \quad (3.64)$$

and  $\xi_i(0)$  is the solution of (3.61) with transversality data

$$\xi_i(T) = [(U(T, x_i+) - U_d(T, x_i+)) - (U(T, x_i-) - U_d(T, x_i-))] \Delta_i U(T, x_i). \quad (3.65)$$

Moreover, the system (3.59) solved backwards in time with transversality data (3.64) and satisfies the system (3.61). Equations (3.51), (3.59), (3.62) and (3.64) represent the first-order optimality conditions for the problem (3.1). However, the system (3.51) is diagonalisable with eigenvalues  $\lambda_{1,2} = \pm a$  and characteristic variables

$$\eta_1 = V + aU, \quad \text{and} \quad \eta_2 = V - aU. \quad (3.66)$$

where  $\eta = (\eta_1, \eta_2)$ . Further, according to condition (3.61), we consider the minimisation problem for  $J(\cdot)$  in characteristic variables. Given (3.63), system (3.59) will be solved backwards for given transversality data  $w(T, x)$ . For  $\bar{w} = w(T - t, x)$  and we obtain the system

$$\begin{aligned} \frac{\partial \bar{w}}{\partial t} - \begin{pmatrix} 0 & 1 \\ a^2 & 0 \end{pmatrix} \frac{\partial \bar{w}}{\partial x} &= \begin{pmatrix} 0 \\ \frac{1}{\varepsilon} (\bar{w}_2 - f'(U(T - t, x)) \bar{w}_1) \end{pmatrix}, \\ \bar{w}(0, x) &= w(T, x), \end{aligned} \quad (3.67)$$

with eigenvalues  $\lambda_{1,2} = \mp a$  and characteristic variables are given by

$$\phi_1 = \bar{w}_2 + a\bar{w}_1, \quad \text{and} \quad \phi_2 = \bar{w}_2 - a\bar{w}_1. \quad (3.68)$$

In the subsequence section, we briefly present numerical schemes that have been proposed by Herty and Piccoli in [71].

### 3.7.1 Numerical scheme

Let the cost functional, equation (3.53) and initial data  $\mathbf{Y}_0 = (U_0, V_0)^T$  be given. For  $\varepsilon$  sufficiently small and  $\Omega = [0, 1]$ , we consider the following optimisation problem

$$\min_{\mathbf{Y}_0} J \quad \text{s.t.} \quad \text{system (3.51),} \quad \mathbf{Y}(0, x) = \mathbf{Y}_0(x), \quad \mathbf{Y}(t, 0) = \mathbf{Y}(t, 1), \quad x \in \Omega, \quad t \in [0, T]. \quad (3.69)$$

Here,  $a^2$  is fixed and satisfy the sub-characteristic condition (3.5). We introduce an equidistant spatial grid on  $\Omega$  with  $\Delta x = x_{i+1} - x_i, i = 0, 1, \dots, N$  and the time level  $t^n = n\Delta t, \quad n = 0, 1, \dots, M$ , where a time step  $\Delta t$  is choosing to be satisfying the CFL condition, that is,  $\Delta t = |a|\Delta x$ . Assume  $x_{i+\frac{1}{2}} = x_i + \frac{\Delta x}{2}, x_N = 1, t^M = T$ , and  $\mathcal{T}^{-1} \in \mathbb{R}^{2 \times 2}$  is the transformation to characteristic



variables (3.66), i.e.

$$\mathcal{T}^{-1} \begin{pmatrix} 0 & 1 \\ a^2 & 0 \end{pmatrix} \mathcal{T} = \begin{pmatrix} a & 0 \\ 0 & -a \end{pmatrix} \quad \text{and} \quad \eta = \mathcal{T}^{-1} \mathbf{Y}. \quad (3.70)$$

We may use an operator splitting [9] to split the transport part and stiff-source term. The splitting of the characteristic variables (3.66) for  $t \in [t^n, t^{n+1}]$  gives

$$\begin{aligned} \frac{\partial}{\partial t} \eta_1 + a \frac{\partial}{\partial x} \eta_1 &= 0, & \frac{\partial}{\partial t} \eta_2 &= 0, \\ \frac{\partial}{\partial t} \eta_1 &= -\frac{1}{\varepsilon} ((\mathcal{T}\eta)_2 - f((\mathcal{T}\eta)_1)), \\ \frac{\partial}{\partial t} \eta_2 &= -\frac{1}{\varepsilon} ((\mathcal{T}\eta)_2 - f((\mathcal{T}\eta)_1)), \\ \frac{\partial}{\partial t} \eta_2 - a \frac{\partial}{\partial x} \eta_2 &= 0, & \frac{\partial}{\partial t} \eta_1 &= 0. \end{aligned} \quad (3.71)$$

A discontinuity at time  $t^n$  in any component of  $\eta_0$  moves with speed  $a$  and  $-a$ , respectively. Further, consider  $\phi(t, x) = \phi(T - t, x)$  and  $J(\eta; U_d) = J(\mathcal{T}^{-1} \mathbf{Y}; U_d)$ , we can get the gradient of the cost functional  $J(\cdot)$  in terms of characteristic variables and the associated tangent vector  $(\phi, \xi)$  as

$$\begin{aligned} \nabla_{\eta_0} J &= \frac{1}{2} \int \chi_{\Omega}(x) \left( \frac{\eta_1(T, x) - \eta_2(T, x)}{2a} - U_d(x) \right) \phi_1(T, x) dx \\ &\quad - \frac{1}{2} \int \chi_{\Omega}(x) \left( \frac{\eta_1(T, x) - \eta_2(T, x)}{2a} - U_d(x) \right) \phi_2(T, x) dx \\ &\quad + \frac{1}{2} \sum_{j=1}^N \left( \frac{\eta_1(T, x_{j+}) - \eta_2(T, x_{j+})}{2a} - U_d(x_{j+}) \right) \xi_j(T) (\Delta_j \eta_1(T, x_j) - \Delta_j \eta_2(T, x_j)) \\ &\quad + \frac{1}{2} \sum_{j=1}^N \left( \frac{\eta_1(T, x_{j-}) - \eta_2(T, x_{j-})}{2a} - U_d(x_{j-}) \right) \xi_j(T) (\Delta_j \eta_1(T, x_j) - \Delta_j \eta_2(T, x_j)), \end{aligned} \quad (3.72)$$

where  $\Delta_j v(T, x_j) = v(T, x_{j+}) - v(T, x_{j-})$  and the evaluation of  $\eta$  at time  $T$  is at point  $x_j(T)$  in the last two terms. The value of  $x_j$  is computed by moving the  $j$ th discontinuity with speed  $\pm a$  depending on whether it is a jump in the first or second component. Assume that the cell average on  $[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$  at time  $t^n$  for any function  $U(t, x)$  denoted by

$$U_i^n = \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} U(t^n, x) dx.$$

Thus, for  $i = 0, 1, \dots, N$  and  $n = 1, 2, \dots, M$ , a first-order Upwind scheme for the solution of system (3.51) using the characteristic variables (3.66) is given by

$$\begin{aligned}
\mathbf{Y}_i^0 &= (\mathbf{Y}_0)_i, \\
\eta_{1,i} &= (\mathcal{T}^{-1}\mathbf{Y}_{i-1}^{n-1})_1, \quad \eta_{1,0} = (\mathcal{T}^{-1}\mathbf{Y}_N^{n-1})_1, \quad \eta_{2,i} = (\mathcal{T}^{-1}\mathbf{Y}_i^{n-1})_2, \\
\tilde{Y}_{1,i} &= (\mathcal{T}\eta_i)_1, \quad \tilde{Y}_{2,i} = \exp\left(-\frac{\Delta t}{\varepsilon}\right)(\mathcal{T}\eta_i)_2 + (1 - \exp\left(-\frac{\Delta t}{\varepsilon}\right))f(U_i), \\
\eta_{2,i} &= (\mathcal{T}^{-1}\tilde{Y}_{i+1})_2, \quad \eta_{2,N} = (\mathcal{T}^{-1}\tilde{Y}_0)_2, \quad \eta_{1,i} = (\mathcal{T}^{-1}\tilde{Y}_i)_1, \\
\mathbf{Y}_i^{n+1} &= \mathcal{T}\eta_i,
\end{aligned} \tag{3.73}$$

where  $\mathbf{Y} = (\tilde{Y}_1, \tilde{Y}_2)^\mathbf{T}$ . The scheme (3.73) uses equation (3.66), which leads to a different scheme compared to [9]. Therefore, the transformation to characteristic variables  $\eta$  and the CFL condition allows resolving transport parts exactly.

Similarly, we can discretise equation (3.67) instead of equation (3.59) for the variations  $\bar{w}$  and transformed into characteristic variables (3.68). Hence, for given a discretised initial data  $\bar{w}_i^0 = (\bar{w}_0)_i$ , we have

$$\begin{aligned}
\phi_{2,i} &= (\mathcal{T}^{-1}\bar{w}_{i-1}^{n-1})_2, \quad \phi_{2,0} = (\mathcal{T}^{-1}\bar{w}_N^{n-1})_2, \quad \phi_{1,i} = (\mathcal{T}^{-1}\bar{w}_i^{n-1})_1, \\
\tilde{w}_{1,i} &= (\mathcal{T}\phi_i)_1, \quad \tilde{w}_{2,i} = \exp\left(\frac{\Delta t}{\varepsilon}\right)(\mathcal{T}\phi_i)_2 + (1 - \exp\left(\frac{\Delta t}{\varepsilon}\right))f'(U_i^{M-n})\tilde{w}_{1,i}, \\
\phi_{1,i} &= (\mathcal{T}^{-1}\tilde{w}_{i+1})_1, \quad \phi_{1,N} = (\mathcal{T}^{-1}\tilde{w}_0)_1, \quad \phi_{2,i} = (\mathcal{T}^{-1}\tilde{w}_i)_2, \\
\bar{w}_i^n &= \mathcal{T}\phi_i.
\end{aligned} \tag{3.74}$$

To discretise system (3.61), we can use a piecewise constant approximation

$$\mathbf{Y}(t, x) = \sum_{i=0}^N \chi_{[t^n, t^{n+1}] \times [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]}(t, x) \mathbf{Y}_i^n, \tag{3.75}$$

to reconstruct the solution and similarly for the initial data. The numerical solution to the problem (3.69) leads to considering piecewise constant controls  $\mathbf{Y}_0 \in \mathcal{U}$  having discontinuities at each cell boundary  $x_{i+\frac{1}{2}}$ . Further, a shift in the position of the discontinuity  $\xi_i$  may occur at each boundary  $x_{i+\frac{1}{2}}$ . However, the number of discontinuities is fixed as long as the spatial resolution is unchanged, which is a crucial assumption in Theorem 3.7.2. Still, we can solve (3.69) with respect to  $\eta_0 = \mathcal{T}^{-1}\mathbf{Y}_0$  instead of  $\mathbf{Y}_0$  since  $\mathbf{Y}$  and characteristic variables  $\eta$  are equivalent through the

linear transformation  $\mathcal{T}$ . For fixed grid points  $N$ , the set of all admissible controls consists of all piecewise constant functions  $\eta_0(x)$  given by

$$\eta_{j,0} = \sum_{i=0}^N \chi_{[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]}(x) \eta_{j,0,i}, \quad j = 1, 2. \quad (3.76)$$

Moreover,  $\eta_{j,0} \in \mathcal{U}$  for  $j = 1, 2$  having each only  $\frac{N}{2}$  points of discontinuity. When the grid points  $N$  sufficiently large, the first component  $\eta_{1,0}$  may have a discontinuity only at  $x_{i+\frac{1}{2}}$  for some odd values  $i$  and the second component  $\eta_{2,0}$  may only have a discontinuity at  $x_{i+\frac{1}{2}}$  for some even values  $i$ . Thus, to approximate any piecewise constant function, we can choose  $\eta_{j,0}$  as

$$(\eta_{1,0})_{2i} = (\eta_{1,0})_{2i+1}, \quad (\eta_{2,0})_{2i-1} = (\eta_{2,0})_{2i}, \quad i = 0, \dots, \frac{N}{2}. \quad (3.77)$$

For given cell  $i$ , structures (3.76) and (3.77) allow either the first or second component to have a discontinuity across the cell boundary  $i + \frac{1}{2}$ . The other component is constant across the cell boundary, which guarantees that the second part of equation (3.61) is satisfied. Furthermore, the jump is parallel to the eigenvectors and does not split under advection. This construction is preserved in the splitting scheme then, (3.61) is automatically satisfied after the transport and application of the source term. Similarly, for  $\varphi$  provided that  $\varphi_{j,0} = \mathcal{T}^{-1} \bar{w}_{j,0}$ ,  $j = 1, 2$  satisfies (3.77). We have the  $L^1$ -variations  $\varphi_0$  and the variation in the position of discontinuities  $\xi_i$  when computing the tangent vector to  $\eta_0$ . Denote by  $\xi_i$ ,  $i = 0, 1, \dots, N$ , the variation of the discontinuity at position  $x_{i+\frac{1}{2}}$ . Hence,  $\xi_i$  for  $i$  odd (even) is the variation of the discontinuity in the first (second) component of  $\eta_0$ . Under assumption (3.77), the position of discontinuities in the first and second components of  $\eta$  at time  $t^n$  are given by

$$x_{2i-1}(t^n) = x_{2i-1}(0) + at^n, \quad x_{2i}(t^n) = x_{2i}(0) - at^n, \quad i = 0, 1, \dots, N \quad (3.78)$$

where  $x_k(0) = x_{k+\frac{1}{2}}$  and we consider the discontinuities exiting at  $x = 1(x = 0)$  enter again at  $x = 0(x = 1)$ . Here, we can show that the second part of equation (3.61) is always satisfied, where  $L_k$  is the  $k$ th unit vector. Let  $i$  be odd and  $\eta$  be computed by the previous scheme. Then,  $\Delta_i \frac{\partial}{\partial x} \eta_2 = 0$  since  $\eta_2$  is constant across the position of the discontinuity in the first family  $x_{2i-1}(t^n)$ . In addition,  $\varphi_{0,i} = \mathcal{T}^{-1} \bar{w}_{0,i}$ ,  $i = 0, 1, \dots, N$  in (3.74) satisfies (3.77) where  $\Delta_i \varphi_2 = 0$ . Similarly, the second part of equation (3.61) is satisfied trivially for  $i$  is even.

To determine a descent direction for  $J(\cdot)$ , we use the following discretisation to discretise the gradient (3.72) for  $\varphi(t, x) = \phi(T - t, x)$

$$\begin{aligned}
\varphi_{1,2i}^0 &= \frac{\eta_{1,2i}^M - \eta_{2,2i}^M}{2a} - (Ud)_{2i}, & \varphi_{1,2i+1}^0 &= \varphi_{1,2i}^0, \\
\varphi_{2,2i-1}^0 &= - \left( \frac{\eta_{1,2i-1}^M - \eta_{2,2i-1}^M}{2a} - (Ud)_{2i-1} \right), & \varphi_{2,2i}^0 &= \varphi_{2,2i-1}^0, \\
\xi_{2i-1} &= \frac{1}{2a} (\hat{\Delta}_{2i-1}(\mathcal{T}\varphi^M)_1 - \hat{\Delta}_{2i-1}Ud) \Delta_{2i-1}\varphi_1^M, \\
\xi_{2i} &= -\frac{1}{2a} (\hat{\Delta}_{2i}(\mathcal{T}\varphi^M)_1 - \hat{\Delta}_{2i}Ud) \Delta_{2i}\varphi_2^M,
\end{aligned} \tag{3.79}$$

where  $\Delta_j v(x) = v(x_{j+1}) - v(x_j)$  and  $\hat{\Delta}_j v(x) = \frac{1}{2}(v(x_{j+1}) + v(x_j))$ .

Finally, to update the initial control  $\eta^0$ , we use (3.63) for  $\mathcal{T}^{-1}\eta^0 \in \mathcal{U}$  that satisfies (3.77) since the tangent vector  $(\varphi(T, x), \xi(T))$  to  $\eta^0$  describes the  $L_1$ -variation and variation of the position of discontinuities. Thus, for  $j \in \{0, \dots, N\}$  is odd and  $k \in \{0, \dots, N\}$  is even, the new control  $\tilde{\eta}_i^0$  can be computed by

$$\begin{aligned}
\Psi_{1,j-1} &= \Psi_{1,j} = \min \left( (-\tilde{\xi}_j)^+, \Delta x \right) \eta_{1,j+1}^0 + \max \left( \tilde{\xi}_{j-2}^+ - \Delta x, 0 \right) \eta_{1,j-1}^0 \\
&\quad + \left[ \Delta x - \min \left( (-\tilde{\xi}_j)^+, \Delta x \right) - \max \left( \tilde{\xi}_{j-2}^+ - \Delta x, 0 \right) \right] \eta_{1,j}^0, \\
\Psi_{2,k-1} &= \Psi_{2,k} = \min \left( (-\tilde{\xi}_k)^+, \Delta x \right) \eta_{2,k+1}^0 + \max \left( \tilde{\xi}_{k-2}^+ - \Delta x, 0 \right) \eta_{2,k-1}^0 \\
&\quad + \left[ \Delta x - \min \left( (-\tilde{\xi}_k)^+, \Delta x \right) - \max \left( \tilde{\xi}_{k-2}^+ - \Delta x, 0 \right) \right] \eta_{2,k}^0, \\
\tilde{\eta}_{1,i}^0 &= \frac{\Psi_{1,i}}{\Delta x} - \varphi_{1,i}^M, \\
\tilde{\eta}_{2,i}^0 &= \frac{\Psi_{2,i}}{\Delta x} - \varphi_{2,i}^M,
\end{aligned} \tag{3.80}$$

where the volume-averaged shifted control in cell  $j$  is denoted by  $\Psi_{k,j}\Delta x$  for the  $k$ th-component,  $x^+ = \max\{0, x\}$  and  $\tilde{\xi}_i = \mathcal{P}(-\xi_i)$  where  $\mathcal{P}$  is the projection on  $[-2\Delta x, 2\Delta x]$  given by

$$\mathcal{P}(z) = \begin{cases} -2\Delta x & \text{if } z \leq -2\Delta x, \\ z & \text{if } -2\Delta x \leq z \leq 2\Delta x, \\ 2\Delta x & \text{if } z \geq 2\Delta x. \end{cases} \tag{3.81}$$

Equation (3.80) corresponds to a gradient step in the  $L^1$ -variation but with a scaled gradient step in the variation of the shock position to prevent shock variations to interact. With the first component, the discontinuity at  $x_{i+\frac{1}{2}}$  is moved by  $\xi_i$  and at  $x_{i-\frac{3}{2}}$  by  $\xi_{i-2}$  where  $\varphi$  satisfies (3.77) and also  $\tilde{\eta}$ .

### 3.7.2 Optimisation algorithm

Here we present a numerical algorithm to solve the problem (3.69). For given a time  $T > 0$ , the sub-characteristic condition  $a \geq \max_U |f'(U)|$  is satisfied. Let  $i = 0, 1, \dots, N$  and  $n = 0, 1, \dots, M$  be spatial and time grid points, respectively and  $(U_d)_i$  be a discrete desired function of  $U_d$ . Hence, we have the following steps:

1. Choose  $\Delta t = \Delta x/a$  and  $k = 0$ . Assume  $\eta_{0,i}^k = (\eta_{1,0,i}^k, \eta_{2,0,i}^k)$  is an arbitrary initial control, where  $\eta_0^k$  satisfies (3.77).
2. Solve system (3.73) with initial data  $(\mathbf{Y}_0)_i = \mathcal{T} \eta_{0,i}^k$  forward in time, to obtain  $\mathbf{Y}_i^M = \mathcal{T} \eta_i^M$ .
3. Set an initial data  $\varphi_i^0$  and shock variations  $\xi_i$  as in the system (3.79), with  $x_i(t^M)$  given by equation (3.78).
4. Solve system (3.74) with initial data  $\bar{w}_i^0 = \mathcal{T} \varphi_i^0$ , backwards in time to obtain  $\bar{w}_i^M = \mathcal{T} \varphi_i^M$ .
5. Update the control  $\eta_0$  using the formula (3.80) with values at old iteration  $\eta_{0,i}^k = \eta_i^0$  to obtain  $\eta_{0,i}^{k+1} = \tilde{\eta}_i^0$  in the new iteration  $k + 1$ .
6. If  $J(\mathcal{T} \eta; U_d)$  is sufficiently small, convergence is achieved. Otherwise, replace  $k$  by  $k + 1$  and repeat Steps 2 to 5.

### **3.8 Concluding remarks**

We derived the adjoint equations and optimality conditions for the inverse design optimal control problem constrained by hyperbolic systems of conservation laws. Due to the issues related to non-linear systems, we used a continuous optimisation approach based on relaxation approximations. Also, we presented the general theory and the numerical algorithm of the tangent vectors approach for solving the optimal control of systems of conservation laws. It can be noted that the numerical schemes and optimisation algorithm presented in this chapter are easier to extend to the multi-dimensional systems of conservation laws. Finally, we illustrated theoretical findings by presenting numerical results related to two optimal control problems constrained by linear advection and Burgers equations in smooth and non-smooth initial data. The computational time and convergence of minimisation problems were also reported. In the next chapter, we consider optimal control problems constrained by multi-dimensional systems of conservation laws.

# Chapter 4

## Optimal control of multi-dimensional systems of conservation laws

This chapter deals with optimal control problems governed by multi-dimensional systems of conservation laws. It extends the results of the previous chapter that were concerned with the one-dimensional case. We start the chapter by recalling results on the existence and uniqueness of solutions to the multi-dimensional systems of conservation laws. Then, we derive the optimality conditions for the multi-dimensional case using a relaxation approximation for the equations. We discretise the proposed systems for both time and space to obtain schemes for solving the forward and backwards equations. We then present the numerical algorithm that solves the optimal control problems. We illustrate our results on examples related to the two-dimensional inviscid Burger's equation. Multi-dimensional conservation laws are challenging [26, 86, 116] and their treatment requires special methods.

### 4.1 Problem formulation

We consider the optimal control problem with matching type objective functional formulated as

$$\min_{U_0} J(U(\cdot, T), U_0; U_d) = \frac{1}{2} \int_{\mathbb{R}^d} \|U(T, \mathbf{x}; U_0) - U_d(\mathbf{x})\|^2 d\mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^d, \quad (4.1)$$

where  $U_d$  is a fixed desired state and  $U(t, \mathbf{x})$  is the unique entropy solution of the multi-dimensional (MD) system of conservation laws

$$\frac{\partial U}{\partial t} + \nabla \cdot f(U) = 0, \quad (4.2)$$

with the initial data

$$U(0, \mathbf{x}) = U_0(\mathbf{x}), \quad \text{at } t = 0, \quad (4.3)$$

where  $(t, \mathbf{x}) \in \mathbb{R}_+^{d+1} = \mathbb{R}_+ \times \mathbb{R}^d = [0, \infty) \times \mathbb{R}^d$ ,  $\nabla = \left( \frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_d} \right)$  with  $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ , and  $f(U)$  is the flux function, which is a non-linear with the components  $f_i : \mathbb{R}^m \rightarrow \mathbb{R}^m$  for  $i = 1, \dots, d$ .

We can solve in principle (4.2) and (4.3) to obtain a solution  $U(t, \mathbf{x}, U_0)$ , which depends on the initial condition  $U_0$ . Upon substituting the solution in (4.1) and discretising, we obtain an optimisation problem for the control  $U_0$ . Hence, existence and uniqueness of solutions of (4.1) - (4.3) depends on the solutions of the underlying problems.

## 4.2 Multi-dimensional conservation laws

Here we review some results on the existence and uniqueness of weak entropy solutions of the system of conservation laws in the form (4.2). These results are mainly due to Bressan et al. [117] and Zheng [118]. For the one-dimensional case, the existence and uniqueness of weak entropy solutions have been introduced in Chapter 2. Let  $\mathcal{D}$  be a state domain which is a subset of  $\mathbb{R}^m$  and let  $\mathcal{S}$  be a hypersurface in  $\mathbb{R}^{d-1}$ . The hyperbolicity of (4.2) requires that for any  $\mathbf{n} \in \mathcal{S}$ ,

$$\nabla_U f(U) \cdot \mathbf{n} \quad \text{have } m \text{ real eigenvalues } \lambda_i(U; \mathbf{n}), \quad 1 \leq i \leq m. \quad (4.4)$$

We say that system (4.2) is hyperbolic in a state domain  $\mathcal{D}$  if condition (4.4) holds for any  $U \in \mathcal{D}$  and  $\mathbf{n} \in \mathcal{S}$ . A function  $\eta : \mathcal{D} \rightarrow \mathbb{R}$  is called an entropy of the system (4.2) if there exists a vector function  $\mathbf{q} : \mathcal{D} \rightarrow \mathbb{R}^d$ ,  $\mathbf{q} = (q_1, \dots, q_d)$  satisfying

$$\nabla q_i(U) = \nabla \eta \nabla f_i(U), \quad i = 1, 2, \dots, d. \quad (4.5)$$



Then,  $\mathbf{q}$  is called the corresponding entropy flux and  $(\eta, \mathbf{q})$  is simply called an entropy pair. An entropy  $\eta(U)$  is called a convex entropy in  $\mathcal{D}$  if  $\nabla^2 \eta(U)$  is positive definite for any  $U \in \mathcal{D}$  and a strictly convex entropy in  $\mathcal{D}$  if  $\nabla^2 \eta(U) - c_0 I$  is positive definite with  $c_0 > 0$  a constant and  $I$  is the  $m \times m$  identity matrix. A solution of the conservation laws (4.2) is said to be entropic if there exist an entropy pair  $(\eta, \mathbf{q})$  such that the so-called Lax entropy inequality is satisfied

$$\frac{\partial \eta(U)}{\partial t} + \nabla \cdot \mathbf{q}(U) \leq 0 \quad (4.6)$$

We have the following result related to the existence of classical solution of Cauchy problems that are due to Dafermos [107, 119].

**Theorem 4.2.1.** *Consider the Cauchy problem for a general hyperbolic system (4.2) with a strictly convex entropy and initial data in the form (4.3). Assume that  $U_0 : \mathbb{R}^d \rightarrow \mathcal{D}$  is in  $H^s \cap L^\infty$  with  $s > d/2 + 1$ . Then there exists a finite time  $T = T(\|U_0\|_s, \|U_0\|_{L^\infty}) \in (0, \infty)$  such that there exists a unique bounded classical solution  $U \in C^1([0, T] \times \mathbb{R}^d)$  with  $U(t, \mathbf{x}) \in \mathcal{D}$  for  $(t, \mathbf{x}) \in [0, T] \times \mathbb{R}^d$  and  $U \in C([0, T]; H^s) \cap C^1([0, T]; H^{s-1})$ .*

*Proof.* The proof can be found on [107, 118–120] and the references therein. □

Furthermore, the existence of the classical solution can also presented based on a sharp continuation principle for  $U_0 \in H^s$  with  $s > d/2 + 1$ ; the time interval  $[0, T)$  is the maximal interval of the classical  $H^s$  existence for the system (4.2), with  $T < \infty$ , if and only if either  $U(t, \mathbf{x})$  escapes every compact subset  $\mathcal{K} \Subset \mathcal{D}$  as  $t$  tends to  $T$  or

$$\left\| \left( \frac{\partial U}{\partial t}, \nabla U \right) (t, \mathbf{x}) \right\|_{L^\infty} \rightarrow \infty \quad \text{as } t \rightarrow T.$$

We have the following stability result for classical solutions. It applies to the set of entropy solutions, Lipschitz solutions and solutions containing rarefaction waves and vacuum states.

**Theorem 4.2.2.** *Suppose that  $U, W$  are two entropy solutions of (4.2) on  $[0, T)$ , taking values in a convex compact subset  $\mathcal{K}$  of  $\mathcal{D}$ , with initial data  $U_0$  and  $W_0$ , respectively. Assume that  $W$  is Lipschitz with Lipschitz constant  $L$ , then*

$$\int_{|x| < R} |U(t, \mathbf{x}) - W(t, \mathbf{x})|^2 dx \leq C(T) \int_{|x| < R+Lt} |U_0(\mathbf{x}) - W_0(\mathbf{x})|^2 dx \quad (4.7)$$

holds for any  $R > 0$  and  $t \in [0, T)$ , and with  $L > 0$  depending solely on  $\mathcal{K}$ .

*Proof.* For the proof, we can refer to [107, 117, 119, 120] and the references therein. □

Shock-front solutions, the simplest type of discontinuous solutions, are the most important discontinuous non-linear progressing wave solutions of conservation laws (4.2). Shock-front solutions are discontinuous piecewise smooth entropy solutions with the following structure:

- (i) There exist a  $C^2$  space-time hypersurface  $\mathcal{S}(t)$  defined in  $(t, \mathbf{x})$  for  $0 \leq t \leq T$  with space-time normal  $(\mathbf{n}_t, \mathbf{n}_x) = (\mathbf{n}_t, \mathbf{n}_1, \dots, \mathbf{n}_d)$  and two  $C^1$  vector valued functions  $U^\pm(t, \mathbf{x})$ , defined on respective domains  $\mathcal{D}^\pm$  on either side of the hypersurface  $\mathcal{S}(t)$ , and satisfying

$$\frac{\partial U^\pm}{\partial t} + \nabla \cdot f(U^\pm) = 0, \quad \text{in } \mathcal{D}^\pm, \quad (4.8)$$

- (ii) The jump across  $\mathcal{S}(t)$  satisfies the Rankine-Hugoniot condition:

$$\mathbf{n}_t(U^+ - U^-) + \mathbf{n}_x \cdot (f(U^+) - f(U^-)) = 0 \quad \text{on } \mathcal{S} \quad (4.9)$$

Since (4.2) is nonlinear, the surface  $\mathcal{S}$  is unknown in advance and must be determined as a part of the solution of the problem; thus, the equations in (4.8) - (4.9) describe a multi-dimensional free-boundary value problem for (4.2). The notions of genuine-nonlinearity and linearly degeneracy can be introduced as in the one-dimensional case (cf. Chapter 2) and lead to an existence and uniqueness result for the solution of generalised Riemann problems, see [121].

### 4.3 Optimality conditions

In this section, we derive the optimality conditions for the optimal control problem (4.1). For a given domain  $\mathcal{D} \subseteq \mathbb{R}^d$ , we can write the Lagrangian function of the problem (4.1) - (4.3) as

$$\begin{aligned} L(\cdot) = L(U(T, \cdot), U_0, P; U_d) &= \frac{1}{2} \int_{\mathcal{D}} (U(T, \mathbf{x}; U_0) - U_d(\mathbf{x}))^2 d\mathbf{x} \\ &+ \int_0^T \int_{\mathcal{D}} P' \left( \frac{\partial U}{\partial t} + \nabla \cdot f(U) \right) d\mathbf{x} dt, \end{aligned} \quad (4.10)$$

where  $P$  is the adjoint variable or Lagrange multiplier, which is assumed to be a smooth function with compact support in  $\mathcal{D}$  and vanishing on the boundaries of  $\mathcal{D}$  and  $'$  denotes the transpose. Integrating by parts, the last two terms on the right-hand side of (4.10) and since  $f(U)$  is vanishing on the boundaries of  $\mathcal{D}$ , we have

$$L(\cdot) = \frac{1}{2} \int_{\mathcal{D}} (U(T, \mathbf{x}; U_0) - U_d(\mathbf{x}))^2 d\mathbf{x} + \int_{\mathcal{D}} (P'(T, \mathbf{x})U(T, \mathbf{x}) - P'(0, \mathbf{x})U(0, \mathbf{x})) d\mathbf{x} \\ - \int_0^T \int_{\mathcal{D}} U' \frac{\partial}{\partial t} P d\mathbf{x} dt - \int_0^T \int_{\mathcal{D}} f'(U) \nabla \cdot P d\mathbf{x} dt \quad (4.11)$$

Setting the variation of the Lagrange functional (4.11) with respect to  $U$  equal to zero, we get

$$\frac{\partial L(\cdot)}{\partial U} = \int_0^T \int_{\mathcal{D}} \left( -\frac{\partial}{\partial t} P - Df(U) \nabla \cdot P \right) d\mathbf{x} dt = 0, \quad (4.12)$$

where  $Df(U)$  is the Jacobian matrix of the flux function  $f(U)$  with respect to  $U$ . Thus, from (4.12) we obtain the adjoint system as

$$-\frac{\partial}{\partial t} P(t, \mathbf{x}) - Df(U) \nabla \cdot P(t, \mathbf{x}) = 0, \quad \text{a.e. } \mathbf{x} \in \mathcal{D}. \quad (4.13)$$

Setting the partial derivatives of  $L(\cdot)$  in (4.11) with respect to  $U(T, \mathbf{x})$  equal to zero, we have the terminal data as

$$P(T, \mathbf{x}) = U(T, \mathbf{x}; U_0) - U_d(\mathbf{x}), \quad \text{a.e. } \mathbf{x} \in \mathcal{D}, \quad (4.14)$$

where a.e. stands for almost everywhere. Moreover, a gradient of the cost functional can be obtained by setting the partial derivatives of  $L(\cdot)$  in (4.11) with respect to  $U_0$  equal zero and given as

$$\nabla_{U_0} \tilde{J}(\cdot) = \int_{\mathcal{D}} P(0, \mathbf{x}) d\mathbf{x}, \quad \text{a.e. } \mathbf{x} \in \mathcal{D}, \quad (4.15)$$

where  $\tilde{J}(\cdot) = \tilde{J}(U_0; U_d)$  is the reduced cost functional. The coupled systems (4.1)-(4.3) and (4.13), (4.14) together with the gradient (4.15) represent the formal first-order optimality conditions for the problem (4.1)-(4.3), in which (4.2) should be solved forward in time while the adjoint system (4.13) with (4.14) should be solved backwards in time.

## 4.4 Space and time discretisation

Briefly, this section presents the numerical schemes for solving the optimal control problems governed by the 2D hyperbolic systems of conservation laws. Given  $\mathcal{D} \subseteq \mathbb{R}^2$ , we have

$$\frac{\partial U}{\partial t} + \frac{\partial}{\partial x} f(U) + \frac{\partial}{\partial y} g(U) = 0, \quad U(0, x, y) = U_0(x, y), \quad (4.16)$$

$$-\frac{\partial P}{\partial t} - Df(U) \frac{\partial P}{\partial x} - Dg(U) \frac{\partial P}{\partial y} = 0, \quad P(T, x, y) = U(T, x, y) - U_d(x, y), \quad (4.17)$$

$$\nabla_{U_0} \tilde{J}(\cdot) = \iint_{\mathcal{D}} P(0, x, y) dx dy \quad \text{a.e. } (x, y) \in \mathcal{D}, \quad (4.18)$$

where  $U(t, x, y) \in \mathbb{R}^m$ ,  $t \in [0, T]$ ,  $(x, y) \in \mathcal{D}$  and  $f(U)$  and  $g(U)$  are flux functions with Jacobian matrices  $Df(U)$  and  $Dg(U)$ , respectively. However, to minimise the cost functional

$$J(\cdot) = \frac{1}{2} \iint_{\mathcal{D}} (U(T, x, y; U_0) - U_d(x, y))^2 dx dy, \quad (4.19)$$

numerically, we must solve equation (4.16) forward in time and equation (4.17) backwards in time, then, we update the initial function using equation (4.18).

### 4.4.1 Forward equations

For the space discretisation of the initial value problem (4.16), we consider the space domain with rectangular cells  $C_{j,k} = [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}] \times [y_{k-\frac{1}{2}}, y_{k+\frac{1}{2}}]$  of uniform sizes  $\Delta x = x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}$  and  $\Delta y = y_{k+\frac{1}{2}} - y_{k-\frac{1}{2}}$  and a uniform time step  $\Delta t = t^{n+1} - t^n$ . The cells,  $C_{j,k}$ , are centred at  $(x_j, y_k)$  with  $x_j = j\Delta x$  and  $y_k = k\Delta y$  and the grid points are defined as  $x_j = (x_j + x_{j+1})/2$  and  $y_k = (y_k + y_{k+1})/2$ . We assume that cell averages are introduced as

$$U_{j,k}^n \equiv \bar{U}(t, x_j, y_k) = \frac{1}{\Delta x \Delta y} \iint_{C_{j,k}} U(t, x, y) dx dy, \quad (4.20)$$

and the approximations of the mean value of the fluxes,  $F_{j+\frac{1}{2},k}^n$  and  $G_{j,k+\frac{1}{2}}^n$ , are defined as

$$F_{j+\frac{1}{2},k}^n := \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} f(U(t, x_{j+\frac{1}{2}}, y_k)) dt, \quad G_{j,k+\frac{1}{2}}^n := \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} g(U(t, x_j, y_{k+\frac{1}{2}})) dt. \quad (4.21)$$

Thus, the finite volume scheme that approximates the equation (4.16) given by the following conservative formulation

$$U_{j,k}^{n+1} = U_{j,k}^n - \frac{\Delta t}{\Delta x} (F_{j+\frac{1}{2},k}^n - F_{j-\frac{1}{2},k}^n) - \frac{\Delta t}{\Delta y} (G_{j,k+\frac{1}{2}}^n - G_{j,k-\frac{1}{2}}^n), \quad (4.22)$$

where the numerical fluxes  $F_{j+\frac{1}{2},k}^n$  and  $G_{j,k+\frac{1}{2}}^n$  are given in the sense of Lax-Friedrichs as

$$F_{j+\frac{1}{2},k}^n = \frac{1}{2} \left( f(U_{j,k}^n) + f(U_{j+1,k}^n) \right) - \frac{\alpha_x}{2} \left( U_{j+1,k}^n - U_{j,k}^n \right), \quad (4.23)$$

and

$$G_{j,k+\frac{1}{2}}^n = \frac{1}{2} \left( g(U_{j,k}^n) + g(U_{j,k+1}^n) \right) - \frac{\alpha_y}{2} \left( U_{j,k+1}^n - U_{j,k}^n \right), \quad (4.24)$$

where  $\alpha_x = \frac{\Delta x}{\Delta t}$  and  $\alpha_y = \frac{\Delta y}{\Delta t}$ . Also, different numerical fluxes can be used with the scheme (4.22), such as the semi-discrete scheme [47, 52, 122, 123], the WENO-ZQ scheme [124, 125] and weighted compact central schemes [126, 127].

Moreover, using equations (4.22), (4.23) and (4.24), we can obtain the finite volume method (Lax-Friedrichs scheme) for solving the flow equation (4.16)

$$U_{j,k}^{n+1} = \frac{1}{2} \left( U_{j+1,k}^n + U_{j-1,k}^n \right) + \frac{1}{2} \left( U_{j,k+1}^n + U_{j,k-1}^n \right) - U_{j,k}^n - \frac{\Delta t}{2\Delta x} \left( f(U_{j+1,k}^n) + f(U_{j-1,k}^n) \right) - \frac{\Delta t}{2\Delta y} \left( g(U_{j,k+1}^n) + g(U_{j,k-1}^n) \right), \quad n = 0, 1, \dots, K. \quad (4.25)$$

#### 4.4.2 Backward equations

To solve the adjoint equations (4.17), numerically, we use similar discretisation that presented for the flow equations in Subsection 4.4.1 to discretise the adjoint equations (4.17), which can be solved backwards in time. Furthermore, we have the numerical scheme, which approximates the solution of equations (4.17) in the form

$$P_{j,k}^n = P_{j,k}^{n+1} - \frac{\Delta t}{\Delta x} Df(U_{j,k}^{n+1}) \left( P_{j+\frac{1}{2},k}^{n+1} - P_{j-\frac{1}{2},k}^{n+1} \right) - \frac{\Delta t}{\Delta y} Dg(U_{j,k}^{n+1}) \left( P_{j,k+\frac{1}{2}}^{n+1} - P_{j,k-\frac{1}{2}}^{n+1} \right), \quad (4.26)$$

where  $P_{j+\frac{1}{2},k}^n = P(t^{n+1}, x_{j+\frac{1}{2}}, y_k)$  and  $P_{j,k+\frac{1}{2}}^{n+1} = P(t^{n+1}, x_j, y_{k+\frac{1}{2}})$ , defined respectively as

$$P_{j+\frac{1}{2},k}^{n+1} = \frac{1}{2} Df(U_{j,k}^{n+1}) \left( P_{j,k}^{n+1} + P_{j+1,k}^{n+1} \right) - \frac{\alpha_x}{2} \left( P_{j+1,k}^{n+1} - P_{j,k}^{n+1} \right), \quad (4.27)$$

and

$$P_{j,k+\frac{1}{2}}^{n+1} = \frac{1}{2}Dg(U_{j,k}^{n+1}) \left( P_{j,k}^{n+1} + P_{j,k+1}^{n+1} \right) - \frac{\alpha_y}{2} \left( P_{j,k+1}^{n+1} - P_{j,k}^{n+1} \right). \quad (4.28)$$

Equations (4.26), (4.27), and (4.28) gives the numerical approximation as

$$\begin{aligned} P_{j,k}^n &= P_{j,k}^{n+1} - \left( Df(U_{j,k}^{n+1}) + Dg(U_{j,k}^{n+1}) \right) P_{j,k}^{n+1} - \frac{\Delta t}{2\Delta x} \left( Df(U_{j,k}^{n+1}) \right)^2 \left( P_{j+1,k}^{n+1} - P_{j-1,k}^{n+1} \right) \\ &+ \frac{1}{2}Df(U_{j,k}^{n+1}) \left( P_{j+1,k}^{n+1} + P_{j-1,k}^{n+1} \right) - \frac{\Delta t}{2\Delta y} \left( Dg(U_{j,k}^{n+1}) \right)^2 \left( P_{j,k+1}^{n+1} - P_{j,k-1}^{n+1} \right) \\ &+ \frac{1}{2}Dg(U_{j,k}^{n+1}) \left( P_{j,k+1}^{n+1} + P_{j,k-1}^{n+1} \right), \quad n = K, K-1, \dots, 0. \end{aligned} \quad (4.29)$$

Finally, to complete the solution of the underlining optimisation problem, we apply the numerical algorithm in the same way as proposed below in Section 4.6. From equation (4.18), we have the following discrete version of the reduced cost functional

$$\nabla_{U_{0,j,k}} \tilde{J}(\cdot) = \sum_{j=1}^N \sum_{k=1}^M P(0, x_j, y_k) \Delta x \Delta y. \quad (4.30)$$

The discrete terminal data with the terminal time,  $T$ , are given by

$$P(T, x_j, y_k) = U(T, x_j, y_k) - U_d(x_j, y_k). \quad (4.31)$$

## 4.5 Three-dimensional relaxation approach

The nonlinearity appearing in equation (4.2) makes the computation challenging. Moreover, the presence of shock and other discontinuities in the solution of the flow solver poses problems to the backward equation (4.13). Therefore, in our approach, we replace the flow equation (4.2) with the corresponding relaxation system given by

$$\begin{aligned} \frac{\partial}{\partial t} U + \sum_{i=1}^d \frac{\partial}{\partial x_i} V_i &= 0, \\ \frac{\partial}{\partial t} V_i + A_i^2 \frac{\partial}{\partial x_i} U &= -\frac{1}{\varepsilon} (V_i - f_i(U)), \quad i = 1, 2, \dots, d \end{aligned} \quad (4.32)$$

where  $V_i \in \mathbb{R}^m$  are the artificial relaxation variables;  $\varepsilon$  is a small positive parameter that measures the relaxation rate;  $A_i = \text{diag}\{a_{i,1}, \dots, a_{i,m}\}$  are positive diagonal matrices that satisfy the sub-characteristic condition [9]

$$\sum_{i=1}^m \frac{(\lambda_i(Df_i(U)))^2}{a_i^2} \leq 1, \quad \forall U \in \mathbb{R}^m, \quad (4.33)$$

where  $\lambda_i$  are eigenvalues of  $Df_i(U)$ , with the initial data for the relaxation system (4.32) stated as

$$U(0, \mathbf{x}) = U_0(\mathbf{x}), \quad V_i(0, \mathbf{x}) = f_i(U_0(\mathbf{x})), \quad i = 1, 2, \dots, d. \quad (4.34)$$

In [69], optimal control problems governed by two-dimensional conservation laws were considered based on a relaxation approximation algorithm. We extend here the optimal control problems to the three-dimensional case following the same approach. The three-dimensional (3D) optimisation problem of minimising the cost functional of tracking type given as

$$\begin{aligned} \min J(\cdot) = \frac{1}{2} \iiint_{\mathcal{D}} & \left( (U(T, x, y, z) - U_d(x, y, z))^2 + (V(T, x, y, z) - f(U_d(x, y, z)))^2 \right. \\ & \left. + (W(T, x, y, z) - g(U_d(x, y, z)))^2 + (H(T, x, y, z) - s(U_d(x, y, z)))^2 \right) dx dy dz, \end{aligned} \quad (4.35)$$

for a given open bounded domain  $\mathcal{D}$  in  $\mathbb{R}^3$ . Where constraints are provided by 3D relaxation system

$$\begin{aligned} \frac{\partial U}{\partial t} + \frac{\partial V}{\partial x} + \frac{\partial W}{\partial y} + \frac{\partial H}{\partial z} &= 0, \\ \frac{\partial V}{\partial t} + a^2 \frac{\partial U}{\partial x} &= -\frac{1}{\varepsilon} (V - f(U)), \\ \frac{\partial W}{\partial t} + b^2 \frac{\partial U}{\partial y} &= -\frac{1}{\varepsilon} (W - g(U)), \\ \frac{\partial H}{\partial t} + c^2 \frac{\partial U}{\partial z} &= -\frac{1}{\varepsilon} (H - s(U)), \end{aligned} \quad (4.36)$$

with the initial data given as

$$\begin{aligned} U(0, x, y, z) &= U_0(x, y, z), \quad V(0, x, y, z) = f(U_0(x, y, z)), \quad W(0, x, y, z) = g(U_0(x, y, z)), \\ \text{and } H(0, x, y, z) &= s(U_0(x, y, z)). \end{aligned} \quad (4.37)$$

Here  $(x, y, z) \in \mathcal{D}$  are space coordinates;  $V, W, H \in \mathbb{R}$  are relaxation variables and  $a, b$ , and  $c$  are characteristic speeds. Then,  $U(t, x, y, z) \in \mathbb{R}$  is the conserved quantity with the corresponding nonlinear flux functions  $f(U)$ ,  $g(U)$  and  $s(U)$ .

To derive the first-order optimality conditions for the problem (4.35), we consider the Lagrangian

function as

$$\begin{aligned}
L(\cdot) = & J(\cdot) + \int_0^T \iiint_{\mathcal{D}} p \left( \frac{\partial U}{\partial t} + \frac{\partial V}{\partial x} + \frac{\partial W}{\partial y} + \frac{\partial H}{\partial z} \right) dx dy dz dt \\
& + \int_0^T \iiint_{\mathcal{D}} q \left( \frac{\partial V}{\partial t} + a^2 \frac{\partial U}{\partial x} + \frac{1}{\varepsilon} (V - f(U)) \right) dx dy dz dt \\
& + \int_0^T \iiint_{\mathcal{D}} h \left( \frac{\partial W}{\partial t} + b^2 \frac{\partial U}{\partial y} + \frac{1}{\varepsilon} (W - g(U)) \right) dx dy dz dt \\
& + \int_0^T \iiint_{\mathcal{D}} r \left( \frac{\partial H}{\partial t} + c^2 \frac{\partial U}{\partial z} + \frac{1}{\varepsilon} (H - s(U)) \right) dx dy dz dt,
\end{aligned} \tag{4.38}$$

where  $p, q, h, r \in \mathcal{D}$  are Lagrange multipliers, assumed to be smooth functions with compact support in  $\mathcal{D}$  and vanish on the boundaries of  $\mathcal{D}$ , denoted by  $\partial\mathcal{D} = (\partial\mathcal{D}_x, \partial\mathcal{D}_y, \partial\mathcal{D}_z)$ . Therefore, using the integration by parts in equation (4.38), we have

$$\begin{aligned}
L(\cdot) = & \iiint_{\mathcal{D}} \left( pU|_0^T - \int_0^T U \frac{\partial p}{\partial t} dt \right) dx dy dz + \iiint_{\mathcal{D}} \left( qV|_0^T - \int_0^T V \frac{\partial q}{\partial t} dt \right) dx dy dz \\
& + \iiint_{\mathcal{D}} \left( hW|_0^T - \int_0^T W \frac{\partial h}{\partial t} dt \right) dx dy dz + \iiint_{\mathcal{D}} \left( rH|_0^T - \int_0^T H \frac{\partial r}{\partial t} dt \right) dx dy dz \\
& + \int_0^T \iint_{(\mathcal{D}_y, \mathcal{D}_z)} \left( pV|_{\partial\mathcal{D}_x} - \int_{\mathcal{D}_x} V \frac{\partial p}{\partial x} dx \right) dy dz dt \\
& + a^2 \int_0^T \iint_{(\mathcal{D}_y, \mathcal{D}_z)} \left( qU|_{\partial\mathcal{D}_x} - \int_{\mathcal{D}_x} U \frac{\partial q}{\partial x} dx \right) dy dz dt \\
& + \int_0^T \iint_{(\mathcal{D}_x, \mathcal{D}_z)} \left( pW|_{\partial\mathcal{D}_y} - \int_{\mathcal{D}_y} W \frac{\partial p}{\partial y} dy \right) dx dz dt \\
& + b^2 \int_0^T \iint_{(\mathcal{D}_x, \mathcal{D}_z)} \left( hU|_{\partial\mathcal{D}_y} - \int_{\mathcal{D}_y} U \frac{\partial h}{\partial y} dy \right) dx dz dt \\
& + \int_0^T \iint_{(\mathcal{D}_x, \mathcal{D}_y)} \left( pH|_{\partial\mathcal{D}_z} - \int_{\mathcal{D}_z} H \frac{\partial p}{\partial z} dz \right) dx dy dt \\
& + c^2 \int_0^T \iint_{(\mathcal{D}_x, \mathcal{D}_y)} \left( rU|_{\partial\mathcal{D}_z} - \int_{\mathcal{D}_z} U \frac{\partial r}{\partial z} dz \right) dx dy dt \\
& + \frac{1}{\varepsilon} \int_0^T \iiint_{\mathcal{D}} (qV) dx dy dz dt - \frac{1}{\varepsilon} \int_0^T \iiint_{\mathcal{D}} (qf(U)) dx dy dz dt \\
& + \frac{1}{\varepsilon} \int_0^T \iiint_{\mathcal{D}} (hW) dx dy dz dt - \frac{1}{\varepsilon} \int_0^T \iiint_{\mathcal{D}} (hg(U)) dx dy dz dt \\
& + \frac{1}{\varepsilon} \int_0^T \iiint_{\mathcal{D}} (rH) dx dy dz dt - \frac{1}{\varepsilon} \int_0^T \iiint_{\mathcal{D}} (rs(U)) dx dy dz dt.
\end{aligned} \tag{4.39}$$

This implies that



$$\begin{aligned}
L(\cdot) = & J(\cdot) + \iiint_{\mathcal{D}} (p(T,x,y,z)U(T,x,y,z)) dx dy dz - \iiint_{\mathcal{D}} (p(0,x,y,z)U(0,x,y,z)) dx dy dz \\
& + \iiint_{\mathcal{D}} (q(T,x,y,z)V(T,x,y,z)) dx dy dz - \iiint_{\mathcal{D}} (q(0,x,y,z)V(0,x,y,z)) dx dy dz \\
& + \iiint_{\mathcal{D}} (h(T,x,y,z)W(T,x,y,z)) dx dy dz - \iiint_{\mathcal{D}} (h(0,x,y,z)W(0,x,y,z)) dx dy dz \\
& + \iiint_{\mathcal{D}} (r(T,x,y,z)H(T,x,y,z)) dx dy dz - \iiint_{\mathcal{D}} (r(0,x,y,z)H(0,x,y,z)) dx dy dz \\
& - \int_0^T \iiint_{\mathcal{D}} \left( U \frac{\partial p}{\partial t} + V \frac{\partial q}{\partial t} + W \frac{\partial h}{\partial t} + H \frac{\partial r}{\partial t} \right) dx dy dz dt \\
& - \int_0^T \iiint_{\mathcal{D}} \left( V \frac{\partial p}{\partial x} + a^2 U \frac{\partial q}{\partial x} + W \frac{\partial p}{\partial y} + b^2 U \frac{\partial h}{\partial y} + H \frac{\partial p}{\partial z} + c^2 U \frac{\partial r}{\partial z} \right) dx dy dz dt \\
& + \frac{1}{\varepsilon} \int_0^T \iiint_{\mathcal{D}} (qV + hW + rH) dx dy dz dt \\
& - \frac{1}{\varepsilon} \int_0^T \iiint_{\mathcal{D}} (qf(U) + hg(U) + rs(U)) dx dy dz dt
\end{aligned} \tag{4.40}$$

Thus, we can rewrite equation (4.40) as follows

$$\begin{aligned}
L(\cdot) = & J(\cdot) + \iiint_{\mathcal{D}} (p(T,x,y,z)U(T,x,y,z)) dx dy dz - \iiint_{\mathcal{D}} (p(0,x,y,z)U(0,x,y,z)) dx dy dz \\
& + \iiint_{\mathcal{D}} (q(T,x,y,z)V(T,x,y,z)) dx dy dz - \iiint_{\mathcal{D}} (q(0,x,y,z)V(0,x,y,z)) dx dy dz \\
& + \iiint_{\mathcal{D}} (h(T,x,y,z)W(T,x,y,z)) dx dy dz - \iiint_{\mathcal{D}} (h(0,x,y,z)W(0,x,y,z)) dx dy dz \\
& + \iiint_{\mathcal{D}} (r(T,x,y,z)H(T,x,y,z)) dx dy dz - \iiint_{\mathcal{D}} (r(0,x,y,z)H(0,x,y,z)) dx dy dz \\
& + \int_0^T \iiint_{\mathcal{D}} U \left( -\frac{\partial p}{\partial t} - a^2 \frac{\partial q}{\partial x} - b^2 \frac{\partial h}{\partial y} - c^2 \frac{\partial r}{\partial z} - \frac{1}{\varepsilon} qf'(U) - \frac{1}{\varepsilon} hg'(U) - \frac{1}{\varepsilon} rs'(U) \right) dx dy dz dt \\
& + \int_0^T \iiint_{\mathcal{D}} V \left( -\frac{\partial q}{\partial t} - \frac{\partial p}{\partial x} + \frac{1}{\varepsilon} q \right) dx dy dz dt \\
& + \int_0^T \iiint_{\mathcal{D}} W \left( -\frac{\partial h}{\partial t} - \frac{\partial p}{\partial y} + \frac{1}{\varepsilon} h \right) dx dy dz dt \\
& + \int_0^T \iiint_{\mathcal{D}} H \left( -\frac{\partial r}{\partial t} - \frac{\partial p}{\partial z} + \frac{1}{\varepsilon} r \right) dx dy dz dt.
\end{aligned} \tag{4.41}$$

Setting the derivatives of  $L(\cdot)$  in (4.41) with respect to  $U$ ,  $V$ ,  $W$  and  $H$  equal to zero, respectively,

we have the adjoint system in 3D as

$$\begin{aligned}
-\frac{\partial p}{\partial t} - a^2 \frac{\partial q}{\partial x} - b^2 \frac{\partial h}{\partial y} - c^2 \frac{\partial r}{\partial z} &= \frac{1}{\varepsilon} (f'(U)q + g'(U)h + s'(U)r), \\
-\frac{\partial q}{\partial t} - \frac{\partial p}{\partial x} &= -\frac{1}{\varepsilon} q, \\
-\frac{\partial h}{\partial t} - \frac{\partial p}{\partial y} &= -\frac{1}{\varepsilon} h, \\
-\frac{\partial r}{\partial t} - \frac{\partial p}{\partial z} &= -\frac{1}{\varepsilon} r,
\end{aligned} \tag{4.42}$$

where  $f'(U)$ ,  $g'(U)$  and  $s'(U)$  are derivatives of the flux functions  $f(U)$ ,  $g(U)$  and  $s(U)$  with respect to  $U$ , respectively.

With the terminal data  $p(T, x, y, z)$ ,  $q(T, x, y, z)$ ,  $h(T, x, y, z)$  and  $r(T, x, y, z)$  that can be obtained by setting the derivatives of  $L(\cdot)$  in (4.41) with respect to  $U(T, x, y, z)$ ,  $V(T, x, y, z)$ ,  $W(T, x, y)$  and  $H(T, x, y, z)$ , respectively, equal to zero as

$$\begin{aligned}
P(T, x, y, z) &= U(T, x, y, z) - U_d(x, y, z), \\
q(T, x, y, z) &= V(T, x, y, z) - V_d(x, y, z), \\
h(T, x, y, z) &= W(T, x, y, z) - W_d(x, y, z), \\
r(T, x, y, z) &= H(T, x, y, z) - H_d(x, y, z).
\end{aligned} \tag{4.43}$$

Moreover, the last three equations of the adjoint system (4.42) give

$$q = \varepsilon \frac{\partial}{\partial x} p + O(\varepsilon^2), \quad h = \varepsilon \frac{\partial}{\partial y} p + O(\varepsilon^2), \quad \text{and} \quad r = \varepsilon \frac{\partial}{\partial z} p + O(\varepsilon^2). \tag{4.44}$$

By substituting (4.44) into the first equation of (4.42), we have

$$-\frac{\partial p}{\partial t} - f'(U) \frac{\partial p}{\partial x} - g'(U) \frac{\partial p}{\partial y} - s'(U) \frac{\partial p}{\partial z} = \varepsilon \left( a^2 \frac{\partial^2 p}{\partial x^2} + b^2 \frac{\partial^2 p}{\partial y^2} + c^2 \frac{\partial^2 p}{\partial z^2} \right), \tag{4.45}$$

which is a viscous approximation to the adjoint system (4.13). Since  $U$  may develop discontinuities, we have to deal with discontinuous derivatives of the flux functions  $f'(U)$ ,  $g'(U)$  and  $s'(U)$ . However, since we use the relaxation approximation, the derivative functions  $f'(U)$ ,  $g'(U)$  and  $s'(U)$  appear as source terms and not as a discontinuous transport coefficient as in (4.45).

Finally, the gradient of the reduced cost functional in the 3D problem can be found by setting the

derivatives of  $L(\cdot)$  in (4.41) with respect to  $U(0,x,y,z)$ ,  $V(0,x,y,z)$ ,  $W(0,x,y,z)$  and  $H(0,x,y,z)$  equal to zero, respectively, giving

$$\begin{aligned} \nabla_{U_0} \tilde{J}(\cdot) = & \iiint_{\mathcal{D}} (p(0,x,y,z) + f'(U(0,x,y,z))q(0,x,y,z) + g'(U(0,x,y,z))h(0,x,y,z) \\ & + s'(U(0,x,y,z))r(0,x,y,z)) dx dy dz, \end{aligned} \quad (4.46)$$

where  $\tilde{J}(\cdot)$  represents the reduced cost functional.

### 4.5.1 Three-dimensional scheme

In this section, the discretisations of 3D problems are considered. For the 2D problems, we use the relaxation schemes proposed by Herty et al. in [69] to solve the problem numerically. The numerical scheme for the flow equations proposed by Seaid in [128] can also be considered for both forward and adjoint equations, with the adjoint equation solved backward in time. In 3D discretisations, we use spatial splitting based on the characteristic variables of the relaxation system (4.36)

$$V \pm aU, \quad W \pm bU \quad \text{and} \quad H \pm cU. \quad (4.47)$$

However, the obtained adjoint formulation in the characteristic form is the same as the adjoint of the characteristic form when applying a dimensional splitting in the spatial variable. We divide the spatial domain into cells  $C_{i,j,k} = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \times [y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}] \times [z_{k-\frac{1}{2}}, z_{k+\frac{1}{2}}]$  with uniform sizes  $\Delta x$ ,  $\Delta y$  and  $\Delta z$  and a uniform time step  $\Delta t = t^{n+1} - t^n$ . The cells,  $C_{i,j,k}$ , are centred at  $(x_i, y_j, z_k)$  with  $x_i = i\Delta x$ ,  $y_j = j\Delta y$ , and  $z_k = k\Delta z$ . As in [82], we use the point-values of a generic function  $\omega$ , denoted as  $\omega_{i\pm\frac{1}{2},j,k} := \omega(t, x_{i\pm\frac{1}{2}}, y_j, z_k)$ ,  $\omega_{i,j\pm\frac{1}{2},k} := \omega(t, x_i, y_{j\pm\frac{1}{2}}, z_k)$  and  $\omega_{i,j,k\pm\frac{1}{2}} := \omega(t, x_i, y_j, z_{k\pm\frac{1}{2}})$  at  $(t, x_{i\pm\frac{1}{2}}, y_j, z_k)$ ,  $(t, x_i, y_{j\pm\frac{1}{2}}, z_k)$  and  $(t, x_i, y_j, z_{k\pm\frac{1}{2}})$ , respectively. The approximate cell-average  $\omega_{i,j,k}$  at  $(t, x_i, y_j, z_k)$  given by

$$\omega_{i,j,k}(t) := \frac{1}{\Delta x \Delta y \Delta z} \iiint_{C_{i,j,k}} \omega(t, x, y, z) dx dy dz. \quad (4.48)$$

Moreover, we define the following finite differences

$$D_x \omega_{i,j,k} = \frac{\omega_{i+\frac{1}{2},j,k} - \omega_{i-\frac{1}{2},j,k}}{\Delta x}, \quad D_y \omega_{i,j,k} = \frac{\omega_{i,j+\frac{1}{2},k} - \omega_{i,j-\frac{1}{2},k}}{\Delta y}, \quad D_z \omega_{i,j,k} = \frac{\omega_{i,j,k+\frac{1}{2}} - \omega_{i,j,k-\frac{1}{2}}}{\Delta z}. \quad (4.49)$$

Thus, the semi-discrete approximation of the relaxation system (4.36) reads

$$\begin{aligned}
\frac{dU_{i,j,k}}{dt} + D_x V_{i,j,k} + D_y W_{i,j,k} + D_z H_{i,j,k} &= 0, \\
\frac{dV_{i,j,k}}{dt} + a^2 D_x U_{i,j,k} &= -\frac{1}{\varepsilon} (V_{i,j,k} - f(U_{i,j,k})), \\
\frac{dW_{i,j,k}}{dt} + b^2 D_y U_{i,j,k} &= -\frac{1}{\varepsilon} (W_{i,j,k} - g(U_{i,j,k})), \\
\frac{dH_{i,j,k}}{dt} + c^2 D_z U_{i,j,k} &= -\frac{1}{\varepsilon} (H_{i,j,k} - s(U_{i,j,k})).
\end{aligned} \tag{4.50}$$

Analogously, the semi-discrete approximation of the adjoint system (4.42) is

$$\begin{aligned}
-\frac{dp_{i,j,k}}{dt} - a^2 D_x q_{i,j,k} - b^2 D_y h_{i,j,k} - c^2 D_z r_{i,j,k} \\
&= \frac{1}{\varepsilon} (f'(U_{i,j,k})q_{i,j,k} + g'(U_{i,j,k})h_{i,j,k} + s'(U_{i,j,k})r_{i,j,k}), \\
-\frac{dq_{i,j,k}}{dt} - D_x p_{i,j,k} &= -\frac{1}{\varepsilon} q_{i,j,k}, \\
-\frac{dh_{i,j,k}}{dt} - D_y p_{i,j,k} &= -\frac{1}{\varepsilon} h_{i,j,k}, \\
-\frac{dr_{i,j,k}}{dt} - D_z p_{i,j,k} &= -\frac{1}{\varepsilon} r_{i,j,k}.
\end{aligned} \tag{4.51}$$

The semi-discrete formulations (4.50) or (4.51) can be rewritten in the standard notation of ordinary differential equations as

$$\frac{d\mathcal{Y}}{dt} = \mathcal{F}(\mathcal{Y}) - \frac{1}{\varepsilon} \mathcal{G}(\mathcal{Y}). \tag{4.52}$$

where the time-dependent vector functions  $\mathcal{Y}$ ,  $\mathcal{F}(\mathcal{Y})$  and  $\mathcal{G}(\mathcal{Y})$  are defined accordingly for the forward problem (4.50) or the backward problem (4.51). The non-stiff stage of the splitting for  $\mathcal{F}$  is treated by an explicit scheme while a diagonally implicit scheme will be applied for the stiff stage for  $\mathcal{G}$ . Then, we can formulate the first-order implicit- explicit scheme (IMEX) as presented in [82] for the forward system (4.52) as

$$\begin{aligned}
K_1 &= \mathcal{Y}^n - \frac{\Delta t}{\varepsilon} \mathcal{G}(K_1), \\
\mathcal{Y}^{n+1} &= \mathcal{Y}^n + \Delta t \mathcal{F}(K_1) - \frac{\Delta t}{\varepsilon} \mathcal{G}(K_1),
\end{aligned} \tag{4.53}$$

with functions  $\mathcal{Y}$ ,  $\mathcal{F}(\mathcal{Y})$  and  $\mathcal{G}(\mathcal{Y})$  given by

$$\mathcal{Y} = \begin{bmatrix} U_{i,j,k} \\ V_{i,j,k} \\ W_{i,j,k} \\ H_{i,j,k} \end{bmatrix}, \quad \mathcal{F} = \begin{bmatrix} -K_2 \\ -a^2 D_x U_{i,j,k} \\ -b^2 D_y U_{i,j,k} \\ -c^2 D_z U_{i,j,k} \end{bmatrix}, \quad \mathcal{G} = \begin{bmatrix} 0 \\ V_{i,j,k} - f(U_{i,j,k}) \\ W_{i,j,k} - g(U_{i,j,k}) \\ H_{i,j,k} - s(U_{i,j,k}) \end{bmatrix}, \quad (4.54)$$

where  $K_2 = D_x V_{i,j,k} + D_y W_{i,j,k} + D_z H_{i,j,k}$ .

For the backward system (4.52), the IMEX scheme is read as

$$\begin{aligned} K_1 &= \mathcal{Y}^n + \Delta t \mathcal{F}(K_1), \\ \mathcal{Y}^{n+1} &= \mathcal{Y}^n + \Delta t \mathcal{F}(K_1) - \frac{\Delta t}{\varepsilon} \mathcal{G}(K_1), \end{aligned} \quad (4.55)$$

here  $\mathcal{Y}$ ,  $\mathcal{F}(\mathcal{Y})$  and  $\mathcal{G}(\mathcal{Y})$  are stated as

$$\mathcal{Y} = \begin{bmatrix} p_{i,j,k} \\ q_{i,j,k} \\ h_{i,j,k} \\ r_{i,j,k} \end{bmatrix}, \quad \mathcal{F} = \begin{bmatrix} -K_3 \\ -D_x p_{i,j,k} \\ -D_y p_{i,j,k} \\ -D_z p_{i,j,k} \end{bmatrix}, \quad \mathcal{G} = \begin{bmatrix} K_4 \\ -q_{i,j,k} \\ -h_{i,j,k} \\ -r_{i,j,k} \end{bmatrix}, \quad (4.56)$$

where

$$K_3 = a^2 D_x q_{i,j,k} + b^2 D_y h_{i,j,k} + c^2 D_z r_{i,j,k}, \quad K_4 = f'(U_{i,j,k}) q_{i,j,k} + g'(U_{i,j,k}) h_{i,j,k} + s'(U_{i,j,k}) r_{i,j,k}.$$

When using the relaxation strategy described above, neither linear algebraic equations nor nonlinear source terms are required. Furthermore, the relaxation scheme is stable independently of  $\varepsilon$  so that the choice of  $\Delta t$  is based only on the usual CFL condition

$$\max \left( \frac{\Delta t}{\delta}, a^2 \frac{\Delta t}{\Delta x}, b^2 \frac{\Delta t}{\Delta y}, c^2 \frac{\Delta t}{\Delta z} \right) \leq 1, \quad (4.57)$$

where  $\delta = \max(\Delta x, \Delta y, \Delta z)$  denotes the maximum cell size.

However, for the space discretisation, applying a first-order upwind scheme to the characteristic variables (4.47) gives

$$\begin{aligned} (V + aU)_{i+\frac{1}{2},j,k} &= (V + aU)_{i,j,k}, & (V - aU)_{i+\frac{1}{2},j,k} &= (V - aU)_{i+1,j,k}, \\ (W + bU)_{i,j+\frac{1}{2},k} &= (W + bU)_{i,j,k}, & (W - bU)_{i,j+\frac{1}{2},k} &= (W - bU)_{i,j+1,k}, \\ (H + cU)_{i,j,k+\frac{1}{2}} &= (H + cU)_{i,j,k}, & (H - cU)_{i,j,k+\frac{1}{2}} &= (H - cU)_{i,j,k+1}. \end{aligned} \quad (4.58)$$

Solving (4.58), we obtain a first-order reconstruction of the forward problem (4.50) as

$$\begin{aligned}
V_{i+\frac{1}{2},j,k} &= \frac{1}{2}(V_{i,j,k} + v_{i+1,j,k}) - \frac{a}{2}(U_{i+1,j,k} - U_{i,j,k}), \\
U_{i+\frac{1}{2},j,k} &= \frac{1}{2}(U_{i,j,k} + U_{i+1,j,k}) - \frac{1}{2a}(V_{i+1,j,k} - V_{i,j,k}), \\
W_{i,j+\frac{1}{2},k} &= \frac{1}{2}(W_{i,j,k} + W_{i,j+1,k}) - \frac{b}{2}(U_{i,j+1,k} - U_{i,j,k}), \\
U_{i,j+\frac{1}{2},k} &= \frac{1}{2}(U_{i,j,k} + U_{i,j+1,k}) - \frac{1}{2b}(W_{i,j+1,k} - W_{i,j,k}), \\
H_{i,j,k+\frac{1}{2}} &= \frac{1}{2}(H_{i,j,k} + H_{i,j,k+1}) - \frac{c}{2}(U_{i,j,k+1} - U_{i,j,k}), \\
U_{i,j,k+\frac{1}{2}} &= \frac{1}{2}(U_{i,j,k} + U_{i,j,k+1}) - \frac{1}{2c}(H_{i,j,k+1} - H_{i,j,k}).
\end{aligned} \tag{4.59}$$

Similarly, applying a first-order upwind scheme to the characteristic variables  $p \pm aq$ ,  $p \pm bh$  and  $p \pm cr$ , respectively, we have

$$\begin{aligned}
(p + aq)_{i+\frac{1}{2},j,k} &= (p + aq)_{i,j,k}, & (p - aq)_{i+\frac{1}{2},j,k} &= (p - aq)_{i+1,j,k}, \\
(p + bh)_{i,j+\frac{1}{2},k} &= (p + bh)_{i,j,k}, & (p - bh)_{i,j+\frac{1}{2},k} &= (p - bh)_{i,j+1,k}, \\
(p + cr)_{i,j,k+\frac{1}{2}} &= (p + cr)_{i,j,k}, & (p - cr)_{i,j,k+\frac{1}{2}} &= (p - cr)_{i,j,k+1}.
\end{aligned} \tag{4.60}$$

Solving (4.60) and  $p$ ,  $q$ ,  $h$  and  $r$  should be replaced by  $-p$ ,  $-q$ ,  $-h$  and  $-r$ , respectively. We get a first-order reconstruction of the backward problem (4.51) as

$$\begin{aligned}
p_{i+\frac{1}{2},j,k} &= -\frac{1}{2}(p_{i,j,k} + p_{i+1,j,k}) - \frac{a}{2}(q_{i+1,j,k} - q_{i,j,k}), \\
q_{i+\frac{1}{2},j,k} &= -\frac{1}{2}(q_{i,j,k} + q_{i+1,j,k}) - \frac{1}{2a}(V_{i+1,j,k} - V_{i,j,k}), \\
p_{i,j+\frac{1}{2},k} &= -\frac{1}{2}(p_{i,j,k} + p_{i,j+1,k}) - \frac{b}{2}(h_{i,j+1,k} - h_{i,j,k}), \\
h_{i,j+\frac{1}{2},k} &= -\frac{1}{2}(h_{i,j,k} + h_{i,j+1,k}) - \frac{1}{2b}(p_{i,j+1,k} - p_{i,j,k}), \\
p_{i,j,k+\frac{1}{2}} &= -\frac{1}{2}(p_{i,j,k} + p_{i,j,k+1}) - \frac{c}{2}(r_{i,j,k+1} - r_{i,j,k}), \\
r_{i,j,k+\frac{1}{2}} &= -\frac{1}{2}(r_{i,j,k} + r_{i,j,k+1}) - \frac{1}{2c}(p_{i,j,k+1} - p_{i,j,k}).
\end{aligned} \tag{4.61}$$

Substituting (4.59) into (4.50) and using the notation (4.49), we obtain the first-order relaxation scheme to approximate the solution of the forward system (4.36). In the same way, we substitute equations (4.61) and the notation (4.49) to have a first-order relaxation scheme for the backward

system (4.42). The characteristic speeds  $a$ ,  $b$  and  $c$  in the relaxation systems (4.36) and (4.42) are locally computed at every cell as

$$a_{i+\frac{1}{2},j,k} = \max_{U \in A} |f'(U)|, \quad b_{i,j+\frac{1}{2},k} = \max_{U \in B} |g'(U)|, \quad c_{i,j,k+\frac{1}{2}} = \max_{U \in C} |s'(U)|, \quad (4.62)$$

where  $A = (U_{i,j,k}, U_{i+1,j,k})$ ,  $B = (U_{i,j,k}, U_{i,j+1,k})$ , and  $C = (U_{i,j,k}, U_{i,j,k+1})$ .

Moreover, for the higher-order relaxation schemes, an interpolating polynomial used in the MUSCL-type formulation that has been presented by Banda and Seaid in [82] can be used to obtain a higher-order reconstruction of the numerical fluxes.

Finally, to solve the optimal control problem (4.1) subject to 3D equations (4.36), we present an algorithm that describes the optimisation procedure based on relaxation approximation in the subsequence section.

## 4.6 Numerical algorithm

In this section, we present the iterative optimisation algorithm that solves the optimal control problem (4.1) numerically. The implementation is performed based on the numerical schemes proposed in the previous sections. A similar algorithm has been used in literature; see [32, 69] for two-dimensional and [129] for one-dimensional problems. It should be noted that the focus is on the proposed numerical method for solving the problems (4.36) and (4.42) and the solution  $U(t, x, y, z)$  does not have to be stored during the iterations by using the developed method. Thus, we do not need to approximate the generalised tangent vectors as presented in the previous Chapter to improve the gradient descent method. The procedure of the optimisation algorithm can be stated as follows: Given an initial datum  $U_0(x, y, z)$ , obtain a terminal data  $U_d(x, y, z)$  which by time  $t = T$  will either evolve into  $U(T, x, y, z) = U_d(x, y, z)$  or will be as close as possible to  $U_d$  in the  $L^2$ -norm. Therefore, we have a sequence  $\{U_0^{(N)}(x, y, z)\}$ ,  $N = 0, 1, 2, \dots$ .

1. For any given tolerance  $\alpha$ , we choose an initial guess  $U_0(x, y, z)$ .
2. We solve the system (4.36) subject to the initial equation (4.37) forward in time from  $t = 0$

to  $t = T$  using the relaxation approach to obtain solutions  $U^{(0)}(T, x, y, z)$ ,  $V^{(0)}(T, x, y, z) = f(U^{(0)}(T, x, y, z))$ ,  $W^{(0)}(T, x, y, z) = g(U^{(0)}(T, x, y, z))$ , and  $H^{(0)}(T, x, y, z) = s(U^{(0)}(T, x, y, z))$ .

3. **Iterations**, for  $N = 0, 1, 2, \dots$

We calculate the cost functional

$$\begin{aligned} J^{(N)}(\cdot) = & \frac{1}{2} \Delta x \Delta y \Delta z \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \sum_{k=1}^{N_z} \left( \left( U^{(N)}(T, x_i, y_j, z_k) - U_d(x_i, y_j, z_k) \right)^2 \right. \\ & + \left( V^{(N)}(T, x_i, y_j, z_k) - V_d(x_i, y_j, z_k) \right)^2 + \left( W^{(N)}(T, x_i, y_j, z_k) - W_d(x_i, y_j, z_k) \right)^2 \\ & \left. + \left( H^{(N)}(T, x_i, y_j, z_k) - H_d(x_i, y_j, z_k) \right)^2 \right). \end{aligned}$$

**While**  $J^{(N)}(\cdot) < \alpha$  **or**  $|J^{(N)}(\cdot) - J^{(N-1)}(\cdot)| < \alpha$

- (i) We solve the adjoint system (4.42) with the terminal data (4.43) backwards in time from  $t = T$  to  $t = 0$  using the relaxation approach to obtain solutions  $p^{(N)}(0, x, y, z)$ ,  $q^{(N)}(0, x, y, z)$ ,  $h^{(N)}(0, x, y, z)$  and  $r^{(N)}(0, x, y, z)$ .
- (ii) We update the initial controls  $U_0(x, y, z)$ ,  $V_0(x, y, z)$ ,  $W_0(x, y, z)$  and  $H_0(x, y, z)$  based on either a gradient descent or quasi-Newton method as introduced in [69].
- (iii) We solve the relaxation system (4.36) with  $U_0(x, y, z) = U_0^{(N+1)}(x, y, z)$ ,  $V_0(x, y, z) = V_0^{(N+1)}(x, y, z)$ ,  $W_0(x, y, z) = W_0^{(N+1)}(x, y, z)$  and  $H_0(x, y, z) = H_0^{(N+1)}(x, y, z)$  forward in time by a relaxation approach to obtain  $U^{(N+1)}(T, x, y, z)$ ,  $V^{(N+1)}(T, x, y, z)$ ,  $W^{(N+1)}(T, x, y, z)$  and  $H^{(N+1)}(T, x, y, z)$ .
- (iv) Then, we set  $N + 1 := N$ .

If the result converges, then stop; otherwise, repeat the procedure.

## 4.7 Numerical results

Here we present our results obtained using the algorithm introduced above to the problem related to Burger's equation with smooth and non-smooth initial data. The algorithm solves Burger's equation forward in time and the adjoint equation backward on the same space grid.



### 4.7.1 Solution of the flow equation

In this section, we discuss the numerical solutions of two-dimensional inviscid Burger's equation

$$\frac{\partial U}{\partial t} + \frac{\partial}{\partial x} \left( \frac{1}{2} U^2 \right) + \frac{\partial}{\partial y} \left( \frac{1}{2} U^2 \right) = 0, \quad (t, x, y) \in [0, T] \times [0, 1] \times [0, 1], \quad (4.63)$$

with the initial data

$$U(0, x, y) = 0.5 \cos\left(\frac{1}{2}(x+y)\right) + \sin^2(\pi x) \sin^2(\pi y), \quad (x, y) \in [0, 1] \times [0, 1]. \quad (4.64)$$

The numerical results of Burger's equation are presented in Figures 4.1 and 4.2 using Lax-Friedrichs scheme and relaxation scheme, respectively, where we also included the initial conditions (4.64).

These results are computed with final time  $T = 0.5$  on  $100 \times 100$  control volumes,  $\text{CFL} = 0.05$ , and  $\varepsilon = 10^{-6}$  for the relaxation scheme. Besides, we solved Burger's equation using the non-smooth initial data (4.65).

$$U(0, x, y) = \begin{cases} 0 & \text{if } x \geq K_1 y, \\ x - k_1 y & \text{if } -K_2 y \leq x < K_1 y, \\ \left(1 + \frac{k_1}{k_2}\right)x & \text{if } x < -K_2 y, \end{cases} \quad (4.65)$$

where  $k_1$  and  $k_2$  are positive parameters. The results are presented in Figures 4.3 and 4.4 using the Lax-Friedrichs scheme and relaxation scheme, respectively, where again, we have included the initial conditions. These results are computed with final time  $T = 0.2$  on  $50 \times 50$  control volumes,  $\text{CFL} = 0.05$ ,  $\varepsilon = 10^{-6}$  for the relaxation scheme and the constants  $k_1 = k_2 = 0.5$ . All the results are in good agreement with the proposed numerical algorithm as we expected.

### 4.7.2 Solution of the optimal control problem

Now we consider the solution of the optimal control problem. We will consider two cases and for the smooth case, we use for the solution of the Burger's equation or the adjoint equation a uniform grid of  $100 \times 100$  grid points and  $50 \times 50$  for a non-smooth case. In our first example, the desired state is obtained as the solution of Burger's equation with the initial data

$$U_d(0, x, y) = 0.7 \cos\left(\frac{1}{2}(x+y)\right) + \sin^2(\pi x) \sin^2(\pi y). \quad (4.66)$$

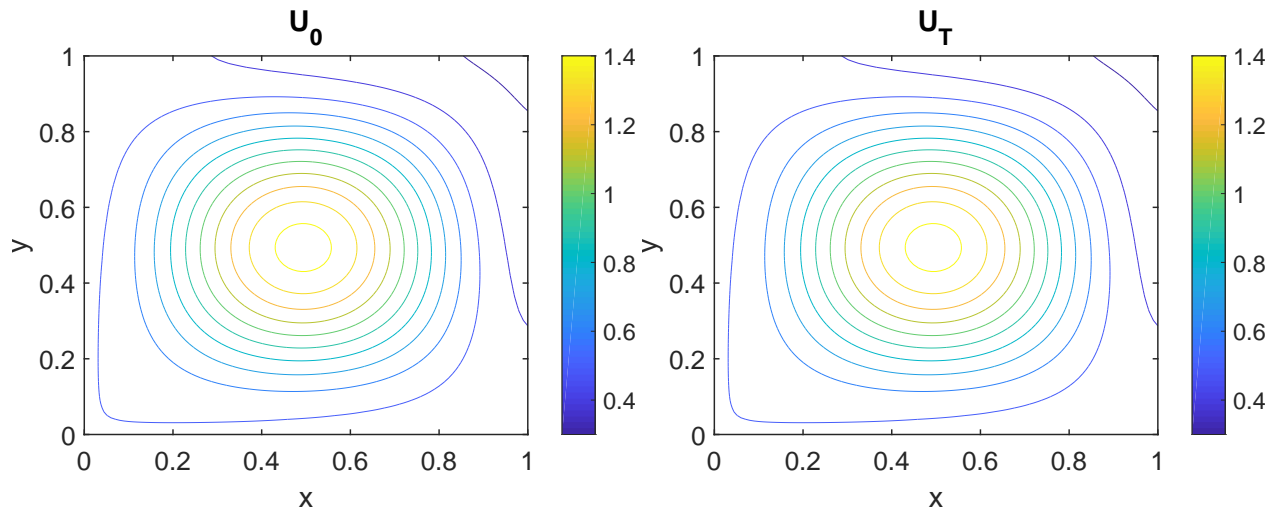


Figure 4.1: Solution of the Cauchy problem for Burger's equation (4.63) with the initial data (4.64) on the rectangle  $[0, 1] \times [0, 1]$  with grid points  $100 \times 100$ , the results obtained with the finite volume scheme (Right) and the initial data (Left).

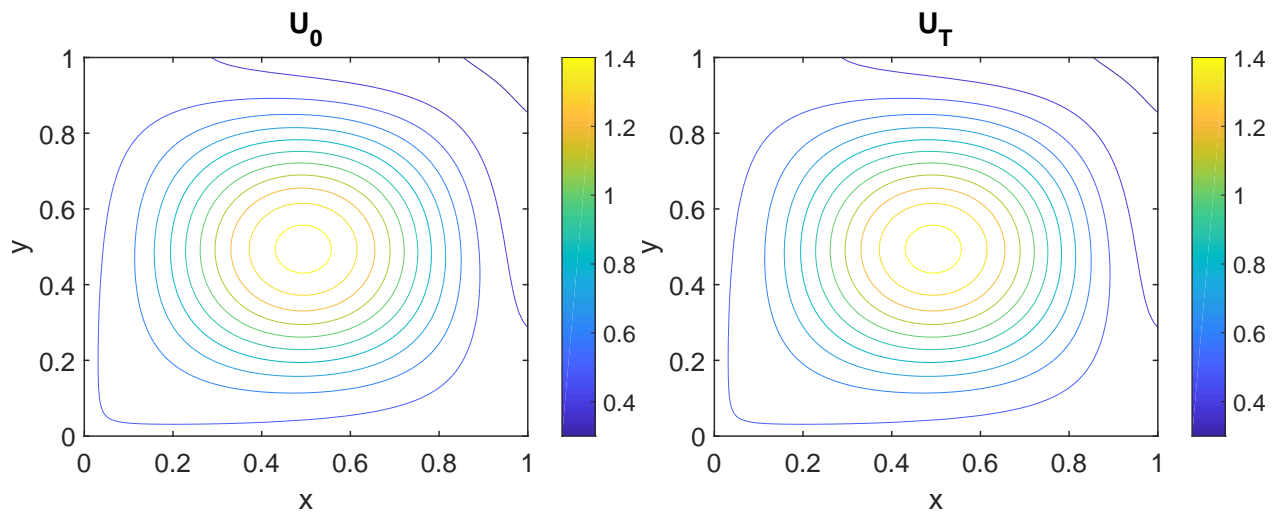


Figure 4.2: Solution of the Cauchy problem for Burger's equation (4.63) with the initial data (4.64) on the rectangle  $[0, 1] \times [0, 1]$  with grid points  $100 \times 100$ , the results obtained with the relaxation scheme (Right) and the initial data (Left).

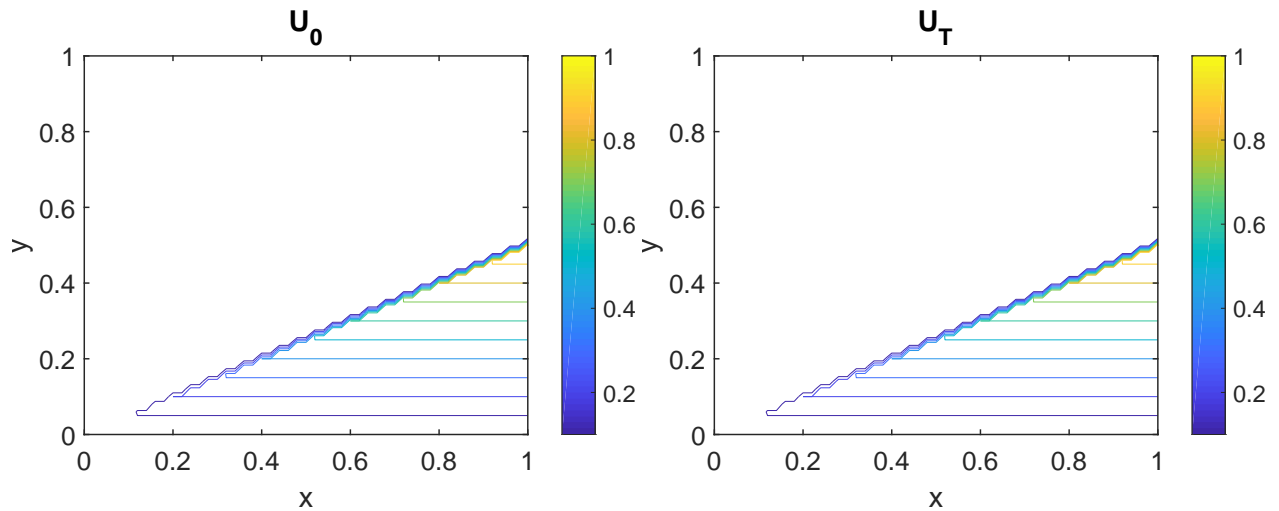


Figure 4.3: Solution of the Riemann problem for Burger's equation (4.63) with the initial data (4.65) on the rectangle  $[0, 1] \times [0, 1]$  with grid points  $50 \times 50$ , the results obtained with the finite volume scheme (Right) and the initial data (Left).

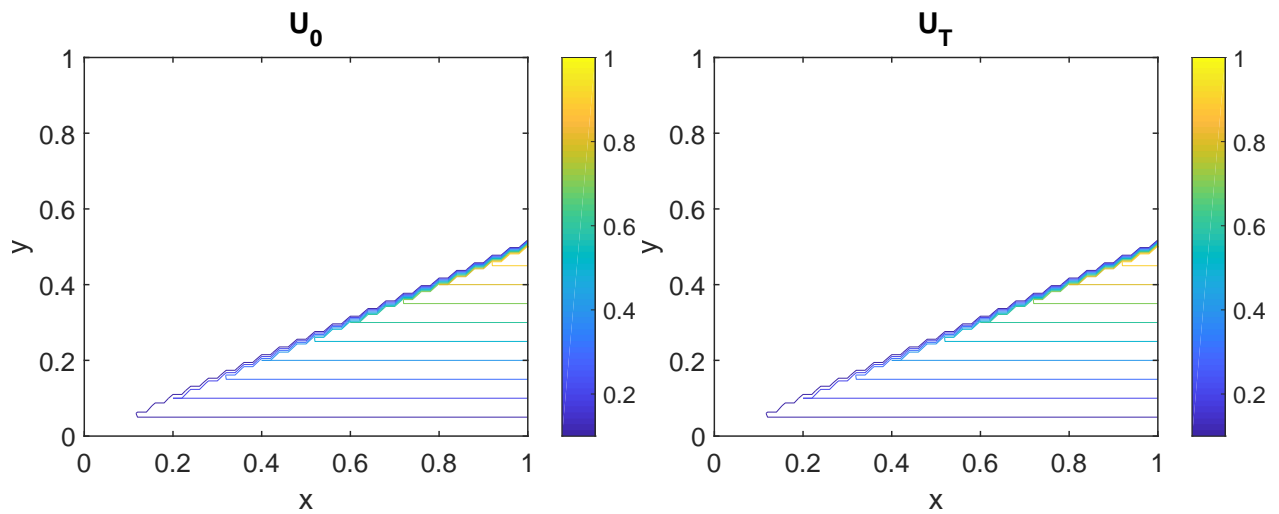


Figure 4.4: Solution of the Riemann problem for Burger's equation (4.63) with the initial data (4.65) on the rectangle  $[0, 1] \times [0, 1]$  with grid points  $50 \times 50$ , the results obtained with the relaxation scheme (Right) and the initial data (Left).

The initial control, which is the initial data for Burger's equation, is given as in equation (4.64). The aim is to drive the solution computed at time  $T$  to the desired state with our optimal control algorithm. We present in Figure 4.5 the numerical results of the optimal solution, desired solution, the gradient of reduced cost functional and the initial for the desired, using the Lax-Friedrichs's scheme. The results are computed at time  $T = 0.5$  and convergence is achieved after 100 iterations. Also, Figure 4.6 shows the results with the relaxation scheme. Our algorithm, which is the steepest descent method successfully drives the initial state to the desired state as can be seen in the graph of the cost functional against the number of iteration in Figure 4.7 related with the Lax-Friedrichs scheme and Figure 4.8 with the relaxation scheme. Now we consider a Riemann case, where the desired state is obtained as the solution at time  $T$  of the initial value problem for the Burger's equation computed using both the Lax-Friedrichs method and the relaxation scheme with the initial data

$$U_d(0, x, y) = \begin{cases} 0.2 & \text{if } x \geq K_1 y, \\ x - k_1 y + 0.2 & \text{if } -K_2 y \leq x < K_1 y, \\ (1.2 + \frac{k_1}{k_2})x & \text{if } x < -K_2 y. \end{cases} \quad (4.67)$$

The initial control is taken as in equation (4.65). We report in Figure 4.9 the numerical results of the optimal solution, desired solution, the gradient of reduced cost functional the initial for the desired, using the Lax-Friedrichs method. The results are computed at time  $T = 0.2$  and convergence is achieved after 50 iterations. Moreover, Figure 4.10 displays the results obtained with the relaxation method. In this case, the convergence of the algorithm with a tolerance of  $\epsilon = 10^{-3}$  occurs after 50 iterations as can be seen in Figure 4.11 based on the Lax-Friedrichs method, and Figure 4.12 on the relaxation method. In the Riemann problem, we used a fixed perturbation in the initial control function with the CFL = 0.05. The optimal solutions and the desired solutions appear to be similar, confirming the convergence of the proposed numerical procedure presented in both approaches. However, the reported solutions are free of spurious oscillations and the shocks are well resolved by the proposed approach without the nonlinear computational techniques. These results demonstrate the efficiency achieved by the proposed algorithm for solving optimal control problems for the two-dimensional Burger's equation. The performance of both approaches is attractive since the

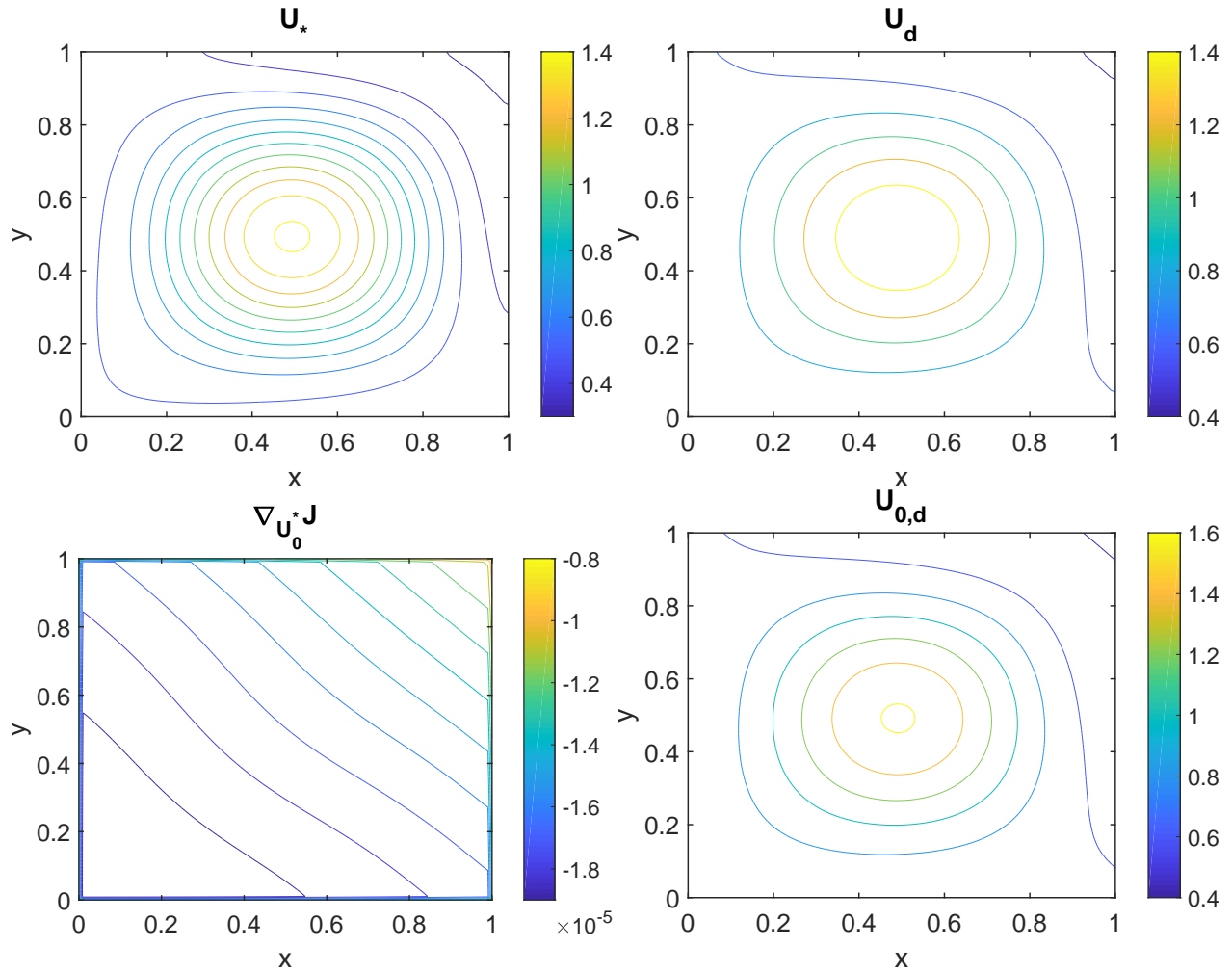


Figure 4.5: Results of the optimal solution (Top-Left), the desired solution (Top-Right), the gradient (Bottom-Left) and the initial for desired (Bottom-Right) of the optimal control problem (4.1) for the Cauchy case. The results are obtained using the finite volume scheme with grid points  $100 \times 100$  and time  $T = 0.5$ .

computed solutions remain stable and accurate even when various perturbations as in the Cauchy problem with the CFL = 0.5. Also, it is clear that some shocks are formed in the optimal solutions even we used initial smooth data.

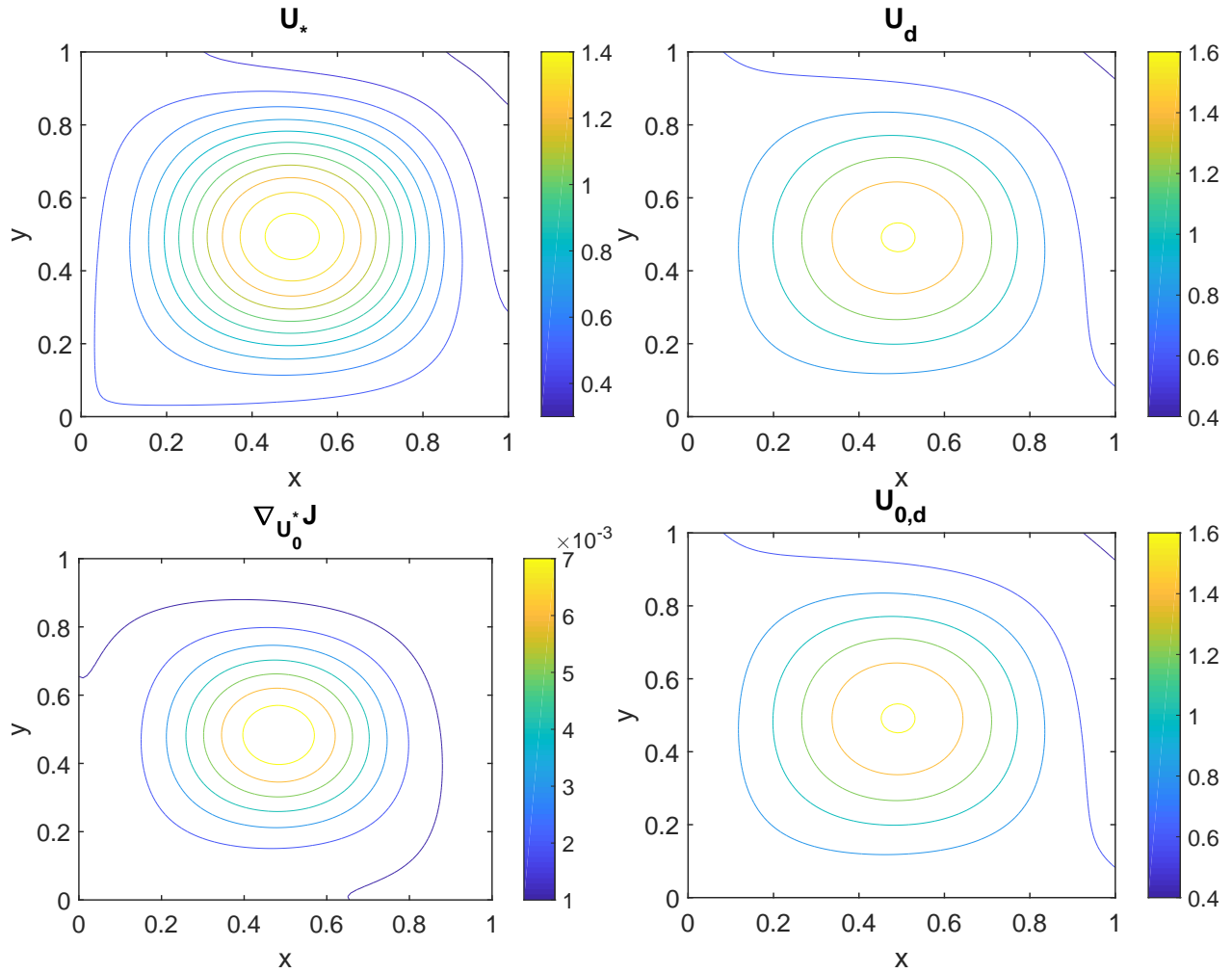


Figure 4.6: Results of the optimal solution (Top-Left), the desired solution (Top-Right), the gradient (Bottom-Left) and the initial for desired (Bottom-Right) of the optimal control problem (4.1) for the Cauchy case. The results are obtained using the relaxation scheme with grid points  $100 \times 100$  and time  $T = 0.5$ .

## 4.8 Concluding remarks

In this chapter, we successfully extended the optimal control problems to involve multi-dimensional systems of conservation laws. We used an adjoint-based approach to derive optimality conditions

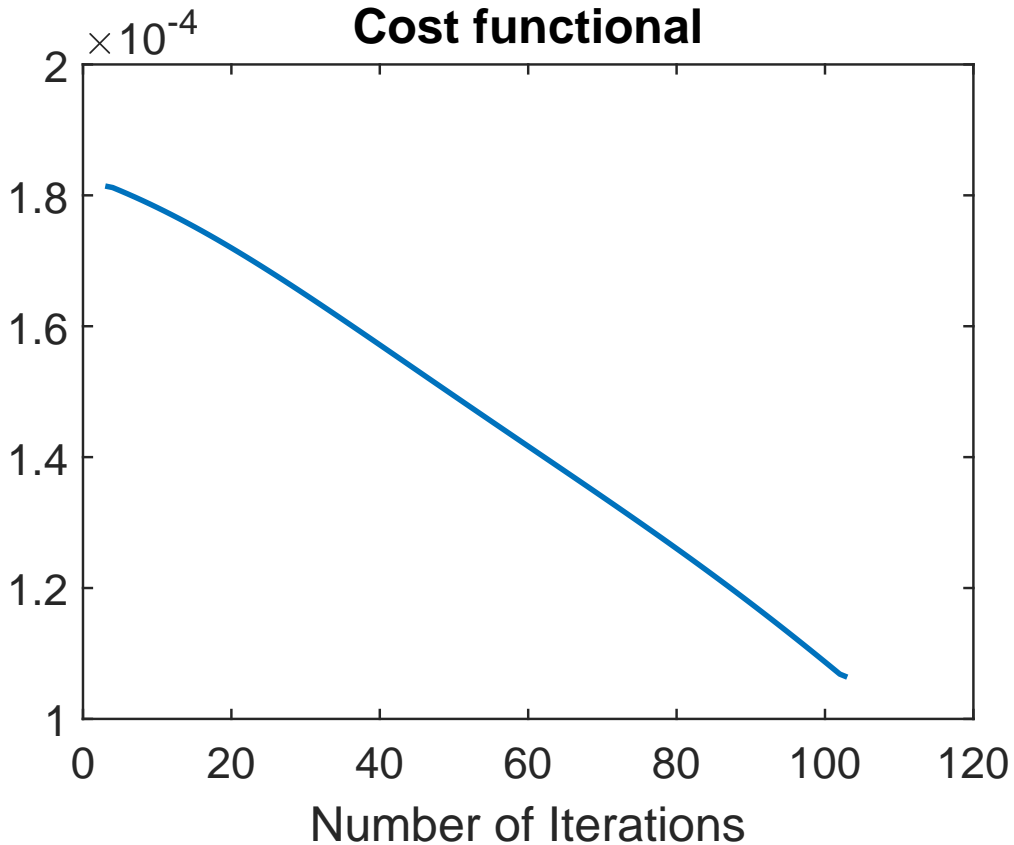


Figure 4.7: Convergence history for the solution of the optimal control problem based on the finite volume method with the tolerance  $\varepsilon = 10^{-3}$  and the initial control (4.64).

based on both multi-dimensional systems of conservation laws and their relaxation approximations. We applied these schemes to construct an optimisation algorithm that solves our optimal control problem numerically. Therefore, we presented numerical results of optimal control problems governed by the two-dimensional inviscid Burgers equation using both finite volume method and relaxation approximation. It can be observed that the numerical simulations successfully examined our algorithm and the results are in good agreement using either finite volume method or relaxation approximation. Besides, we can observe that the optimisation algorithm presented above made the computed optimal solutions very close to the desired solutions in both problems. However, the results are not the same using both methods, probably because of the parameters used in the relaxation approximation or the type of numerical flux in the finite volume method. The presented algorithm can solve not only the smooth problems but also the problems involving

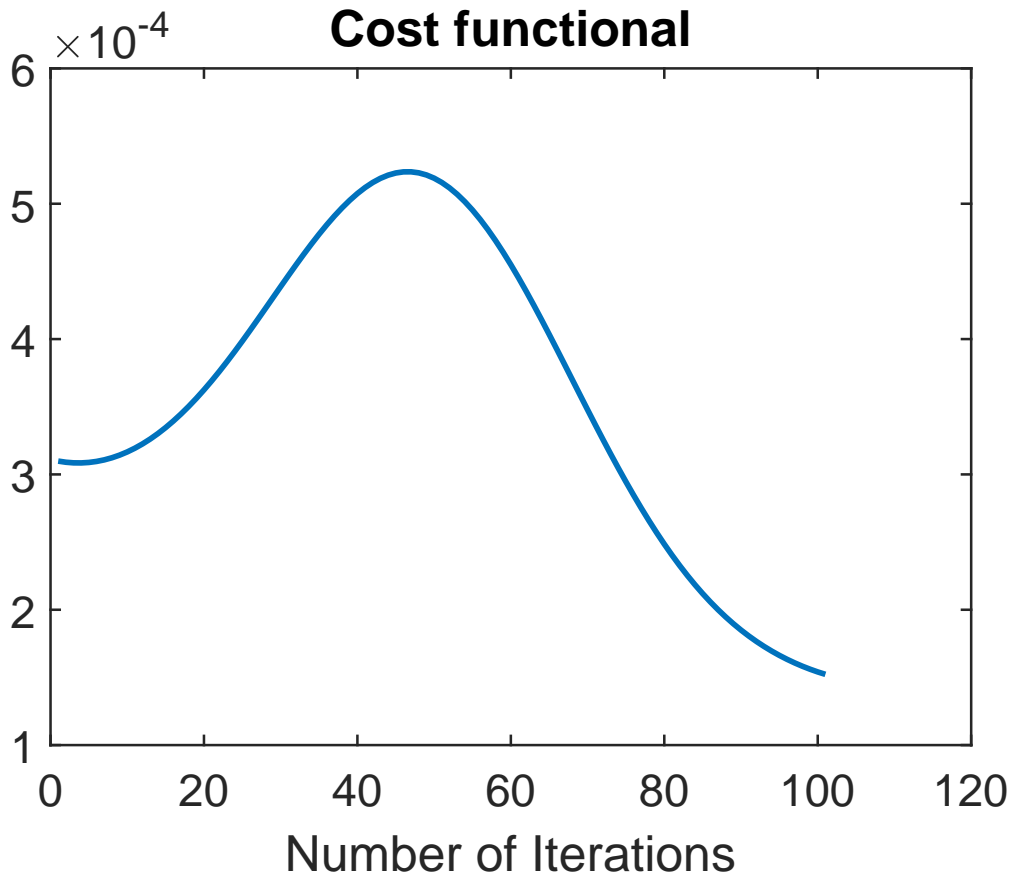


Figure 4.8: Convergence history for the solution of the optimal control problem based on the relaxation method with the tolerance  $\varepsilon = 10^{-3}$  and the initial control (4.64).

discontinuities in either the initial data for the desired state or in the desired state itself. The extension to the three-dimensional case might also be considered.

In the next chapter, the optimal control problems constrained by multi-dimensional Euler equations based on Lattice Boltzmann approximation will be considered.



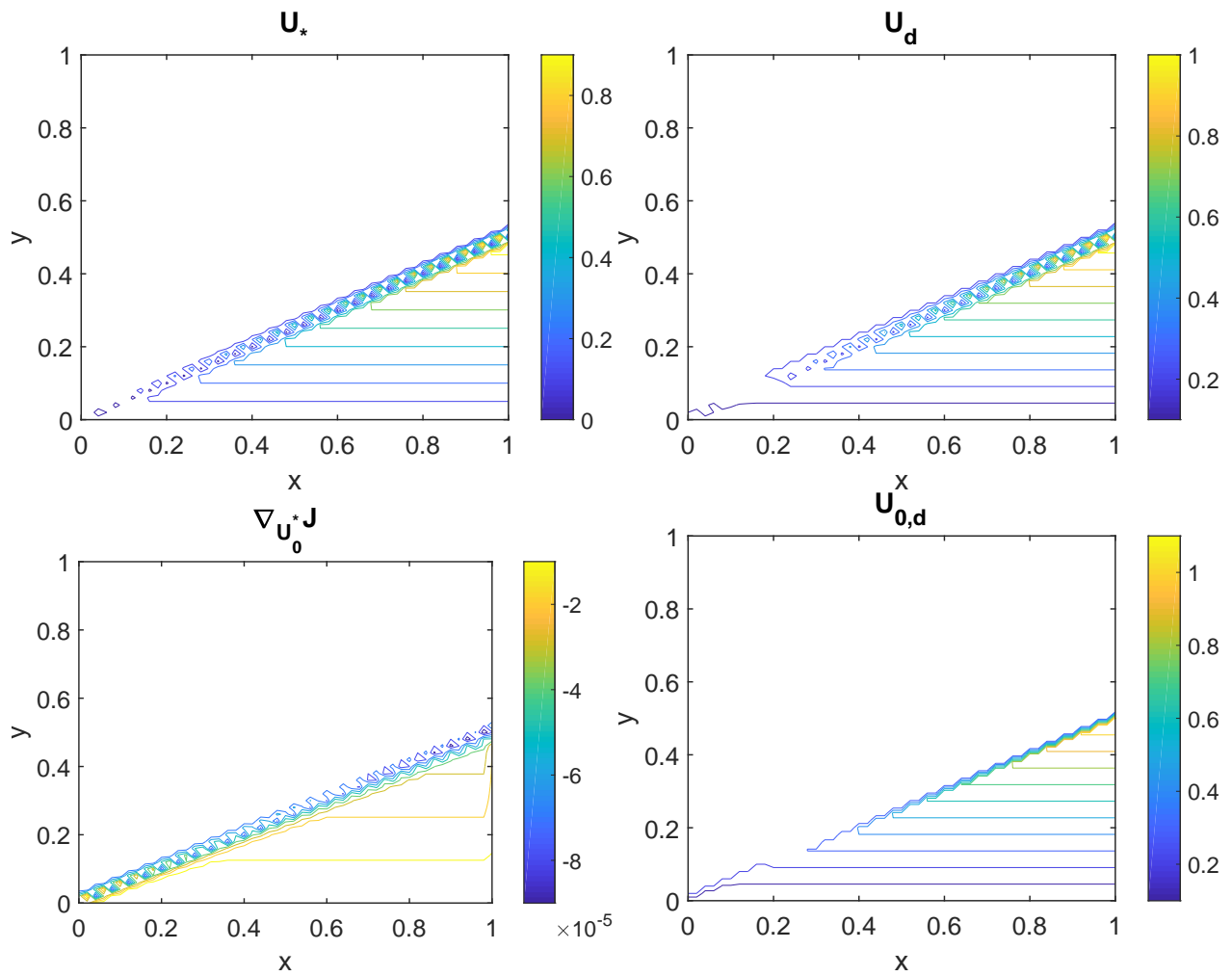


Figure 4.9: Plot of the optimal solution (Top-Left), the desired solution (Top-Right), the gradient (Bottom-Left) and the initial for desired (Bottom-Right) of the optimal control problem (4.1) for the Riemann case. The results are computed using the finite volume method with grid points  $50 \times 50$  and time  $T = 0.2$ .

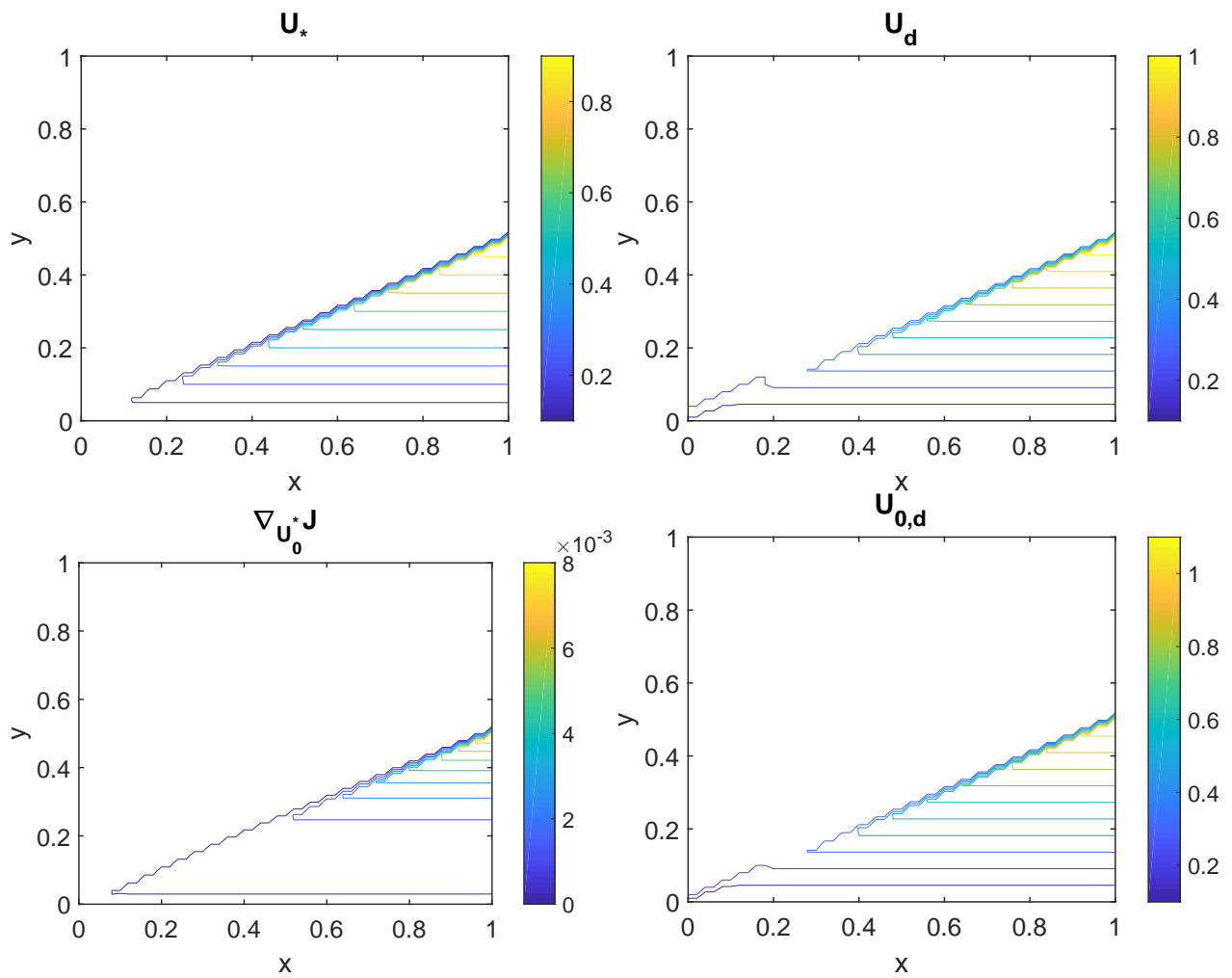


Figure 4.10: Plot of the optimal solution (Top-Left), the desired solution (Top-Right), the gradient (Bottom-Left) and the initial for desired (Bottom-Right) of the optimal control problem (4.1) for the Riemann case. The results are computed using the relaxation method with grid points  $50 \times 50$  and time  $T = 0.2$ .

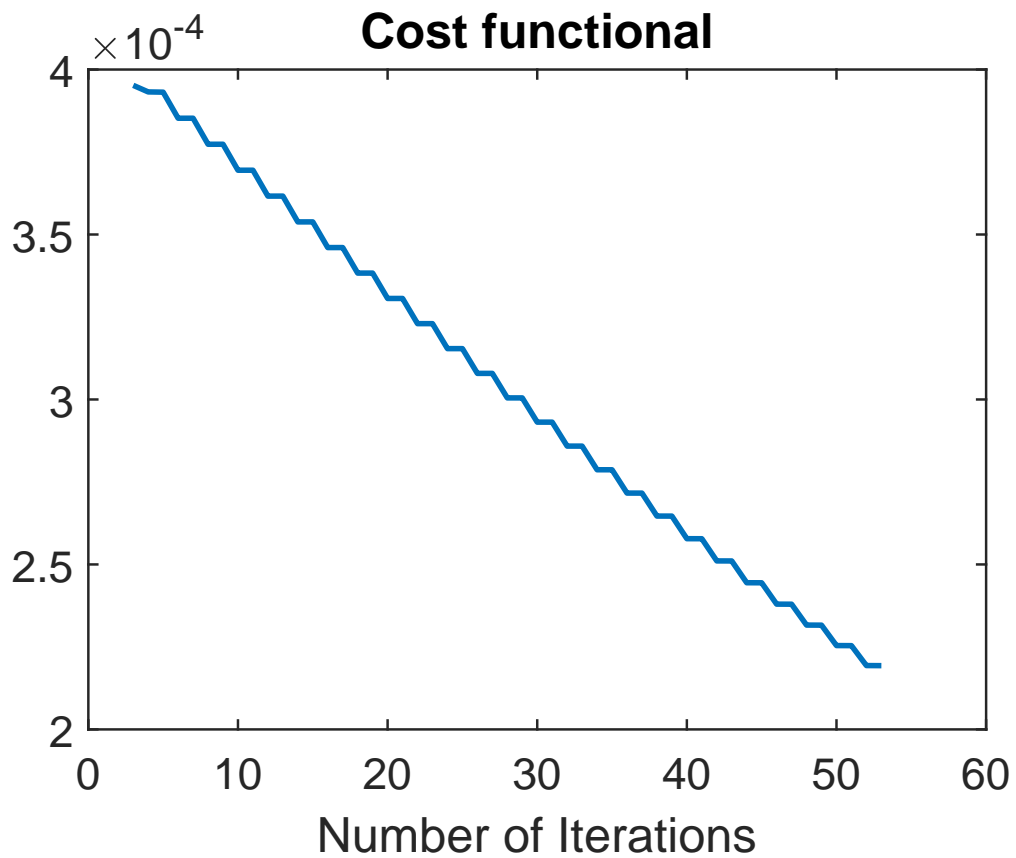


Figure 4.11: Convergence history for the solution of the optimal control problem related to the finite volume scheme with the tolerance  $\varepsilon = 10^{-3}$  and the initial control (4.65).

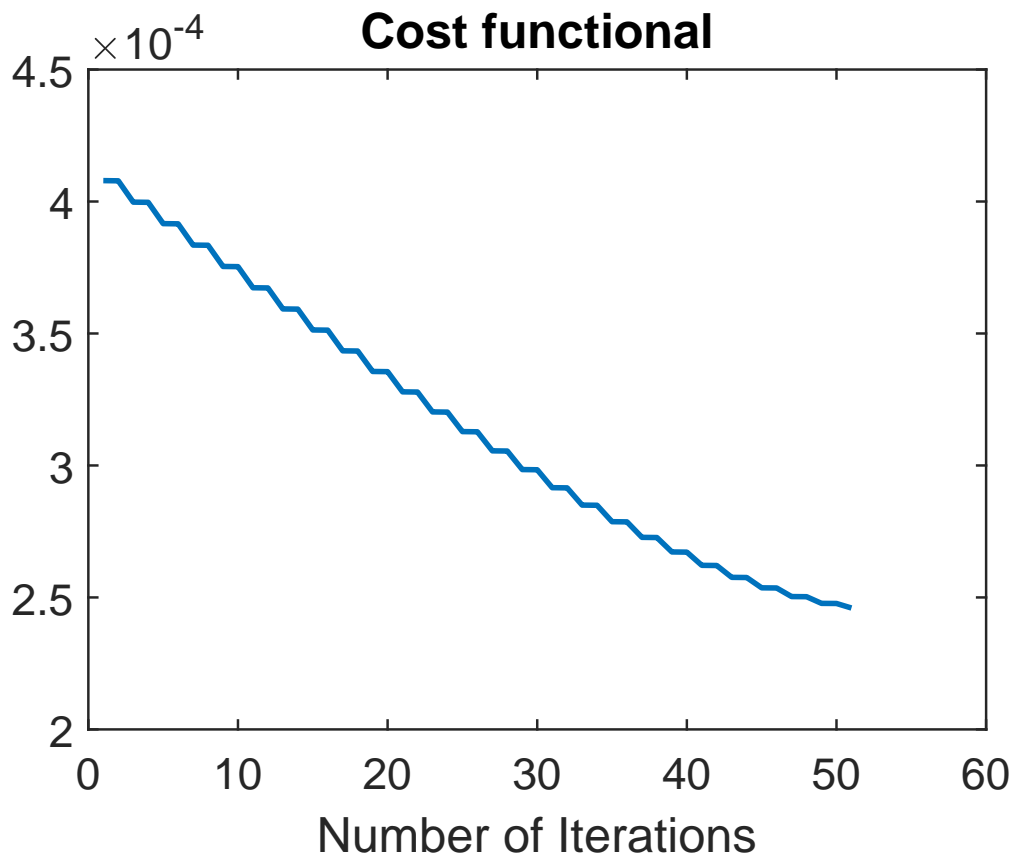


Figure 4.12: Convergence history for the solution of the optimal control problem related to the relaxation scheme with the tolerance  $\varepsilon = 10^{-3}$  and the initial control (4.65).

# Chapter 5

## Optimal control problem governed by the multi-dimensional system of Euler equations

This chapter deals with a flow matching optimal control problem constrained by the multi-dimensional Euler equations. The control variable is the initial condition of the flow equations. In the optimise-then-discretise framework, we consider a lattice Boltzmann approximation of the Euler equations and derive the optimality conditions at the kinetic level. This is important as the original nonlinear equations exhibit discontinuities such as shock or contact discontinuities that pose problems to the adjoint solver. Our analysis focuses on the two-dimensional nine velocities (D2Q9) lattice Boltzmann approximation of the Euler equations and we present some numerical results in the two-dimensional cases.

### 5.1 Introduction

Optimal control problems governed by partial differential equations have attracted a lot of attention recently [31–34]. Amongst the many applications considered in the literature, we mention those related to shape optimisation that is involved in the design of aero-dynamical objects such as cars and planes. In this chapter, we propose a numerical solution to a flow matching optimal control problem constrained by the multi-dimensional Euler equations. Precisely we consider the problem

$$\text{Minimise}_{U^0} J(U(T, \cdot), U^0; U_d) = \frac{1}{2} \int_{\mathbb{R}^d} \|U(T, \mathbf{x}) - U_d(\mathbf{x})\|^2 d\mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^d, \quad (5.1)$$

where the function  $U = (\rho, \rho u_\alpha, E)$ , with  $E = \rho(bRT + u_\alpha^2)$  solves the multi-dimensional Euler equations of the form [10]

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_\alpha}(\rho u_\alpha) &= 0, \\ \frac{\partial}{\partial t}(\rho u_\alpha) + \frac{\partial}{\partial x_\beta}(\rho u_\alpha u_\beta) + \frac{\partial P}{\partial x_\alpha} &= 0, \\ \frac{\partial}{\partial t}[\rho(bRT + u_\alpha^2)] + \frac{\partial}{\partial x_\beta}[\rho u_\alpha(bT + u_\beta^2) + 2Pu_\alpha] &= 0, \end{aligned} \quad (5.2)$$

where  $t$  is the time,  $x_\alpha$  is the spatial coordinate, with  $\mathbf{x} = (x_\alpha)_{\alpha=1}^d$ ;  $\rho$ ,  $u_\alpha$ ,  $T$  and  $p = \rho RT$  are respectively the density, the flow velocity in the  $x_\alpha$  direction, the temperature and the pressure of the gas.  $R$  is the specific gas constant and  $b = \frac{2}{\gamma-1}$  is a given constant with  $\gamma$  being the specific heat ratio. Note that the subscripts  $\alpha$  and  $\beta$  represent the spatial coordinates and the Einstein summation convention is applied. The initials conditions, which are the control variables in our optimal control problem are given by

$$\rho = \rho^0, \quad u_\alpha = u_\alpha^0, \quad T = T^0 \quad \text{at} \quad t = 0, \quad (5.3)$$

where  $\rho^0$ ,  $u_\alpha^0$  and  $T^0$  are given functions of the space variable  $\mathbf{x}$ . In (5.1),  $U_d$  is the desired state that has to be achieved approximately at the final time  $T$ .

Our approach for the solution of problem (5.1) consists of replacing in the Lagrangian formulation leading to the optimality conditions, the nonlinear equations (5.2) with a lattice Boltzmann approximation due to Kataoka and Tsutahara [10]. This is an extension to the multi-dimensional case of the work done in [70].

It is well-known that classical solutions to the Cauchy problem associated with multi-dimensional conservation laws develop discontinuities in finite time even if the initial conditions are smooth. Those discontinuities may be classified as vorticity waves, focusing waves, complicated wave interactions and concentration waves. Generally, we then seek weak solutions as introduced in [117]. In the realm of weak solutions, discontinuities are accepted provided that at their front, the so-called Rankine-Hugoniot condition is satisfied. The discontinuous solutions are not unique and the choice of the physically relevant solution that leads to some unique results is done via some

entropy conditions. Then, we talk about the uniqueness of the weak entropy solution of conservation laws [13, 23, 26]. The numerical solution of a system of conservation laws has attracted a lot of interest in the literature. Different methods have been proposed among others by Kurganov et al [47], Wang et al [48] and Gottlieb et al. [49]. See also [32, 51]. For the solution of our optimal control problem (5.1), we propose in the optimise-then-discretise framework [1–3] a set of optimality conditions that leads to a numerical algorithm. The method involves, as in other related research literature, the solution of the flow equations forward in time, the solution of the adjoint equation backward in time and an update of the control using the gradient of the reduced cost functional. The discontinuities in the solution of the flow equations as mentioned above poses a serious problem to the backward solver. We propose to replace the flow equation given by (5.2) with a lattice Boltzman (LB) approximation, which is a linear conservation law with a stiff source term. In general, the lattice Boltzmann (LB) method solves the kinetic equation of the discrete-molecular-velocity type such that the macroscopic variable satisfies the fluid dynamics type equations. In the lattice Boltzmann formulation of a system of conservation laws, the nonlinearity of the macroscopic equation is captured by the so-called collision operator that appears as a source term in the LB equation. The solution of optimal control problems in computational fluid dynamics using the lattice Boltzmann equations has been done before by Nørgaard et al [64] for the solution of a shape and topology optimisation problem. Li et al [68] considered a problem of Airfoil design optimisation and combined the LB methods and the adjoint method. Morales-Hernández and Zuazua [35] considered a computational method for the two-dimensional inverse design of linear transport equations on unstructured grids. They suggested the use of lower-order methods for the solution of the adjoint equation because the use of higher-order schemes leads to spurious high-frequency numerical components that slow down the convergence process. It is important to note that in the numerical algorithm for the solution of optimal control problems related to conservation laws, the flow equations and the adjoint equations have to be solved on the same grid. In our derivation of the optimality conditions, we focus on the two-dimensional nine velocities (D2Q9) lattice Boltzmann model as proposed by [10]. We obtain an efficient method that performs well for many problems of interest.

The rest of this chapter is organised as follows. In Section 5.2, the formulation of the optimal control problem is proposed. In Section 5.3, we present the multi-dimensional lattice Boltzmann approximation of the Euler equations. In Section 5.4, we derive the optimality conditions for the optimal control at the kinetic level where we replace the original constraints by the approximating lattice Boltzmann equations. Some numerical results are presented in Section 5.5 and we present some concluding remarks and an outlook in Section 5.6.

## 5.2 Problem formulation

We consider the optimal control problem

$$\text{Minimise}_{U^0} J(U(T, \cdot), U^0; U_d) = \frac{1}{2} \int_{\mathbb{R}^d} \|U(T, \mathbf{x}) - U_d(\mathbf{x})\|^2 d\mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^d, \quad (5.4)$$

constrained by the multi-dimensional Euler equations (5.2). The initial conditions  $U^0 = (\rho^0, u_\alpha^0, E^0)$ , with  $E^0 = \rho^0(bRT^0 + (u_\alpha^0)^2)$  play the role of the control variable. Problem (5.4) can be seen as an inverse problem. It is often referred to in the literature as an inverse design problem. In the one-dimensional case, the constraints equations take the form

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho u) &= 0, \\ \frac{\partial}{\partial t}(\rho u) + \frac{\partial}{\partial x}(\rho u^2 + P) &= 0, \\ \frac{\partial}{\partial t}[\rho(b\theta + u^2)] + \frac{\partial}{\partial x}[\rho u(b\theta + u^2) + 2Pu] &= 0. \end{aligned} \quad (5.5)$$

In this case, the solution of the problem has been extensively solved in [70]. In the two-dimensional case, the flow equations amount to

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho u) + \frac{\partial}{\partial y}(\rho v) &= 0, \\ \frac{\partial}{\partial t}(\rho u) + \frac{\partial}{\partial x}(\rho u^2 + P) + \frac{\partial}{\partial y}(\rho uv) &= 0, \\ \frac{\partial}{\partial t}(\rho v) + \frac{\partial}{\partial x}(\rho vu) + \frac{\partial}{\partial y}(\rho v^2 + P) &= 0, \\ \frac{\partial}{\partial t}[\rho(b\theta + u^2 + v^2)] + \frac{\partial}{\partial x}[\rho u(b\theta + u^2 + v^2) + 2Pu] + \frac{\partial}{\partial y}[\rho v(b\theta + u^2 + v^2) + 2Pv] &= 0. \end{aligned} \quad (5.6)$$



where we have denoted  $x_1 = x$ ,  $x_2 = y$ ,  $u_1 = u$ ,  $u_2 = v$ ,  $\theta = RT$ ,  $\partial_t = \frac{\partial}{\partial t}$  etc. The flow equations in all these cases can be written as a multi-dimensional conservation law of the form

$$\frac{\partial U}{\partial t} + \nabla \cdot \mathbf{f}(U) = 0, \quad (5.7)$$

where the flux function  $\mathbf{f} = (f_1, f_2, \dots, f_l)$  maps  $\mathbb{R}^k$  where  $k$  is the number of conservative variables to  $\mathbb{R}^l$  where  $l$  is the number of equations,  $\nabla = (\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \dots, \frac{\partial}{\partial x_d})$  is the derivative operator. In the two-dimensional case, which we will use for the illustration of our results, the system of equations (5.6) can be written in the form (5.7) with

$$U = \begin{bmatrix} \rho \\ m_1 \\ m_2 \\ E \end{bmatrix}, \quad f_1(U) = \begin{bmatrix} \frac{m_1}{b} + \frac{m_1^2(b-1) - m_2^2}{\rho b} \\ \frac{m_2 m_1}{\rho} \\ \frac{m_1 E}{\rho} + \frac{2m_1(\rho E - m_1^2 - m_2^2)}{\rho^2 b} \end{bmatrix}, \quad f_2(U) = \begin{bmatrix} \frac{m_2}{b} + \frac{m_2^2(b-1) - m_1^2}{\rho b} \\ \frac{m_1 m_2}{\rho} \\ \frac{m_2 E}{\rho} + \frac{2m_2(\rho E - m_1^2 - m_2^2)}{\rho^2 b} \end{bmatrix}, \quad (5.8)$$

where,  $\rho$ ,  $m_1 = \rho u$ ,  $m_2 = \rho v$ , and  $E = \rho(b\theta + u^2 + v^2)$  are the density, the momentum in the  $x$ -coordinate, the momentum in the  $y$ -coordinate and the total energy per unit volume, respectively. With the cost functional  $J(U(T, \cdot), U^0; U_d)$  that can be written as

$$J(U(T, \cdot), U^0; U_d) = \frac{1}{2} \int_{\mathbb{R}} \int_{\mathbb{R}} \left[ (\rho(T, x, y) - \rho_d(x, y))^2 + (m_1(T, x, y) - m_{1,d}(x, y))^2 + (m_2(T, x, y) - m_{2,d}(x, y))^2 + (E(T, x, y) - E_d(x, y))^2 \right] dx dy. \quad (5.9)$$

The existence and uniqueness of solutions of the optimal control problem (5.4) rely on the existence of solutions of (5.7) and the existence of solutions of constrained optimisation problems (see Chapter 4 for more details).

### 5.3 A lattice Boltzmann approximation of the Euler equations

Let  $\xi_{i\alpha}$  ( $i = 1, \dots, N-1$ ) be the molecular velocity in the  $x_\alpha$  direction of the  $i$ th particle, with  $N$  the total number of molecular velocities. We introduce the variable  $\eta_i$  to control the specific heat ratio and denote by  $f_i(t, x_\alpha)$  the velocity distribution function of the  $i$ th particle. The macroscopic

variable  $\rho$ ,  $u_\alpha$  and  $T$  are defined as

$$\rho = \sum_{i=0}^{N-1} f_i, \quad \rho u_\alpha = \sum_{i=0}^{N-1} \xi_{i\alpha} f_i, \quad \text{and} \quad E = \rho (bRT + u_\alpha^2) = \sum_{i=0}^{N-1} (\xi_{i\alpha}^2 + \eta_i^2) f_i. \quad (5.10)$$

Now we consider the initial value problem for the kinetic equation in non-dimensional form

$$\frac{\partial f_i}{\partial t} + \xi_{i\alpha} \frac{\partial f_i}{\partial x_\alpha} = \Omega(f_i), \quad i = 0, 1, \dots, N-1, \quad (5.11)$$

where the collision operator  $\Omega(f_i)$  is of the Bhatnager-Gross-Krook (BGK)-type

$$\Omega(f_i) = \frac{f_i^{eq}(\rho, u_\alpha, E) - f_i}{\varepsilon}. \quad (5.12)$$

With the initial conditions

$$f_i(0, x_\alpha) = f_i^{eq}(\rho^0, u_\alpha^0, T^0)(x_\alpha), \quad \text{at} \quad t = 0. \quad (5.13)$$

Therein  $\varepsilon$  is a given constant called the relaxation time and the local equilibrium distribution function  $f_i^{eq}(\rho, u_\alpha, T)$  is a given function of the macroscopic variables. We can integrate the Lattice Boltzmann model (5.11) along characteristics to obtain the often used model

$$\frac{f_i(t + \Delta t, x_\alpha + \xi_{i\alpha} \Delta t) - f_i(t, x_\alpha)}{\Delta t} = \frac{f_i^{eq}(\rho, u_\alpha, T) - f_i}{\varepsilon}, \quad (5.14)$$

where  $\Delta t$  is the discrete-time step of order  $\varepsilon$ . In general, (5.14) is viewed as a two step process made of a collision step

$$\tilde{f}_i(t, x_\alpha) = f_i(t, x_\alpha) + \Delta t \left[ \frac{f_i^{eq}(\rho, u_\alpha, T) - f_i}{\varepsilon} \right], \quad (5.15)$$

and a propagation step

$$f_i(t + \Delta t, x_\alpha + \xi_{i\alpha} \Delta t) = \tilde{f}_i(t, x_\alpha), \quad (5.16)$$

where  $\tilde{f}_i$  is the post-collision distribution function.

The form (5.14) is only one finite difference discretisation of the lattice Boltzmann model (5.11). Therefore, for the sake of generality and the purpose of deriving an adjoint calculus for our optimal control problem, we consider in the sequel the general form (5.11). To recover from the lattice

Boltzmann equation, the Euler equation at the hydrodynamic limits, the following constraints are imposed on the moments of  $f_i^{eq}$ , the equilibrium distribution appearing in the collision operator (5.12):

$$\begin{aligned}
\sum_{i=0}^{N-1} f_i^{eq} &= \rho, \\
\sum_{i=0}^{N-1} \xi_{i\alpha} f_i^{eq} &= \rho u_\alpha, \\
\sum_{i=0}^{N-1} \xi_{i\alpha} \xi_{i\beta} f_i^{eq} &= P \delta_{\alpha\beta} + \rho u_\alpha u_\beta, \\
\sum_{i=0}^{N-1} (\xi_{i\alpha}^2 + \eta_i^2) f_i^{eq} &= \rho (bRT + u_\alpha^2), \\
\sum_{i=0}^{N-1} (\xi_{i\beta}^2 + \eta_i^2) \xi_{i\alpha} f_i^{eq} &= \rho [(b+2)RT + u_\beta^2] u_\alpha.
\end{aligned} \tag{5.17}$$

We will illustrate our results on the two-dimensional nine velocities (D2Q9) lattice Boltzmann model as illustrated in Figure 5.1. The initial value problem (5.11) - (5.12) can be written in two-dimensional cases as

$$\frac{\partial f_i}{\partial t} + \xi_{i1} \frac{\partial f_i}{\partial x} + \xi_{i2} \frac{\partial f_i}{\partial y} = \Omega(f_i), \quad i \in \{0, 1, \dots, N-1\}, \tag{5.18}$$

where  $\xi_{i1}$  and  $\xi_{i2}$  are the molecular velocities in the x-direction and y- direction, respectively and  $\Omega(f_i)$  is given as [130]

$$\Omega_i(f) = \frac{f_i^{eq}(\rho, m_1, m_2, E) - f_i}{\varepsilon}, \tag{5.19}$$

with the initial conditions

$$f_i(0, x, y) = f_i^{eq}(\rho^0, m_1^0, m_2^0, E^0)(x, y), \quad \text{at } t = 0. \tag{5.20}$$

Therein, the macroscopic variables  $\rho$ ,  $u$ ,  $v$  and  $\theta$  are defined as

$$\begin{aligned}
\rho &= \sum_{i=0}^{N-1} f_i, \quad m_1 = \sum_{i=0}^{N-1} \xi_{1,i} f_i, \quad m_2 = \sum_{i=0}^{N-1} \xi_{2,i} f_i, \quad \text{and} \\
\rho (b\theta + u^2 + v^2) &= \sum_{i=0}^{N-1} (\xi_{1,i}^2 + \xi_{2,i}^2 + \eta_i^2) f_i,
\end{aligned} \tag{5.21}$$

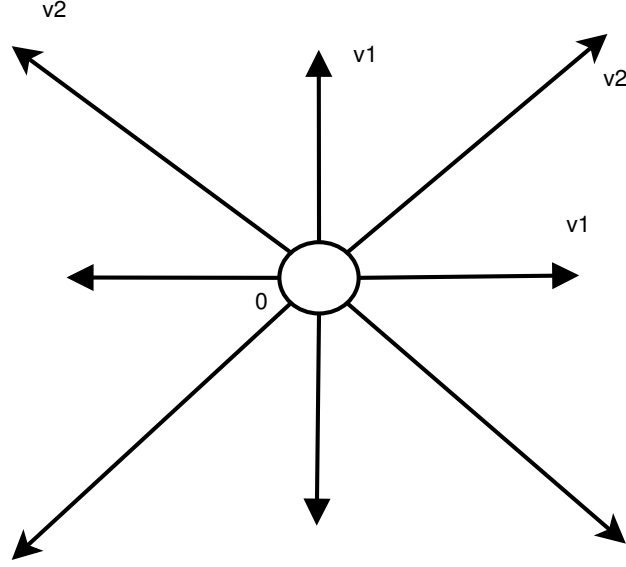


Figure 5.1: Two-dimensional, nine-velocities (D2Q9) square Lattice Boltzmann Model.

For  $v1 = c_1$  and  $v2 = c_1$ , the exact form of the discrete velocities are given as in [10] as

$$\xi_i = \begin{cases} (0,0), & \text{for } i = 0; \\ c_1 \left( \cos\left(\frac{\pi(i+1)}{2}\right), \sin\left(\frac{\pi(i+1)}{2}\right) \right), & \text{for } i = 1, 2, 3, 4; \\ c_2 \left( \cos\left(\frac{\pi(i+1)}{2} + \frac{\pi}{4}\right), \sin\left(\frac{\pi(i+1)}{2} + \frac{\pi}{4}\right) \right), & \text{for } i = 5, 6, 7, 8. \end{cases} \quad (5.22)$$

The constants  $\eta_i$  are given as

$$\eta_i = \begin{cases} \eta_0, & \text{for } i = 0; \\ 0, & \text{for } i = 1, 2, \dots, 8. \end{cases} \quad (5.23)$$

In equation (5.22),  $c_1$  and  $c_2$ , with  $c_1 \neq c_2$ , and  $\eta_0$  are given nonzero constants. The equilibrium distribution is given in the form [10]

$$f_i^{eq} = \rho A_i + (m_1 \xi_{1,i} + m_2 \xi_{2,i}) B_i + \frac{1}{\rho} \left( m_1^2 \xi_{1,i}^2 + m_2^2 \xi_{2,i}^2 \right) D_i, \quad (5.24)$$

where

$$A_i = \begin{cases} \frac{(b-2)\theta}{\eta_0^2}, & \text{for } i = 0, \\ \frac{1}{4(c_1^2 - c_2^2)} \left[ -c_2^2 + \left( (b-2)\frac{c_2^2}{\eta_0^2} + 2 \right) \theta + \frac{c_2^2}{c_1^2} (u^2 + v^2) \right], & \text{for } i = 1, 2, 3, 4, \\ \frac{1}{4(c_2^2 - c_1^2)} \left[ -c_1^2 + \left( (b-2)\frac{c_1^2}{\eta_0^2} + 2 \right) \theta + \frac{c_1^2}{c_2^2} (u^2 + v^2) \right], & \text{for } i = 5, 6, 7, 8; \end{cases} \quad (5.25)$$

$$B_i = \begin{cases} 0, & \text{for } i = 0, \\ \frac{-c_2^2 + (b+2)\theta + u^2 + v^2}{2c_1^2(c_1^2 - c_2^2)}, & \text{for } i = 1, 2, 3, 4, \\ \frac{-c_1^2 + (b+2)\theta + u^2 + v^2}{2c_2^2(c_2^2 - c_1^2)}, & \text{for } i = 5, 6, 7, 8; \end{cases} \quad (5.26)$$

and

$$D_i = \begin{cases} 0, & \text{for } i = 0, \\ \frac{1}{2c_1^4}, & \text{for } i = 1, 2, 3, 4, \\ \frac{1}{2c_2^4}, & \text{for } i = 5, 6, 7, 8. \end{cases} \quad (5.27)$$

Now we briefly discuss the asymptotic convergence of the lattice Boltzmann equation (5.11) with initial conditions (5.13) to the solution of the multi-dimensional Euler equations (5.2) when  $\varepsilon \rightarrow 0$ . We consider both the case where the deviation of  $f_i$  from that of a uniform reference state at rest is of the order of unity throughout and the case where the solution has a steep variation in the form of a shock wave or contact discontinuities. The weak solution of the Euler equations (5.2) satisfies the integral form of the equation, which can be written in the general form

$$\int_0^\infty \int_{\mathbb{R}^d} \left( \frac{\partial \psi}{\partial t}(t, \mathbf{x}) U(t, \mathbf{x}) + \nabla \psi(t, \mathbf{x}) \cdot \mathbf{f}(U) \right) d\mathbf{x} dt + \int_{\mathbb{R}^d} U^0(\mathbf{x}) \cdot \psi(0, \mathbf{x}) d\mathbf{x} = 0, \quad (5.28)$$

where  $\psi : \mathbb{R}^+ \times \mathbb{R}^d \rightarrow \mathbb{R}^k$  is a smooth test function of  $(t, \mathbf{x}) \in \mathbb{R}^+ \times \mathbb{R}^d$  which vanishes for  $t + |\mathbf{x}|$  large enough.

To obtain the weak solution of the Euler equation from the kinetic equation system (5.11), we consider the weak form of (5.11) with the initial conditions (5.13) in the form

$$\int_0^\infty \int_{\mathbb{R}^d} \left[ \left( \frac{\partial \psi}{\partial t} + \xi_{i\alpha} \frac{\partial \psi}{\partial x_\alpha} \right) f_i - \frac{f_i^{eq}(\rho, u_\alpha, T) - f_i}{\varepsilon} \psi \right] d\mathbf{x} dt + \int_{\mathbb{R}^d} f_i^{eq}(\rho^0, u_\alpha^0, T^0) \psi(0, \mathbf{x}) d\mathbf{x} = 0, \quad (5.29)$$

where  $\psi$  is a test function as above and is independent of  $\varepsilon$ . It is well-known that flows with shock waves, contact discontinuities or other singularities can be correctly described by the weak solutions of the compressible Euler equations in the form (5.28) with the subsidiary entropy condition. According to the analysis of the Boltzmann equation, shock waves and contact discontinuities are unreal discontinuities in the realm of Lattice Boltzmann simulations but thin layers of width  $O(\varepsilon^2)$  across which the variable makes an appreciable variation. The following proposition is an adaptation of the statement of such result as in [10].

**Proposition 5.3.1.** [10] *Consider the case where the solution  $f_i$  of the equation (5.11) makes a steep variation in several localised regions that may contain shocks or contact discontinuities. In these regions, the order of variation of  $f_i$  in spaces  $\mathbf{x}$  and time  $t$  variable is  $O(\varepsilon)$ . In other regions, which are called Euler regions,  $f_i$  has a moderate variation in  $\mathbf{x}$  and  $t$  in the order of unity. Then, the solution  $f_i$  of equation (5.29) in the limit  $\varepsilon \mapsto 0$  is given by  $f_i = f_i^{eq}(\rho, u_\alpha, T)$  whose macroscopic variables  $\rho$ ,  $u_\alpha$  and  $T$  satisfy the weak form of the Euler equation (5.28).*

## 5.4 Optimality conditions

In this section, we derive the optimality conditions for the optimal control problem (5.4). Our approach comprises replacing the constraints equations given by (5.2) with the lattice Boltzmann approximation (5.11). Hence, we derive the optimality equation at the kinetic level and obtain an adjoint equation that should be solved backwards in time and the derivative of the reduced cost functional that plays a crucial role in the update of the control, which is the initial condition of the

flow equations. We introduce the Lagrange multiplier  $\lambda : \mathbb{R}^+ \times \mathbb{R}^d \rightarrow \mathbb{R}^N$  and the Lagrangian

$$L(f, \lambda) = J(U(T, \cdot), U^0; U_d) - \sum_{i=0}^{N-1} \int_{\mathbb{R}^d} \left[ \frac{\partial f_i}{\partial t} + \xi_{i\beta} \frac{\partial f_i}{\partial x_\beta} - \Omega(f_i) \right] \lambda_i d\mathbf{x} dt \quad (5.30)$$

where we assume that the function  $\lambda$  is smooth with compact support. To derive the formal first-order optimality system, we first set the variations of  $L(f, \lambda)$  with respect to each of the functions  $\lambda_i$ ,  $f_i$  and  $U^0$  equal to zero. Setting the first partial derivative of  $L(f, \lambda)$  in (5.30) with respect to  $\lambda_i$  equal to zero, we obtain the Lattice Boltzmann equation (5.11). Furthermore, since the Lagrange parameter vanishes at infinity, integration by parts allows us to write

$$\begin{aligned} L(f, \lambda) = J(U(T, \cdot), U^0; U_d) - \sum_{i=0}^{N-1} \int_{\mathbb{R}^d} [\lambda_i(T, \mathbf{x}) f_i(T, \mathbf{x}) - \lambda_i(0, \mathbf{x}) f_i(0, \mathbf{x})] d\mathbf{x} \\ + \sum_{i=0}^{N-1} \int_0^T \int_{\mathbb{R}^d} \left[ f_i \frac{\partial \lambda_i}{\partial t} + \xi_{i\beta} f_i \frac{\partial \lambda_i}{\partial x_\beta} + \lambda_i \Omega(f_i) \right] d\mathbf{x} dt. \end{aligned} \quad (5.31)$$

Setting the derivative of  $L(\cdot, \cdot)$  in (5.31) with respect to  $f_i$  equal to zero and taking into account (5.10) gives

$$\frac{\partial L(\cdot, \cdot)}{\partial f_i} = \sum_{i=0}^{N-1} \int_0^T \int_{\mathbb{R}^d} \left[ \frac{\partial \lambda_i}{\partial t} + \xi_{i\beta} \frac{\partial \lambda_i}{\partial x_\beta} + \sum_{j=0}^{N-1} \frac{\partial \Omega(f_j)}{\partial f_i} \lambda_j \right] d\mathbf{x} dt = 0, \quad (5.32)$$

Therefore, the adjoint variable  $\lambda_i$  satisfies

$$-\frac{\partial \lambda_i}{\partial t} - \xi_{i\beta} \frac{\partial \lambda_i}{\partial x_\beta} = \sum_{j=0}^{N-1} \frac{\partial \Omega(f_j)}{\partial f_i} \lambda_j. \quad (5.33)$$

Equation (5.33) is the adjoint equation that will be solved backwards in time in our numerical algorithm.

In the two-dimensional cases, we write the space variables as  $\mathbf{x} = (x, y)$  and the adjoint equation (5.33) becomes

$$-\frac{\partial \lambda_i}{\partial t} - \xi_{i1} \frac{\partial \lambda_i}{\partial x} - \xi_{i2} \frac{\partial \lambda_i}{\partial y} = \sum_{j=0}^{N-1} \frac{\partial \Omega(f_j)}{\partial f_i} \lambda_j. \quad (5.34)$$

In this case, given the objective function of the matching type given in (5.4), the terminal conditions  $\lambda_i(T, x, y)$  can be obtained by setting the partial derivatives of  $L(\cdot, \cdot)$  in (5.31) with respect to

$f_i(T, x, y)$  equal to zero, which gives

$$\begin{aligned} \lambda_i(T, \cdot) = & (\rho(T, \cdot) - \rho_d(\cdot)) \frac{\partial \rho}{\partial f_i} + (m_1(T, \cdot) - m_{1,d}(\cdot)) \frac{\partial m_1}{\partial f_i} + (m_2(T, \cdot) - m_{2,d}(\cdot)) \frac{\partial m_2}{\partial f_i} \\ & + (E(T, \cdot) - E_d(\cdot)) \frac{\partial E}{\partial f_i}, \quad (x, y) \in \mathbb{R}^2, \end{aligned} \quad (5.35)$$

with

$$\frac{\partial \rho}{\partial f_i} = 1, \quad \frac{\partial m_1}{\partial f_i} = \xi_{1,i}, \quad \frac{\partial m_2}{\partial f_i} = \xi_{2,i}, \quad \text{and} \quad \frac{\partial E}{\partial f_i} = \xi_{1,i}^2 + \xi_{2,i}^2 + \eta_i^2. \quad (5.36)$$

*Remark.* The adjoint equation (5.34) has the same structure as the original model (5.11). Thus, the term on the right-hand side of (5.34) can be referred to as in [70] as the **adjoint collision operator**. In the BGK formulation, the adjoint collision operator has the form

$$\sum_{j=0}^{N-1} \frac{\partial \Omega_j(f)}{\partial f_i} \lambda_j = \frac{1}{\varepsilon} \left( \sum_{j=0}^{N-1} \frac{\partial f_j^{eq}}{\partial f_i} \lambda_j - \lambda_i \right) \quad (5.37)$$

We can derive formally using a standard technique the adjoint system related to the optimization of Euler flows. Then, we obtain obtain a backward linear system of conservation laws in the adjoint variables. On the other hand, we can consider the moment of the adjoint lattice Boltzmann system obtained above. For example, applying  $\sum_{i=0}^{N-1}$  to (5.33) leads to the equation

$$-\frac{\partial \lambda}{\partial t} - \nabla \lambda \tilde{U} = \frac{1}{\varepsilon} \left( \sum_{j=0}^{N-1} \lambda_j^{eq} - \lambda \right), \quad (5.38)$$

where  $\lambda = \sum_{i=0}^{N-1} \lambda_i$  can be seen as an adjoint density and  $\lambda \tilde{U} = \sum_{i=0}^{N-1} \xi_i \lambda_i$  is an adjoint momentum. Also, we can derive an equation for the adjoint momentum and the adjoint energy. The result is a nonlinear system of conservation law with a source term that depends on the moment of the adjoint equilibrium distribution. With a suitable choice of the adjoint equilibrium distribution, this source term can vanish, which requires, for example, that  $\lambda = \sum_{i=0}^{N-1} \lambda_i^{eq}$ . However, we do not have much degree of freedom in the choice of the equilibrium distributions. It is important to keep in mind that, in general, the equilibrium distributions are found as the minimum of the entropy function under the constraints of conservation of mass and conservation of momentum [131]. Since the adjoint collision operator is a linear combination of the derivatives of the equilibrium distributions  $f_j^{eq}$  with respect to the velocity distributions  $f_i$ , this amounts to impose some constraints on both the



equilibrium and its derivatives. We found that this is meaningful only if the equilibrium functional is linear in the density and velocity. But, this case is not of much interest in practical problems. We conjecture that since our direct adjoint equation is linear and the adjoint obtained via LBM is nonlinear, the moment of the adjoint distribution does not have a direct physical meaning.

For the equilibrium distribution function given in (5.24), we obtain

$$\frac{\partial f_j^{eq}}{\partial f_i} = \frac{\partial f_j^{eq}}{\partial \rho} \frac{\partial \rho}{\partial f_i} + \frac{\partial f_j^{eq}}{\partial m_1} \frac{\partial m_1}{\partial f_i} + \frac{\partial f_j^{eq}}{\partial m_2} \frac{\partial m_2}{\partial f_i} + \frac{\partial f_j^{eq}}{\partial E} \frac{\partial E}{\partial f_i}. \quad (5.39)$$

The partial derivatives of the equilibrium distribution function with respect to the macroscopic variables can be obtained as:

$$\begin{aligned} \frac{\partial f_j^{eq}}{\partial \rho} &= A_j + \rho \frac{\partial A_j}{\partial \rho} + (m_1 \xi_{1,j} + m_2 \xi_{2,j}) \frac{\partial B_j}{\partial \rho} - \frac{1}{\rho^2} (m_1^2 \xi_{1,j}^2 + m_2^2 \xi_{2,j}^2) D_j, \\ \frac{\partial f_j^{eq}}{\partial m_1} &= \rho \frac{\partial A_j}{\partial m_1} + \xi_{1,j} \left( B_j + m_1 \frac{\partial B_j}{\partial m_1} \right) + \frac{2m_1 \xi_{1,j}^2}{\rho} D_j, \\ \frac{\partial f_j^{eq}}{\partial m_2} &= \rho \frac{\partial A_j}{\partial m_2} + \xi_{2,j} \left( B_j + m_2 \frac{\partial B_j}{\partial m_2} \right) + \frac{2m_2 \xi_{2,j}^2}{\rho} D_j, \\ \frac{\partial f_j^{eq}}{\partial E} &= \rho \frac{\partial A_j}{\partial E} + (m_1 \xi_{1,j} + m_2 \xi_{2,j}) \frac{\partial B_j}{\partial E}. \end{aligned} \quad (5.40)$$

Therein,

$$\frac{\partial A_j}{\partial \rho} = \begin{cases} \frac{b-2}{\eta_0^2} \left( \frac{2(m_1^2+m_2^2)-\rho E}{b\rho^3} \right), & \text{for } j=0, \\ \frac{1}{4(c_1^2-c_2^2)} \left[ -c_2^2 + \left( (b-2) \frac{c_2^2}{\eta_0^2} + 2 \right) \frac{2(m_1^2+m_2^2)-\rho E}{b\rho^3} - \frac{2c_2^2(m_1^2+m_2^2)}{c_1^2\rho^3} \right], & \text{for } j=1,2,3,4, \\ \frac{1}{4(c_2^2-c_1^2)} \left[ -c_1^2 + \left( (b-2) \frac{c_1^2}{\eta_0^2} + 2 \right) \frac{2(m_1^2+m_2^2)-\rho E}{b\rho^3} - \frac{2c_1^2(m_1^2+m_2^2)}{c_2^2\rho^3} \right], & \text{for } j=5,6,7,8; \end{cases} \quad (5.41)$$

and

$$\frac{\partial B_j}{\partial \rho} = \begin{cases} 0, & \text{for } j = 0, \\ \frac{1}{2c_1^2(c_1^2 - c_2^2)} \left[ -c_2^2 + \frac{(b+2)[2(m_1^2 + m_2^2) - \rho E]}{b\rho^3} - \frac{2(m_1^2 + m_2^2)}{\rho^3} \right], & \text{for } j = 1, 2, 3, 4, \\ \frac{1}{2c_2^2(c_2^2 - c_1^2)} \left[ -c_1^2 + \frac{(b+2)[2(m_1^2 + m_2^2) - \rho E]}{b\rho^3} - \frac{2(m_1^2 + m_2^2)}{\rho^3} \right], & \text{for } j = 5, 6, 7, 8; \end{cases} \quad (5.42)$$

and

$$\frac{\partial A_j}{\partial m_1} = \begin{cases} \frac{b-2}{\eta_0^2} \left( \frac{-2m_1}{b\rho^2} \right), & \text{for } j = 0, \\ \frac{1}{4(c_1^2 - c_2^2)} \left[ -c_2^2 - \left( (b-2)\frac{c_2^2}{\eta_0^2} + 2 \right) \frac{2m_1}{b\rho^2} + \frac{2c_2^2 m_1}{c_1^2 \rho^2} \right], & \text{for } j = 1, 2, 3, 4, \\ \frac{1}{4(c_2^2 - c_1^2)} \left[ -c_1^2 - \left( (b-2)\frac{c_1^2}{\eta_0^2} + 2 \right) \frac{2m_1}{b\rho^2} + \frac{2c_1^2 m_1}{c_2^2 \rho^2} \right], & \text{for } j = 5, 6, 7, 8; \end{cases} \quad (5.43)$$

and

$$\frac{\partial B_j}{\partial m_1} = \begin{cases} 0, & \text{for } j = 0, \\ \frac{1}{2c_1^2(c_1^2 - c_2^2)} \left[ -\left( c_2^2 + \frac{4m_1}{b\rho^2} \right) \right], & \text{for } j = 1, 2, 3, 4, \\ \frac{1}{2c_2^2(c_2^2 - c_1^2)} \left[ -\left( c_1^2 + \frac{4m_1}{b\rho^2} \right) \right], & \text{for } j = 5, 6, 7, 8; \end{cases} \quad (5.44)$$

and

$$\frac{\partial A_j}{\partial m_2} = \begin{cases} \frac{b-2}{\eta_0^2} \left( \frac{-2m_2}{bp^2} \right), & \text{for } j = 0, \\ \frac{1}{4(c_1^2 - c_2^2)} \left[ -c_2^2 - \left( (b-2) \frac{c_2^2}{\eta_0^2} + 2 \right) \frac{2m_2}{bp^2} + \frac{2c_2^2 m_2}{c_1^2 p^2} \right], & \text{for } j = 1, 2, 3, 4, \\ \frac{1}{4(c_2^2 - c_1^2)} \left[ -c_1^2 - \left( (b-2) \frac{c_1^2}{\eta_0^2} + 2 \right) \frac{2m_2}{bp^2} + \frac{2c_1^2 m_2}{c_2^2 p^2} \right], & \text{for } j = 5, 6, 7, 8; \end{cases} \quad (5.45)$$

and

$$\frac{\partial B_j}{\partial m_2} = \begin{cases} 0, & \text{for } j = 0, \\ \frac{1}{2c_1^2(c_1^2 - c_2^2)} \left[ - \left( c_2^2 + \frac{4m_2}{bp^2} \right) \right], & \text{for } j = 1, 2, 3, 4, \\ \frac{1}{2c_2^2(c_2^2 - c_1^2)} \left[ - \left( c_1^2 + \frac{4m_2}{bp^2} \right) \right], & \text{for } j = 5, 6, 7, 8; \end{cases} \quad (5.46)$$

and

$$\frac{\partial A_j}{\partial E} = \begin{cases} \frac{b-2}{\eta_0^2 bp}, & \text{for } j = 0, \\ \frac{1}{4(c_1^2 - c_2^2)} \left[ -c_2^2 + \left( (b-2) \frac{c_2^2}{\eta_0^2} + 2 \right) \frac{1}{bp} \right], & \text{for } j = 1, 2, 3, 4, \\ \frac{1}{4(c_2^2 - c_1^2)} \left[ -c_1^2 + \left( (b-2) \frac{c_1^2}{\eta_0^2} + 2 \right) \frac{1}{bp} \right], & \text{for } j = 5, 6, 7, 8; \end{cases} \quad (5.47)$$

and

$$\frac{\partial B_j}{\partial E} = \begin{cases} 0, & \text{for } j = 0, \\ \frac{1}{2c_1^2(c_1^2 - c_2^2)} \left[ - \left( c_2^2 - \frac{b+2}{b\rho} \right) \right], & \text{for } j = 1, 2, 3, 4, \\ \frac{1}{2c_2^2(c_2^2 - c_1^2)} \left[ - \left( c_1^2 - \frac{b+2}{b\rho} \right) \right], & \text{for } j = 5, 6, 7, 8; \end{cases} \quad (5.48)$$

and  $D_j$  as introduced above.

The **adjoint equilibrium distribution function** takes the form

$$\lambda_i^{eq} = \sum_{j=0}^{N-1} \frac{\partial f_j^{eq}}{\partial f_i} \lambda_j. \quad (5.49)$$

The optimality conditions are obtained by setting the partial derivative of  $L(\cdot, \cdot)$  in (5.31) with respect to  $U^0$  equal to zero

$$\frac{\partial L(\cdot, \cdot)}{\partial U^0} = \frac{\partial \tilde{J}(\cdot)}{\partial U^0} - \sum_{i=0}^{N-1} \int_{\mathbb{R}^d} \left( -\lambda_i(0, \mathbf{x}) \frac{\partial f_i(0, \mathbf{x})}{\partial U_i^0} \right) d\mathbf{x} = 0, \quad (5.50)$$

In the two-dimensional cases, equation (5.50) becomes

$$\frac{\partial L(\cdot, \cdot)}{\partial U^0} = \frac{\partial \tilde{J}(\cdot)}{\partial U^0} - \sum_{i=0}^{N-1} \int_{\omega} \int_{\mathfrak{D}} \left( -\lambda_i(0, x, y) \frac{\partial f_i(0, x, y)}{\partial U_i^0} \right) dx dy = 0, \quad (5.51)$$

where  $\tilde{J}(\cdot)$  is the reduced cost functional, which is the cost functional seen as a function of the control variable  $U^0$  only. Then, we obtain the gradient of the reduced cost functional for our optimal control problem as

$$\nabla_{U_i^0} \tilde{J} = \sum_{i=0}^{N-1} \int_{\omega} \int_{\mathfrak{D}} \frac{\partial f_i^{eq}(\rho^0, m_1^0, m_2^0, E^0)}{\partial U_i^0} \lambda_i(0, x, y) dx dy. \quad (5.52)$$

There is another way to derive the optimality conditions obtained above. We are going to present it herein by introducing the cost functional in one macroscopic variable for convenience, for example, density  $\rho$ , as

$$\begin{aligned} J(U(T, \cdot), U^0; U_d) &= \frac{1}{2} \int_{\mathbb{R}^d} (\rho(T, \mathbf{x}) - \rho_d(\mathbf{x}))^2 d\mathbf{x} \\ &= \frac{1}{2} \int_{\mathbb{R}^d} (\rho - \rho_d)^2 d\mathbf{x}, \end{aligned} \quad (5.53)$$

A variation in the initial control along the distributional level will cause a variation in the cost functional, as

$$\delta J = \sum_{i=0}^{N-1} \int_{\mathbb{R}^d} (\rho - \rho_d) \frac{\partial \rho}{\partial f_i} \delta f_i d\mathbf{x}. \quad (5.54)$$

Now, using the Lagrange multiplier approach to add the Lattice Boltzmann equation (5.11) to the variation of the cost functional. The variation of the cost functional as a constraint can be obtained as

$$\delta J(\cdot) = \sum_{i=0}^{N-1} \int_{\mathbb{R}^d} (\rho - \rho_d) \frac{\partial \rho}{\partial f_i} \delta f_i d\mathbf{x} - \sum_{i=0}^{N-1} \int_0^T \int_{\mathbb{R}^d} \delta \left[ \frac{\partial f_i}{\partial t} + \xi_{i\beta} \frac{\partial f_i}{\partial x_\beta} - \Omega(f_i) \right] \lambda_i d\mathbf{x} dt, \quad (5.55)$$

We can rewrite (5.55) as

$$\begin{aligned} \delta J(\cdot) &= \sum_{i=0}^{N-1} \int_{\mathbb{R}^d} (\rho - \rho_d) \frac{\partial \rho}{\partial f_i} \delta f_i d\mathbf{x} - \sum_{i=0}^{N-1} \int_0^T \int_{\mathbb{R}^d} \delta \left( \frac{\partial f_i}{\partial t} \right) \lambda_i d\mathbf{x} dt - \sum_{i=0}^{N-1} \int_0^T \int_{\mathbb{R}^d} \delta \left( \xi_{i\beta} \frac{\partial f_i}{\partial x_\beta} \right) \lambda_i d\mathbf{x} dt \\ &\quad + \sum_{i=0}^{N-1} \int_0^T \int_{\mathbb{R}^d} \delta(\Omega(f_i)) \lambda_i d\mathbf{x} dt. \end{aligned} \quad (5.56)$$

From the second term of equation (5.56), we have

$$\begin{aligned} \delta I_1(\cdot) &= \sum_{i=0}^{N-1} \int_0^T \int_{\mathbb{R}^d} \delta \left( \frac{\partial f_i}{\partial t} \right) \lambda_i d\mathbf{x} dt \\ &= \sum_{i=0}^{N-1} \int_{\mathbb{R}^d} [\lambda_i(T, \cdot) \delta f_i(T, \cdot) - \lambda_i(0, \cdot) \delta f_i(0, \cdot)] d\mathbf{x} - \sum_{i=0}^{N-1} \int_0^T \int_{\mathbb{R}^d} \delta f_i \frac{\partial \lambda_i}{\partial t} d\mathbf{x} dt, \end{aligned} \quad (5.57)$$

from the third term of equation (5.56), we obtain

$$\begin{aligned} \delta I_2(\cdot) &= \sum_{i=0}^{N-1} \int_0^T \int_{\mathbb{R}^d} \delta \left( \xi_{i\beta} \frac{\partial f_i}{\partial x_\beta} \right) \lambda_i d\mathbf{x} dt \\ &= \sum_{i=0}^{N-1} \int_0^T \int_{\mathbb{R}^d} \lambda_i \left[ \xi_{i\beta} \frac{\partial \delta f_i}{\partial x_\beta} + \frac{\partial f_i}{\partial x_\beta} \delta(\xi_{i\beta}) \right] d\mathbf{x} dt \\ &= \sum_{i=0}^{N-1} \int_0^T \int_{\mathbb{R}^d} \left[ \frac{\partial(\lambda_i \xi_{i\beta} \delta f_i)}{\partial x_\beta} - \frac{\partial(\lambda_i \xi_{i\beta})}{\partial x_\beta} \delta f_i \right] d\mathbf{x} dt \\ &= \sum_{i=0}^{N-1} \int_0^T \int_{\mathbb{R}^d} \left[ \frac{\partial(\lambda_i \xi_{i\beta} \delta f_i)}{\partial x_\beta} - \lambda_i \frac{\partial \xi_{i\beta}}{\partial x_\beta} \delta f_i - \xi_{i\beta} \frac{\partial \lambda_i}{\partial x_\beta} \delta f_i \right] d\mathbf{x} dt, \end{aligned} \quad (5.58)$$

where the variation of discrete velocities  $\xi_i$  vanishes at the discrete level. Also, the last term of equation (5.56) can be written as

$$\begin{aligned}\delta I_3(\cdot) &= \sum_{i=0}^{N-1} \int_0^T \int_{\mathbb{R}^d} \delta(\Omega(f_i)) \lambda_i d\mathbf{x} dt \\ &= \sum_{i=0}^{N-1} \int_0^T \int_{\mathbb{R}^d} \left( \sum_{j=0}^{N-1} \lambda_j \frac{\partial \Omega(f_j)}{\partial f_i} \delta f_i \right) d\mathbf{x} dt.\end{aligned}\quad (5.59)$$

By inserting equations (5.57), (5.58) and (5.59) into equation (5.56) and ignore the terms with the second-order variation and the partial derivatives of discrete velocities  $\xi_{i\beta}$  which vanish at the discrete level, we have

$$\begin{aligned}\delta J(\cdot) &= \sum_{i=0}^{N-1} \int_{\mathbb{R}^d} (\rho - \rho_d) \frac{\partial \rho}{\partial f_i} \delta f_i d\mathbf{x} - \sum_{i=0}^{N-1} \int_{\mathbb{R}^d} [\lambda_i(T, \cdot) \delta f_i(T, \cdot) - \lambda_i(0, \cdot) \delta f_i(0, \cdot)] d\mathbf{x} \\ &+ \sum_{i=0}^{N-1} \int_0^T \int_{\mathbb{R}^d} \left( \delta f_i \frac{\partial \lambda_i}{\partial t} \right) d\mathbf{x} dt + \sum_{i=0}^{N-1} \int_0^T \int_{\mathbb{R}^d} \left( \xi_{i\beta} \frac{\partial \lambda_i}{\partial x_\beta} \delta f_i \right) d\mathbf{x} dt \\ &+ \sum_{i=0}^{N-1} \int_0^T \int_{\mathbb{R}^d} \left( \sum_{j=0}^{N-1} \lambda_j \frac{\partial \Omega(f_j)}{\partial f_i} \delta f_i \right) d\mathbf{x} dt \\ &= \sum_{i=0}^{N-1} \int_{\mathbb{R}^d} \left[ (\rho - \rho_d) \frac{\partial \rho}{\partial f_i} - \lambda_i(T, \cdot) \right] \delta f_i d\mathbf{x} + \sum_{i=0}^{N-1} \int_{\mathbb{R}^d} (\lambda_i(0, \cdot) \delta f_i(0, \cdot)) d\mathbf{x} \\ &+ \sum_{i=0}^{N-1} \int_0^T \int_{\mathbb{R}^d} \left[ \frac{\partial \lambda_i}{\partial t} \delta f_i + \xi_{i\beta} \frac{\partial \lambda_i}{\partial x_\beta} \delta f_i + \sum_{j=0}^{N-1} \lambda_j \frac{\partial \Omega(f_j)}{\partial f_i} \delta f_i \right] d\mathbf{x} dt.\end{aligned}\quad (5.60)$$

To eliminate the influence of flow field changes on cost functional gradients, we make the variation (5.60) be zero. Therefore, from the last term of equation (5.60), we have

$$\sum_{i=0}^{N-1} \int_0^T \int_{\mathbb{R}^d} \left[ \frac{\partial \lambda_i}{\partial t} + \xi_{i\beta} \frac{\partial \lambda_i}{\partial x_\beta} + \sum_{j=0}^{N-1} \lambda_j \frac{\partial \Omega(f_j)}{\partial f_i} \right] \delta f_i d\mathbf{x} dt = 0, \quad (5.61)$$

hence, the adjoint equation of the optimal control problem (5.4) can be obtained as

$$-\frac{\partial \lambda_i}{\partial t} - \xi_{i\beta} \frac{\partial \lambda_i}{\partial x_\beta} = \sum_{j=0}^{N-1} \lambda_j \frac{\partial \Omega(f_j)}{\partial f_i}, \quad (5.62)$$

while the terminal conditions  $\lambda_i(T, \cdot)$  can be found by setting the variation of equation (5.60) with respect to  $f_i$  at a terminal time to zero

$$\lambda_i(T, \cdot) = (\rho(T, \cdot) - \rho_d(\cdot)) \frac{\partial \rho}{\partial f_i}. \quad (5.63)$$

In general, the optimality condition is obtained by setting the gradient of (5.60) with respect to  $\rho^0$  to zero as

$$\nabla_{\rho_i^0} \tilde{J} = \sum_{i=0}^{N-1} \int_{\mathbb{R}^d} \frac{\partial f_i^{eq}(\rho^0)}{\partial \rho_i^0} \lambda_i(0, \cdot) d\mathbf{x}. \quad (5.64)$$

Furthermore, we can write the optimality conditions (5.62) - (5.64) in the macroscopic variables  $\rho$ ,  $m_1$ ,  $m_2$  and  $E$  of Euler equations (5.6) to obtain the full versions same as in the above derivations. We have then derived the key ingredients of our optimisation algorithm to be described below.

## 5.5 Numerical analysis and results

In this section, we discuss the numerical algorithm for the solution of the optimal control problem. We solve the lattice Boltzmann equations (5.11) to obtain the discrete velocities  $f_i$ . We then take the moments to obtain the macroscopic variables at any time  $t \in [0, T]$ . These variables are then used to solve backward in time the microscopic adjoint equation (5.33) for the adjoint variables  $\lambda_i$ . Thereafter, we use the computed value of  $f_i$  and the  $\lambda_i$  to compute the gradient of the reduced cost functional as in (5.52), which is used for the update of the control  $U^0$ . The algorithm used in this study is presented in Algorithm 1.

We point out that, for each particle  $i$  with speed  $\xi_i$ , the lattice Boltzmann equation (5.11) and its adjoint form (5.33) are transport equations with the source term. Therefore, we discretise them in the finite volume framework using a second-order integration in time and a second-order upwind integration in space with the minmod slope limiters [132] as briefly described below. We consider the two-dimensional advection equation in the general form

$$\begin{cases} \frac{\partial V}{\partial t} + a \frac{\partial V}{\partial x} + b \frac{\partial V}{\partial y} = g(V), & (t, x, y) \in [0, T] \times [0, 1] \times [0, 1], \\ V(0, x, y) = V^0(x, y), & (x, y) \in [0, 1] \times [0, 1], \end{cases} \quad (5.65)$$

where  $a$  and  $b$  are the wave speeds and  $g(V)$  is source term, function of the balanced quantity  $V$ . We consider a uniform grid in spaces  $x_j \doteq j\Delta x$ ,  $y_k \doteq k\Delta y$ ,  $j, k = 0, \dots, K$  and the time grid is denoted as  $t_n = n\Delta t$ ,  $n = 0, \dots, H$  with the space steps  $\Delta x$  and  $\Delta y$  destined to tend to zero and the time step

---

**Algorithm 1:** Numerical algorithm for the solution of the optimal control problem

---

$U_0(t, \mathbf{x})$  : Chosen initial control variable, initial data for the problem (5.33)

$U_d(\mathbf{x})$  : Desired solution

$\tau$  : Given tolerance

$T$  : Final simulation time for the flow solvers.

- Consider an initial distribution  $f_i^0$  such that the equation (5.10) are satisfied and solve the equation (5.11) with the initial conditions (5.13).

Compute the macroscopic variable by averaging the distributions using (5.10).

for  $k = 0, 1, 2, \dots$

Compute the cost functional  $J(U(T, \mathbf{x}), U_d(\mathbf{x}))$

which in the two-dimensional case amounts to

$$J^{(k)} = \frac{1}{2} \Delta x \Delta y \left( \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} (\rho_{ij}^{(k)} - \rho_{ij}^d)^2 + (m_{1ij}^{(k)} - m_{1ij}^d)^2 + (m_{2ij}^{(k)} - m_{2ij}^d)^2 + (E_{ij}^{(k)} - E_{ij}^d)^2 \right)$$

While  $J^{(k)} > \tau$

Solve the adjoint equation (5.34) backward in time using the terminal conditions (5.35)

Update the control variable using the gradient of the reduced cost functional (5.52) as

$$(U^0)^{new} = (U^0)^{old} - \alpha \nabla_{(U^0)^{old}} \tilde{J}$$

and the step length  $\alpha$  is chosen using either the Armijo rule or any step selection procedure.

Set  $k = k + 1$

---



is chosen according to the CFL condition  $CFL \leq 1$  with

$$CFL = \max_{i=0,\dots,N-1} \left( \left| \frac{\Delta t}{\Delta x} a \right|, \left| \frac{\Delta t}{\Delta y} b \right| \right). \quad (5.66)$$

In the finite volume framework, we introduce the rectangular cells  $I_{j,k} = [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}] \times [y_{k-\frac{1}{2}}, y_{k+\frac{1}{2}}]$  where the cells boundary are defined as  $x_{j+1/2} = \frac{1}{2}(x_j + x_{j+1})$  and  $y_{k+1/2} = \frac{1}{2}(y_k + y_{k+1})$ . The cell average of the balanced quantity  $V(t, x, y)$  for rectangular cells at time  $t_n$  is denoted as  $V_{j,k}^n$ ,  $j, k = 0, \dots, K$  with

$$V_{j,k}(t) = \frac{1}{\Delta x \Delta y} \int_{I_{j,k}} V(t, x, y) dx dy.$$

The resulting second-order scheme takes the semi-discrete form [47, 132]

$$\frac{dV_{j,k}}{dt} = -\frac{F_{j+\frac{1}{2},k} - F_{j-\frac{1}{2},k}}{\Delta x} - \frac{G_{j,k+\frac{1}{2}} - G_{j,k-\frac{1}{2}}}{\Delta y} + g_{j,k}, \quad (5.67)$$

where  $g_{j,k}$  is the cell average of the source term and the numerical fluxes are given by

$$F_{j+\frac{1}{2},k} = a^- V_{j+1,k} + a^+ V_{j,k} + \frac{1}{2} |a| \left( 1 - \left| \frac{a \Delta t}{\Delta x} \right| \right) \sigma_{j+\frac{1}{2},k}, \quad (5.68)$$

$$G_{j,k+\frac{1}{2}} = b^- V_{j,k+1} + b^+ V_{j,k} + \frac{1}{2} |b| \left( 1 - \left| \frac{b \Delta t}{\Delta y} \right| \right) \sigma_{j,k+\frac{1}{2}}, \quad (5.69)$$

where  $a^+ = \max\{a, 0\}$ ,  $a^- = \min\{a, 0\}$ ,  $b^+ = \max\{0, b\}$ , and  $b^- = \min\{0, b\}$  and the slope limiters  $\sigma_{j+\frac{1}{2},k}$  and  $\sigma_{j,k+\frac{1}{2}}$  are defined as

$$\sigma_{j+\frac{1}{2},k} = \begin{cases} \text{Minmod} \left( \frac{V_{j,k} - V_{j-1,k}}{\Delta x}, \frac{V_{j+1,k} - V_{j,k}}{\Delta x} \right) & \text{if } a \geq 0, \\ \text{Minmod} \left( \frac{V_{j+1,k} - V_{j,k}}{\Delta x}, \frac{V_{j+2,k} - V_{j+1,k}}{\Delta x} \right) & \text{if } a < 0, \end{cases} \quad (5.70)$$

$$\sigma_{j,k+\frac{1}{2}} = \begin{cases} \text{Minmod} \left( \frac{V_{j,k} - V_{j,k-1}}{\Delta y}, \frac{V_{j,k+1} - V_{j,k}}{\Delta y} \right) & \text{if } b \geq 0, \\ \text{Minmod} \left( \frac{V_{j,k+1} - V_{j,k}}{\Delta y}, \frac{V_{j,k+2} - V_{j,k+1}}{\Delta y} \right) & \text{if } b < 0, \end{cases} \quad (5.71)$$

with

$$\text{Minmod}(A, B) \doteq \frac{1}{2} (\text{sgn}(A) + \text{sgn}(B)) \cdot \min(|A|, |B|).$$

The mesh size in time is set as  $\Delta t = \tau/4$  where  $\tau$  is the Knudsen number. This choice ensures that the CFL condition is satisfied. For our numerical results, we used  $\tau = 10^{-4}$ . For the source term, we use the mid-point quadrature rule.

*Remark.* It is very important in the numerical algorithm to solve the flow equation and the adjoint equation on the same space grid. Also, for the solution of the optimal control problem, we use for simplicity a constant step length.

### 5.5.1 Solution of the flow equations

We consider the two dimensional Euler equations on the space domain  $[0, 1] \times [0, 1]$  with the initial conditions

$$\begin{aligned}
 \rho(0, x, y) &= 1 + x^2 + y^2, \\
 u(0, x, y) &= \sin(\pi x) \cos(\pi y), \\
 v(0, x, y) &= -\cos(\pi x) \sin(\pi y), \\
 P(0, x, y) &= 1 + \frac{1}{4}(\sin(2\pi x) + \cos(2\pi y)).
 \end{aligned}
 \tag{5.72}$$

We solve the Lattice Boltzmann equation using the numerical scheme in (5.67) with the numerical fluxes given in (5.68) and (5.69). We use the forward Euler method for time discretisation. The contour plots of the density, pressure and energy are presented in Figure 5.2 where we also included the initial conditions. We also solved directly the Euler equation using a finite volume method [132]. The results are presented in Figure 5.3 where again, we have included the initial conditions. We expect the results in the two figures to look similar as indicated in Proposition 5.3.1. They mismatch probably because the parameters that we used in the lattice Boltzmann equations are inappropriate. It is noted in the original paper by Kataoka and Tshutahara [10] that the exact values of the parameters were unspecified. Nevertheless, we present some numerical results in the next section that support our numerical solution of the optimal control problem.

### 5.5.2 Solution to the optimal control problem

Now we consider the solution of the optimal control problem. We will consider two examples and for all the examples, we use for the solution of the flow or the adjoint equation a uniform grid of  $50 \times 50$  grid points. The value of the gas ratio is taken as  $\gamma$  is  $7/3$ .

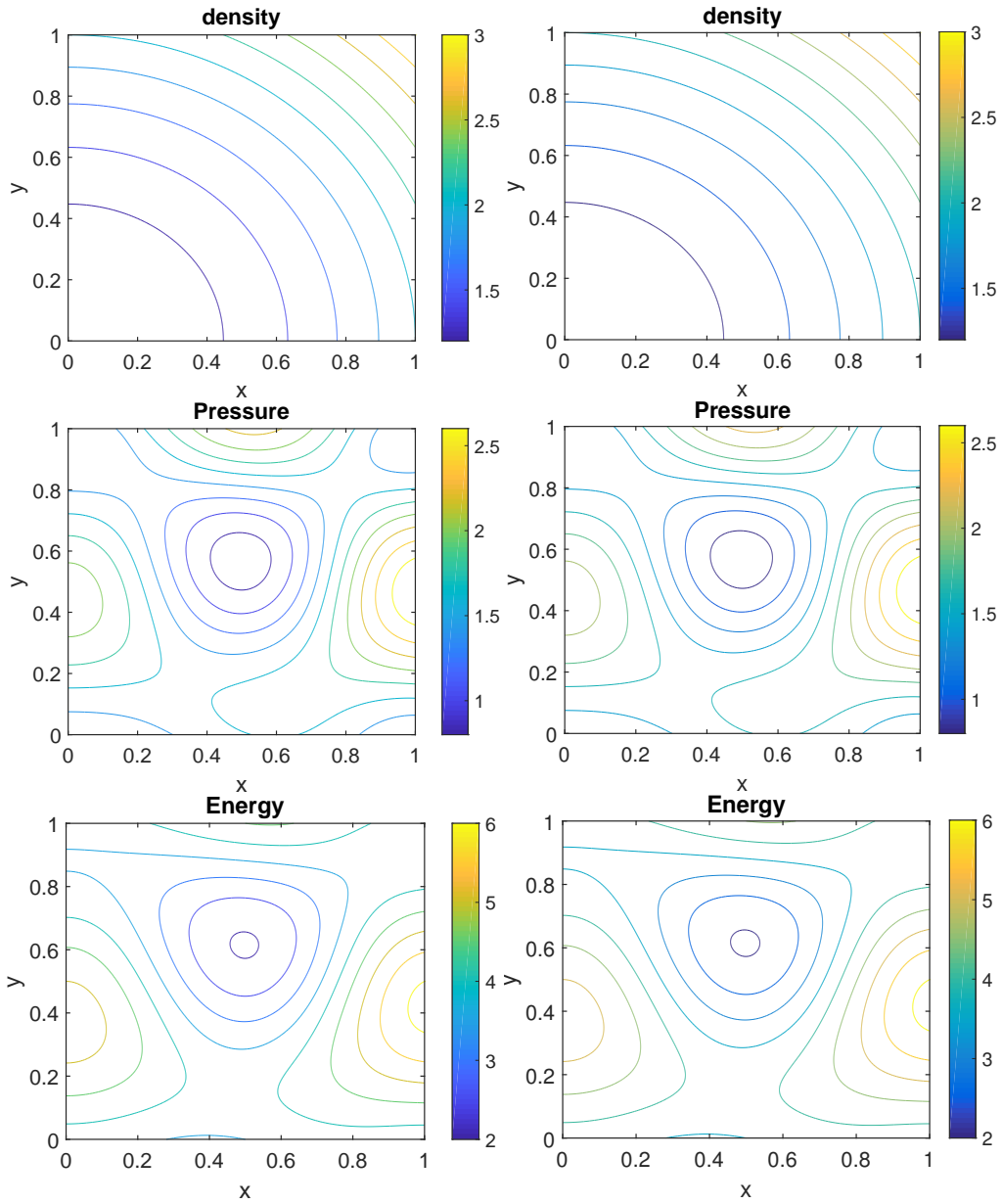


Figure 5.2: Contour plot of the density, pressure and energy for the solution of the Euler equations in two dimensions using the Lattice Boltzmann method. We use 500x500 grid points and the results are computed up to time  $T = 0.0063$ .

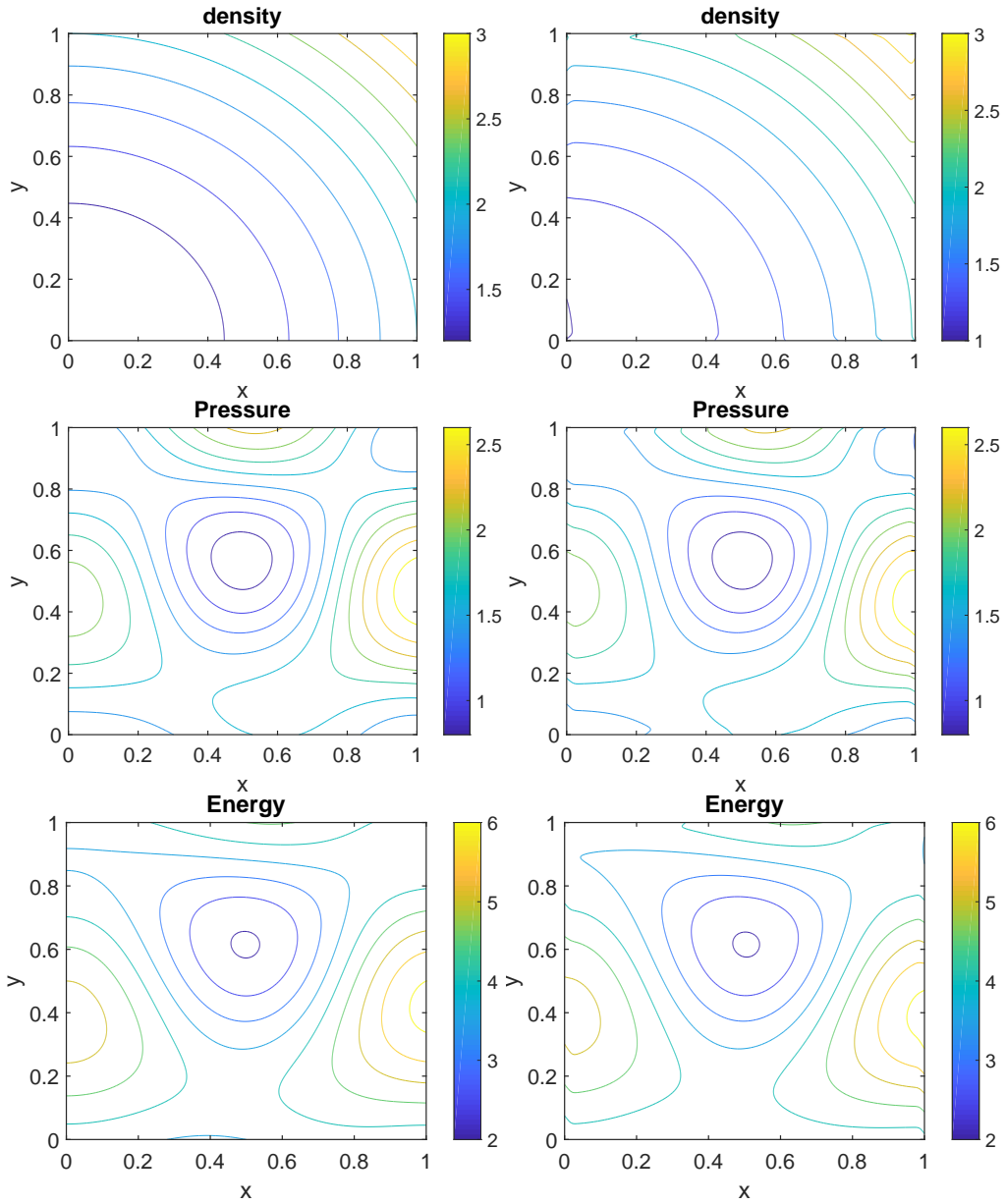


Figure 5.3: Contour plot of the density, pressure and energy for the solution of the Euler equations in two dimensions using the finite volume method. We use 500x500 grid points and the initial condition (left) and the results (right) are computed up to time  $T = 0.0063$ .

In our first example, the desired state is obtained as the solution of the flow equations with the initial data

$$\begin{aligned}
\rho(0, x, y) &= 1.1 + x^2 + y^2, \\
u(0, x, y) &= 0.1(\sin(x)\cos(y)) + \sin(\pi x)\cos(\pi y), \\
v(0, x, y) &= 0.1(\sin(x)\cos(y)) - \cos(\pi x)\sin(\pi y), \\
P(0, x, y) &= 1.1 + \frac{1}{4}(\sin(2\pi x) + \cos(2\pi y)).
\end{aligned} \tag{5.73}$$

The initial controls, which are the initial conditions for the flow equations are given by

$$\begin{aligned}
\rho(0, x, y) &= 1 + x^2 + y^2, \\
u(0, x, y) &= \sin(\pi x)\cos(\pi y), \\
v(0, x, y) &= -\cos(\pi x)\sin(\pi y), \\
P(0, x, y) &= 1 + \frac{1}{4}(\sin(2\pi x) + \cos(2\pi y)).
\end{aligned} \tag{5.74}$$

The aim is to drive the solution computed at time  $T$  to the desired state with our optimal control algorithm. We present in Figure 5.4 the contour plots of the optimal and desired density, pressure and energy. The results are computed at time  $T = 0.0005$  and convergence is achieved after 50 iterations. Our algorithm, which is the steepest descent method successfully drives the initial state to the desired state as can be seen in the graph of the cost functional against the number of iteration in Figure 5.5. We now consider a second example where the desired state is obtained as the solution at time  $T$  of the initial value problem for the two dimensional Euler equations computed using the lattice Boltzmann approximation with the initial data

$$\begin{aligned}
\rho(0, x, y) &= 1.2 + \frac{1}{8}(\sin(\pi xy) + \cos(\pi xy)), \\
u(0, x, y) &= 0.2 + \sin(2\pi xy), \\
v(0, x, y) &= 0.2 + \cos(2\pi xy), \\
P(0, x, y) &= 0.2 - \frac{1}{4}(\sin(\pi xy) + \cos(\pi xy)).
\end{aligned} \tag{5.75}$$

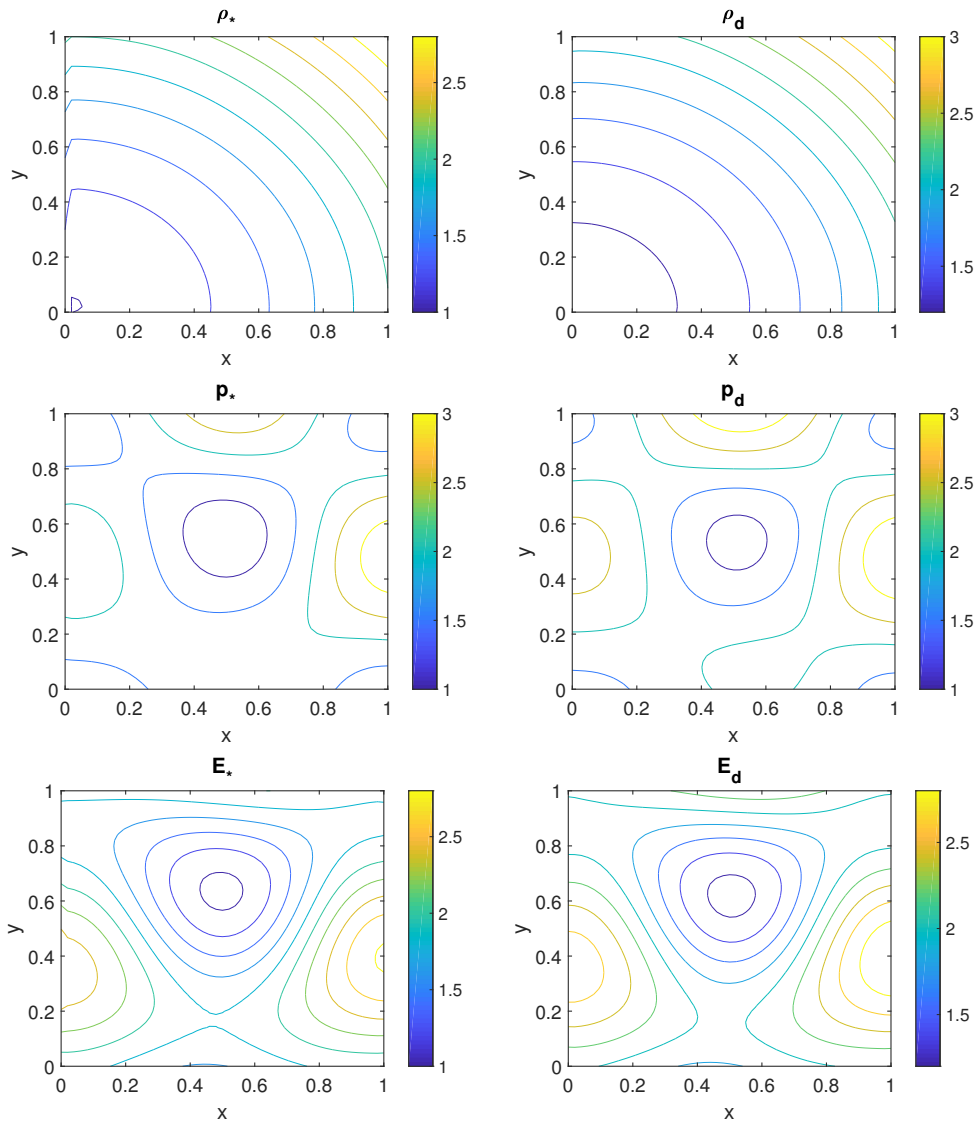


Figure 5.4: Contour plot of the optimal (left) and desired solutions (right) density, pressure and energy for the solution of the optimal control problem for the first example. The results are computed at time  $T = 0.0005$ .

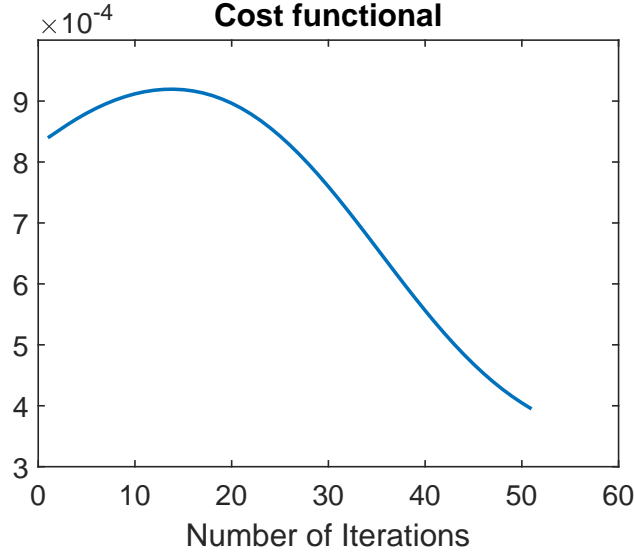


Figure 5.5: Convergence history for the solution of the optimal control problem computed with the tolerance  $\varepsilon = 10^{-3}$  using the initial control (5.74).

The initial control is taken as

$$\begin{aligned}
 \rho(0, x, y) &= 1 + \frac{1}{8} (\sin(\pi xy) + \cos(\pi xy)), \\
 u(0, x, y) &= \sin(2\pi xy), \\
 v(0, x, y) &= \cos(2\pi xy), \\
 P(0, x, y) &= -\frac{1}{4} (\sin(\pi xy) + \cos(\pi xy)).
 \end{aligned} \tag{5.76}$$

We display in Figures 5.6, the optimal (left) and desired (right) contour plots of the density, pressure and energy computed at time  $T = 0.005$ . In this example, just as in the first, the convergence of the algorithm with a tolerance of  $\varepsilon = 10^{-3}$  occurs after 52 iterations as can be seen in Figure 5.7. We note that in Figure 5.4, various perturbations in the desired states were used as initial control for the solution of the optimal control problem. Similar results now shown here were observed. The problem can be solved on a finer grid but will require more time to run.

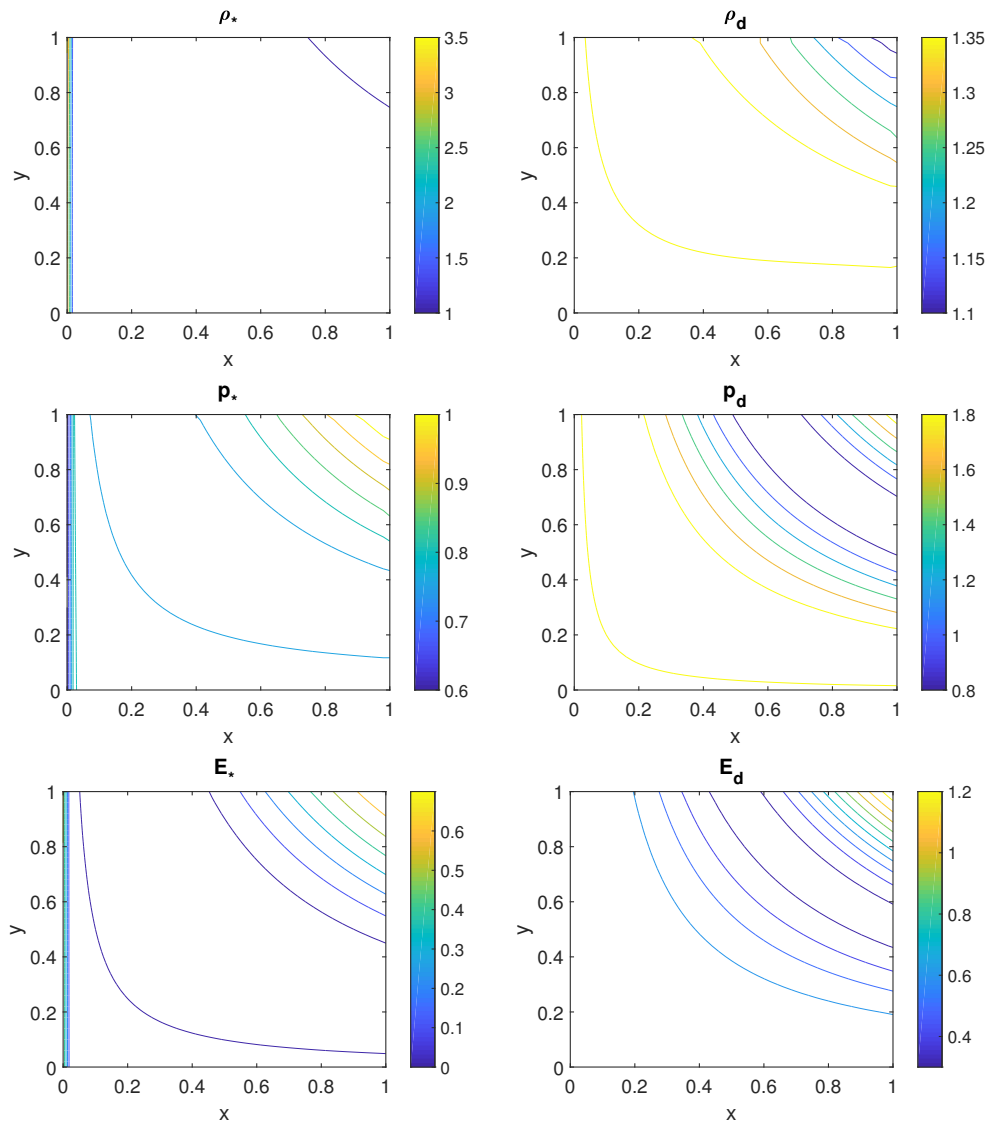


Figure 5.6: Contour plot of the optimal (left) and desired solutions (right) density, pressure and energy for the solution of the optimal control problem for the second example. The results are computed at time  $T = 0.0005$ .



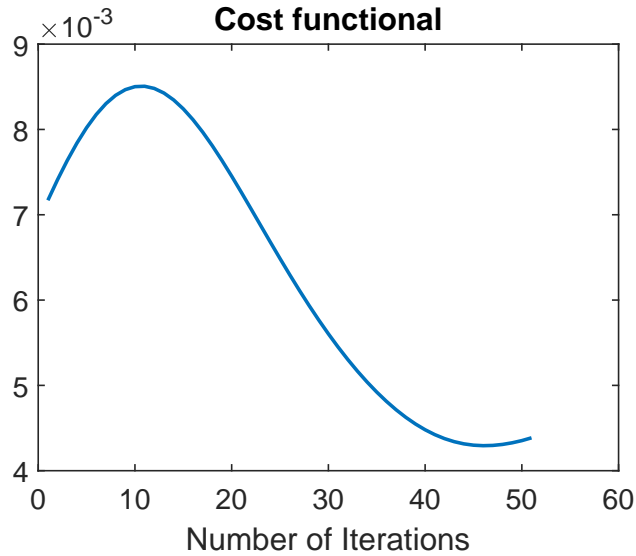


Figure 5.7: Convergence history for the solution of the optimal control problem computed with the tolerance  $\varepsilon = 10^{-3}$  using the initial control (5.76)

## 5.6 Concluding remarks

In this chapter, we have presented a numerical algorithm for the solution of optimal control problems governed by the two-dimensional Euler equations. The particularity of the algorithm comes from the fact that we replaced the Euler equation in our derivation of the optimality condition by a lattice Boltzmann approximation of the Euler equations. For a given level of tolerance, our algorithm allowed us to compute an optimal solution that matches to some extent the desired state, with an error that can be improved. Our algorithm did not do very well in solving problems that involve discontinuities in the initial conditions. This might be attributed to the choice of parameters in the lattice Boltzmann approximation. Our analysis is quite general and can be extended to three-dimensional problems or problems involving discontinuities. This will be one topic in future work.

# Chapter 6

## Conclusion

In this thesis, we presented and assessed a class of numerical methods for solving optimal control problems constrained by systems of conservation laws. Some theoretical perspectives from previously published research on one- and multi-dimensional cases have been substantiated by numerical results. In addition, a review and development of effective numerical methods has been central to our optimisation algorithm. We reviewed the mathematical analysis of the constraint equations and their numerical challenges. Then, we discussed the relaxation approximations for the numerical solutions of the conservation laws. Further, we derived the optimality conditions that lead to the numerical algorithm for solving optimal control problems related to the one-dimensional case. The results obtained with the Jin–Xin relaxation approximations are more accurate than those obtained using the finite volume method (see Sections 2.5 and 2.8 of Chapter 2, as well as Section 3.6 of Chapter 3 for optimal solutions based on relaxation methods). The convergence history against the number of iterations confirmed our findings. Also, it is observed that the algorithm presented made the computed optimal solutions very close to the desired solutions.

The existence and uniqueness results of the entropy solutions to the multi-dimensional conservation laws were presented. If we combine these results with the Karush-Kuhn Tucker conditions of optimisation theory, we have at least formally an existence and uniqueness result for the optimal control problem. Furthermore, we derived optimality conditions using the adjoint approach to the optimal control problems at the continuous level. The optimality conditions were obtained related to the multi-dimensional system of conservation laws and three-dimensional relaxation approximations. These conditions were applied to construct an optimisation algorithm that solves the

optimal control problem numerically. We tested the optimisation algorithm on several examples related to the two-dimensional inviscid Burger's equation. In the numerical algorithm, the flow equations and the adjoint equations were solved on the same grid. The results obtained using both schemes were in excellent agreement with our theoretical analysis. The algorithm presented here can solve smooth problems and problems involving discontinuities (see Section 4.7 of Chapter 4). We applied the results to the multi-dimensional Euler equations of gas dynamics. In the inverse design optimisation, we replaced the Euler equations with the Lattice Boltzmann approximation. Further, we derived the optimality condition based on the two-dimensional, nine velocities Lattice Boltzmann equation. These conditions are used to construct an algorithm that solves our optimal control problem numerically. Also, we presented some numerical results applied to examples related to the smooth initial data.

# References

- [1] S. Nadarajah and A. Jameson. A comparison of the continuous and discrete adjoint approach to automatic aerodynamic optimization. In *38th Aerospace Sciences Meeting and Exhibit*, page 667, 2000.
- [2] W. K. Anderson and V. Venkatakrishnan. Aerodynamic design optimization on unstructured grids with a continuous adjoint formulation. *Computers & Fluids*, 28(4-5):443–480, 1999.
- [3] M. H. Hekmat and M. Mirzaei. A comparison of the continuous and discrete adjoint approach extended based on the standard lattice boltzmann method in flow field inverse optimization problems. *Acta Mechanica*, 227(4):1025–1050, 2016.
- [4] R. J. LeVeque and R. J. Leveque. *Numerical methods for conservation laws*, volume 132. Springer, 1992.
- [5] F. T. Eleuterio. *Riemann solvers and numerical methods for fluid dynamics*. Berlin: Springer Verlag, 1999.
- [6] A. Bressan. Lecture notes on hyperbolic conservation laws. *Department of Mathematics, Penn State University, University park*, pages 1–85, 2009.
- [7] M. E. Vázquez-Cendón. *Solving hyperbolic equations with finite volume methods*, volume 90. Springer, 2015.
- [8] E. F. Toro. *Riemann solvers and numerical methods for fluid dynamics: a practical introduction*. Springer Science & Business Media, 2013.
- [9] S. Jin and Z. Xin. The relaxation schemes for systems of conservation laws in arbitrary space dimensions. *Communications on pure and applied mathematics*, 48(3):235–276, 1995.

- [10] T. Kataoka and M. Tsutahara. Lattice boltzmann method for the compressible euler equations. *Physical review E*, 69(5):056702, 2004.
- [11] P. G. Lefloch. Hyperbolic conservation laws on spacetimes. In *Nonlinear Conservation Laws and Applications*, pages 379–391. Springer, 2011.
- [12] P. Baiti, P. G. LeFloch, and B. Piccoli. Uniqueness of classical and nonclassical solutions for nonlinear hyperbolic systems. *Journal of Differential Equations*, 172(1):59–82, 2001.
- [13] A. Bressan and A. Marson. A variational calculus for discontinuous solutions of systems of conservation laws. *Communications in partial differential equations*, 20(9):1491–1552, 1995.
- [14] S. Bianchini. On the shift differentiability of the flow generated by a hyperbolic system of conservation laws. *Discrete & Continuous Dynamical Systems*, 6(2):329, 2000.
- [15] S. Bianchini and L. Yu. Global structure of admissible bv solutions to piecewise genuinely nonlinear, strictly hyperbolic conservation laws in one space dimension. *Communications in Partial Differential Equations*, 39(2):244–273, 2014.
- [16] J. Glimm. Solutions in the large for nonlinear hyperbolic systems of equations. *Communications on pure and applied mathematics*, 18(4):697–715, 1965.
- [17] A. Bressan, G. Crasta, and B. Piccoli. Well-posedness of the cauchy problem for  $n \times n$  systems of conservation laws. In *Memoirs AMS, no. 694*, American Mathematical Society. Citeseer, 1997.
- [18] A. Bressan, T. Liu, and T. Yang. L1 stability estimates for  $n \times n$  conservation laws. *Archive for rational mechanics and analysis*, 149(1):1–22, 1999.
- [19] A. Bressan and P. G. LeFloch. Structural stability and regularity of entropy solutions to hyperbolic systems of conservation laws. *Indiana University mathematics journal*, pages 43–84, 1999.

- [20] A. Bressan, P. LeFloch, et al. Uniqueness of weak solutions to systems of conservation laws. *Archive for Rational Mechanics and Analysis*, 140(4):301–317, 1997.
- [21] A. Bressan and P. Goatin. Oleinik type estimates and uniqueness for  $n \times n$  conservation laws. *Journal of differential equations*, 156(1):26–49, 1999.
- [22] P. Lions, B. Perthame, and E. Tadmor. A kinetic formulation of multidimensional scalar conservation laws and related equations. *Journal of the American Mathematical Society*, 7(1):169–191, 1994.
- [23] P. G. LeFloch. *Hyperbolic Systems of Conservation Laws: The theory of classical and nonclassical shock waves*. Springer Science & Business Media, 2002.
- [24] Wladimir Neves. Scalar multidimensional conservation laws ibvp in noncylindrical lipschitz domains. *Journal of Differential Equations*, 192(2):360–395, 2003.
- [25] E. Y. Panov. Existence of strong traces for generalized solutions of multidimensional scalar conservation laws. *Journal of Hyperbolic Differential Equations*, 2(04):885–908, 2005.
- [26] G. G. Chen. Multidimensional conservation laws: overview, problems, and perspective. *Nonlinear conservation laws and applications*, pages 23–72, 2011.
- [27] G. Crasta, V. De Cicco, G. De Philippis, and F. Ghiraldin. Structure of solutions of multidimensional conservation laws with discontinuous flux and applications to uniqueness. *Archive for Rational Mechanics and Analysis*, 221(2):961–985, 2016.
- [28] C. Dogbe and C. Bianca. Existence of entropy solutions for multidimensional conservation laws with 11 boundary conditions. *Mathematics in Engineering, Science and Aerospace (MESA)*, 9(4):507–526, 2018.
- [29] F. Tröltzsch. *Optimal control of partial differential equations: theory, methods, and applications*, volume 112. American Mathematical Soc., 2010.
- [30] E. Casas and M. Mateos. Optimal control of partial differential equations. In *Computational mathematics, numerical analysis and applications*, pages 3–59. Springer, 2017.

- [31] M. D. Gunzburger. *Perspectives in Flow Control and Optimization*, volume 5. SIAM, 2003.
- [32] A. Chertock, M. Herty, and A. Kurganov. An eulerian–lagrangian method for optimization problems governed by multidimensional nonlinear hyperbolic pdes. *Computational Optimization and Applications*, 59(3):689–724, 2014.
- [33] I. Cheylan, G. Fritz, D. Ricot, and P. Sagaut. Shape optimization using the adjoint lattice boltzmann method for aerodynamic applications. *AIAA journal*, 57(7):2758–2773, 2019.
- [34] C. Chen, K. Yaji, T. Yamada, K. Izui, and S. Nishiwaki. Local-in-time adjoint-based topology optimization of unsteady fluid flows using the lattice boltzmann method. *Mechanical Engineering Journal*, 4(3):17–00120, 2017.
- [35] M. Morales-Hernández and E. Zuazua. Adjoint computational methods for 2d inverse design of linear transport equations on unstructured grids. *Computational and Applied Mathematics*, 38(4):1–25, 2019.
- [36] R. Lecaros and E. Zuaza. Control of 2d scalar conservation laws in the presence of shocks. 2016.
- [37] M. Herty, A. Kurganov, and D. Kurochkin. On convergence of numerical methods for optimization problems governed by scalar hyperbolic conservation laws. In *XVI International Conference on Hyperbolic Problems: Theory, Numerics, Applications*, pages 691–706. Springer, 2016.
- [38] P. Schäfer Aguilar, J. M. Schmitt, S. Ulbrich, and M. Moos. On the numerical discretization of optimal control problems for conservation laws. *Control and Cybernetics*, 48, 2019.
- [39] S. Pfaff and S. Ulbrich. Optimal boundary control of nonlinear hyperbolic conservation laws with switched boundary data. *SIAM Journal on Control and Optimization*, 53(3):1250–1277, 2015.
- [40] S. Pfaffa and S. Ulbricha. Optimal control of nonlinear hyperbolic conservation laws by on/off-switching. 2016.

- [41] S. Hajian, M. Hintermüller, and S. Ulbrich. Total variation diminishing schemes in optimal control of scalar conservation laws. *IMA Journal of Numerical Analysis*, 39(1):105–140, 2019.
- [42] N. Zeng, L. Cen, Y. Xie, and S. Zhang. Nonlinear optimal control of cascaded irrigation canals with conservation law pdes. *Control Engineering Practice*, 100:104407, 2020.
- [43] D. Frenzel and J. Lang. A third-order weighted essentially non-oscillatory scheme in optimal control problems governed by nonlinear hyperbolic conservation laws. *Computational Optimization and Applications*, pages 1–20, 2021.
- [44] M. Hintermüller and N. Strogies. On the consistency of runge–kutta methods up to order three applied to the optimal control of scalar conservation laws. In *Numerical Analysis and Optimization*, pages 119–154. Springer, 2017.
- [45] M. J. Zahr and P. Persson. An adjoint method for a high-order discretization of deforming domain conservation laws for optimization of flow problems. *Journal of Computational Physics*, 326:516–543, 2016.
- [46] M. Herty, A. Kurganov, and D. Kurochkin. Numerical method for optimal control problems governed by nonlinear hyperbolic systems of pdes. *Communications in Mathematical Sciences*, 13(1):15–48, 2015.
- [47] A. Kurganov and E. Tadmor. Solution of two-dimensional riemann problems for gas dynamics without riemann problem solvers. *Numerical Methods for Partial Differential Equations: An International Journal*, 18(5):584–608, 2002.
- [48] Z. J. Wang, L. Zhang, and Y. Liu. Spectral (finite) volume method for conservation laws on unstructured grids iv: extension to two-dimensional systems. *Journal of Computational Physics*, 194(2):716–741, 2004.
- [49] S. Gottlieb, D. I. Ketcheson, and C. Shu. High order strong stability preserving time discretizations. *Journal of Scientific Computing*, 38(3):251–289, 2009.



- [50] S. Gottlieb, C. Shu, and E. Tadmor. Strong stability-preserving high-order time discretization methods. *SIAM Review*, 43(1):89–112, 2001.
- [51] J. S. Hesthaven. *Numerical methods for conservation laws: From analysis to algorithms*. SIAM, 2018.
- [52] A. Kurganov and C. Lin. On the reduction of numerical dissipation in central-upwind schemes. *Commun. Comput. Phys*, 2(1):141–163, 2007.
- [53] F. A. Dorini and M. C. C. Cunha. A finite volume method for the mean of the solution of the random transport equation. *Applied mathematics and computation*, 187(2):912–921, 2007.
- [54] R. J. LeVeque and D. L. George. High-resolution finite volume methods for the shallow water equations with bathymetry and dry states. In *Advanced numerical models for simulating tsunami waves and runup*, pages 43–73. World Scientific, 2008.
- [55] A. Kurganov, S. Noelle, and G. Petrova. Semidiscrete central-upwind schemes for hyperbolic conservation laws and hamilton–jacobi equations. *SIAM Journal on Scientific Computing*, 23(3):707–740, 2001.
- [56] C. Lattanzio and D. Serre. Convergence of a relaxation scheme for hyperbolic systems of conservation laws. *Numerische Mathematik*, 88(1):121–134, 2001.
- [57] R. Natalini. Convergence to equilibrium for the relaxation approximations of conservation laws. *Communications on pure and applied mathematics*, 49(8):795–823, 1996.
- [58] D. Aregba-Driollet and R. Natalini. Convergence of relaxation schemes for conservation laws. *Applicable Analysis*, 61(1-2):163–193, 1996.
- [59] D. Aregba-Driollet and V. Milišić. Kinetic approximation of a boundary value problem for conservation laws. *Numerische Mathematik*, 97(4):595–633, 2004.
- [60] R. Natalini and A. Terracina. Convergence of a relaxation approximation to a boundary value problem for conservation laws. *Communications in Partial Differential Equations*, 26(7-8):1235–1252, 2001.

- [61] R. Natalini. A discrete kinetic approximation of entropy solutions to multidimensional scalar conservation laws. *Journal of differential equations*, 148(2):292–317, 1998.
- [62] P. L. Bhatnagar, E. P. Gross, and M. Krook. A model for collision processes in gases. i. small amplitude processes in charged and neutral one-component systems. *Physical Review*, 94(3):511, 1954.
- [63] A. Vasseur. Convergence of a semi-discrete kinetic scheme for the system of isentropic gas dynamics with  $\gamma=3$ . *Indiana University mathematics journal*, pages 347–364, 1999.
- [64] S. Nørgaard, O. Sigmund, and B. Lazarov. Topology optimization of unsteady flow problems using the lattice boltzmann method. *Journal of Computational Physics*, 307:291–307, 2016.
- [65] E. M. Yohana and M. K. Banda. High-order relaxation approaches for adjoint-based optimal control problems governed by nonlinear hyperbolic systems of conservation laws. *Journal of Numerical Mathematics*, 24(1):45–71, 2016.
- [66] S. Steffensen, M. Herty, and L. Pareschi. Numerical methods for the optimal control of scalar conservation laws. In *IFIP Conference on System Modeling and Optimization*, pages 136–144. Springer, 2011.
- [67] M. K. Banda and M. Herty. Adjoint imex-based schemes for control problems governed by hyperbolic conservation laws. *Computational optimization and applications*, 51(2):909–930, 2012.
- [68] X. Li, L. Fang, and Y. Peng. Airfoil design optimization based on lattice boltzmann method and adjoint approach. *Applied Mathematics & Mechanics*, 39(6):891–904, 2018.
- [69] M. Herty, L. Salhi, and M. Seaid. A relaxation algorithm for optimal control problems governed by two-dimensional conservation laws. In *International Conference on Computational Science*, pages 122–135. Springer, 2020.

- [70] J. M. T. Ngnotchouye, M. Herty, S. Steffensen, and M. K. Banda. Relaxation approaches to the optimal control of the euler equations. *Computational & Applied Mathematics*, 30(2):399–425, 2011.
- [71] M. Herty and B. Piccoli. A numerical method for the computation of tangent vectors to  $2 \times 2$  hyperbolic systems of conservation laws. *Communications in Mathematical Sciences*, 14(3):683–704, 2016.
- [72] G. Albi, M. Herty, and L. Pareschi. Linear multistep methods for optimal control problems and applications to hyperbolic relaxation systems. *Applied Mathematics and Computation*, 354:460–477, 2019.
- [73] S. N. Kružkov. First order quasilinear equations in several independent variables. *Mathematics of the USSR-Sbornik*, 10(2):217, 1970.
- [74] L. C. Evans. *Partial differential equations*, volume 19. American Mathematical Soc., 2010.
- [75] T. R. Hagen, M. O. Henriksen, J. M. Hjelmervik, and K. Lie. How to solve systems of conservation laws numerically using the graphics processor as a high-performance computational engine. In *Geometric Modelling, Numerical Simulation, and Optimization*, pages 211–264. Springer, 2007.
- [76] F. Bouchut. *Nonlinear stability of finite volume methods for hyperbolic conservation laws: And Well-Balanced schemes for sources*. Springer Science & Business Media, 2004.
- [77] P. D. Lax. *Hyperbolic partial differential equations*, volume 14. American Mathematical Soc., 2006.
- [78] A. Bressan. *Hyperbolic systems of conservation laws: the one-dimensional Cauchy problem*, volume 20. Oxford University Press on Demand, 2000.
- [79] A. Bressan, G. Crasta, and B. Piccoli. *Well-posedness of the Cauchy problem for  $n \times n$  systems of conservation laws*, volume 694. American Mathematical Soc., 2000.

- [80] A. Bressan. Hyperbolic systems of conservation laws. *Revista Matemática Complutense*, 12(1):135–200, 1999.
- [81] S. Mungkasi, B. W. Harini, and L. H. Wiryanto. Jin–xin relaxation method used to solve the one-dimensional inviscid burgers equation. In *Journal of Physics: Conference Series*, volume 856, page 012007. IOP Publishing, 2017.
- [82] M. K. Banda and M. Seaid. Higher-order relaxation schemes for hyperbolic systems of conservation laws. *Journal of Numerical Mathematics jnma*, 13(3):171–196, 2005.
- [83] R. J. LeVeque et al. *Finite volume methods for hyperbolic problems*, volume 31. Cambridge university press, 2002.
- [84] A. Kurganov and E. Tadmor. New high-resolution central schemes for nonlinear conservation laws and convection–diffusion equations. *Journal of Computational Physics*, 160(1):241–282, 2000.
- [85] E. Godlewski and P. Raviart. *Numerical approximation of hyperbolic systems of conservation laws*, volume 118. Springer Science & Business Media, 2013.
- [86] J. S. Hesthaven. *Numerical methods for conservation laws: From analysis to algorithms*. SIAM, 2017.
- [87] A. Kurganov and G. Petrova. Central-upwind schemes on triangular grids for hyperbolic systems of conservation laws. *Numerical Methods for Partial Differential Equations: An International Journal*, 21(3):536–552, 2005.
- [88] S. Bianchini, E. Marconi, et al. On the concentration of entropy for scalar conservation laws. 2016.
- [89] A. Bressan. One dimensional hyperbolic systems of conservation laws. In *Current Developments in Mathematics, 2002*, pages 1–37. International Press of Boston, 2003.
- [90] S. Bianchini and A. Bressan. Vanishing viscosity solutions of nonlinear hyperbolic systems. *Annals of Mathematics*, pages 223–342, 2005.

- [91] R. Pouso and J. Rodriguez-Lopez. Existence and uniqueness of solutions for systems of discontinuous differential equations under localized bressan–shen transversality conditions. *Journal of Mathematical Analysis and Applications*, 492(1):124425, 2020.
- [92] J. Liu, Z. Ma, and Z. Zhou. Explicit and implicit tvd schemes for conservation laws with caputo derivatives. *Journal of Scientific Computing*, 72(1):291–313, 2017.
- [93] F. Benkhaldoun and M. Seaïd. A simple finite volume method for the shallow water equations. *Journal of Computational and Applied Mathematics*, 234(1):58–72, 2010.
- [94] W. F. Ames. *Numerical methods for partial differential equations*. Academic press, 2014.
- [95] M. M. Babatin and Y. H. Zahran. Adaptive multi-resolution central-upwind schemes for systems of conservation laws. *International Journal of Computational Fluid Dynamics*, 23(10):723–735, 2009.
- [96] S. Jin. Runge-kutta methods for hyperbolic conservation laws with stiff relaxation terms. *Journal of Computational Physics*, 122(1):51–67, 1995.
- [97] P. K. Sweby. High resolution schemes using flux limiters for hyperbolic conservation laws. *SIAM journal on numerical analysis*, 21(5):995–1011, 1984.
- [98] D. Luo, W. Huang, and J. Qiu. A quasi-lagrangian moving mesh discontinuous galerkin method for hyperbolic conservation laws. *Journal of Computational Physics*, 396:544–578, 2019.
- [99] X. Meng, C. Shu, Q. Zhang, and B. Wu. Superconvergence of discontinuous galerkin methods for scalar nonlinear conservation laws in one space dimension. *SIAM Journal on Numerical Analysis*, 50(5):2336–2356, 2012.
- [100] C. Shu. Discontinuous galerkin methods: general approach and stability. *Numerical solutions of partial differential equations*, 201, 2009.

- [101] T. Chen and C. Shu. Review of entropy stable discontinuous galerkin methods for systems of conservation laws on unstructured simplex meshes. *CSIAM Transactions on Applied Mathematics*, 1(1):1–52, 2020.
- [102] B. Cockburn, G. E. Karniadakis, and C. Shu. *Discontinuous Galerkin methods: theory, computation and applications*, volume 11. Springer Science & Business Media, 2012.
- [103] J. Chen and Z. Shi. Application of a fourth-order relaxation scheme to hyperbolic systems of conservation laws. *Acta Mechanica Sinica*, 22(1):84–92, 2006.
- [104] H. J. Schroll. Relaxed high resolution schemes for hyperbolic conservation laws. *Journal of Scientific Computing*, 21(2):251–279, 2004.
- [105] G. Naldi and L. Pareschi. Numerical schemes for hyperbolic systems of conservation laws with stiff diffusive relaxation. *SIAM Journal on Numerical Analysis*, 37(4):1246–1270, 2000.
- [106] S. Ulbrich. A sensitivity and adjoint calculus for discontinuous solutions of hyperbolic conservation laws with source terms. *SIAM journal on control and optimization*, 41(3):740–797, 2002.
- [107] C. M. Dafermos and C. M. Dafermos. *Hyperbolic conservation laws in continuum physics*, volume 3. Springer, 2005.
- [108] C. Castro, F. Palacios, and E. Zuazua. Optimal control and vanishing viscosity for the burgers equation. In *Integral Methods in Science and Engineering, Volume 2*, pages 65–90. Springer, 2010.
- [109] R. Lecaros and E. Zuazua. Tracking control of 1d scalar conservation laws in the presence of shocks. *Trends in contemporary mathematics*, pages 195–219, 2014.
- [110] D. Frenzel and J. Lang. A third-order weighted essentially non-oscillatory scheme in optimal control problems governed by nonlinear hyperbolic conservation laws. *arXiv preprint arXiv:2009.12392*, 2020.

- [111] S. Pfaff and S. Ulbrich. Optimal control of scalar conservation laws by on/off-switching. *Optimization Methods and Software*, 32(4):904–939, 2017.
- [112] S. Ulbrich. Adjoint-based derivative computations for the optimal control of discontinuous solutions of hyperbolic conservation laws. *Systems & Control Letters*, 48(3-4):313–328, 2003.
- [113] G. Albi, M. Herty, and L. Pareschi. Relaxation approximation of optimal control problems and applications to traffic flow models. In *AIP Conference Proceedings*, volume 1975, page 020001. AIP Publishing LLC, 2018.
- [114] L. Armijo. Minimization of functions having lipschitz continuous first partial derivatives. *Pacific Journal of Mathematics*, 16(1):1–3, 1966.
- [115] C. D’Apice, R. Manzo, and B. Piccoli. Numerical schemes for the optimal input flow of a supply chain. *SIAM Journal on Numerical Analysis*, 51(5):2634–2650, 2013.
- [116] P. Colella. Multidimensional upwind methods for hyperbolic conservation laws. *Journal of Computational Physics*, 87(1):171–200, 1990.
- [117] A. Bressan, G. G. Chen, M. Lewicka, and D. Wang. *Nonlinear conservation laws and applications*. Springer, 2011.
- [118] Y. Zheng. *Systems of conservation laws: two-dimensional Riemann problems*, volume 38. Springer Science & Business Media, 2012.
- [119] C. M. Dafermos. Entropy and the stability of classical solutions of hyperbolic systems of conservation laws. In *Recent mathematical methods in nonlinear wave propagation*, pages 48–69. Springer, 1996.
- [120] A. Keimer, L. Pflug, and M. Spinola. Existence, uniqueness and regularity of multi-dimensional nonlocal balance laws with damping. *Journal of Mathematical Analysis and Applications*, 466(1):18–55, 2018.

- [121] A. Ben-Israel and T. N. E. Greville. *Generalized inverses: theory and applications*, volume 15. Springer Science & Business Media, 2003.
- [122] A. Kurganov, Z. Qu, O. S. Rozanova, and T. Wu. Adaptive moving mesh central-upwind schemes for hyperbolic system of pdes: Applications to compressible euler equations and granular hydrodynamics. *Communications on Applied Mathematics and Computation*, pages 1–35, 2020.
- [123] A. Kurganov, Y. Liu, and V. Zeitlin. Numerical dissipation switch for two-dimensional central-upwind schemes. *Submitted to ESAIM Math. Model. Numer. Anal*, 2020.
- [124] J. Zhu and J. Qiu. A new type of finite volume weno schemes for hyperbolic conservation laws. *Journal of Scientific Computing*, 73(2):1338–1359, 2017.
- [125] J. Zhu and C. Shu. A new type of multi-resolution weno schemes with increasingly higher order of accuracy. *Journal of Computational Physics*, 375:659–683, 2018.
- [126] D. Levy, G. Puppo, and G. Russo. Compact central weno schemes for multidimensional conservation laws. *SIAM Journal on Scientific Computing*, 22(2):656–672, 2000.
- [127] H. Shen and M. Parsani. A class of high-order weighted compact central schemes for solving hyperbolic conservation laws. *arXiv preprint arXiv:2104.04347*, 2021.
- [128] M. Seaid. Multidimensional relaxation approximations for hyperbolic systems of conservation laws. *Journal of Computational Mathematics*, pages 440–457, 2007.
- [129] C. Castro, F. Palacios, and E. Zuazua. An alternating descent method for the optimal control of the inviscid burgers equation in the presence of shocks. *Mathematical Models and Methods in Applied Sciences*, 18(03):369–416, 2008.
- [130] Y. Gan, A. Xu, G. Zhang, X. Yu, and Y. Li. Two-dimensional lattice boltzmann model for compressible flows with high mach number. *Physica A: Statistical Mechanics and its Applications*, 387(8-9):1721–1732, 2008.



[131] S. S. Chikatamarla and I. V. Karlin. Lattices for the lattice boltzmann method. *Physical Review E*, 79(4):046701, 2009.

[132] R. J. LeVeque. *Numerical methods for conservation laws*, volume 3. Springer, 1992.

**LANGUAGE EDITING CERTIFICATE**



ZANEZ EXPERT EDITING

**Registered with the South African Translators' Institutes (SATI)  
Reference number 1000363**

**SACE REGISTERED**

03 August 2021

***PhD Thesis:* OPTIMAL CONTROL PROBLEMS CONSTRAINED BY  
HYPERBOLIC CONSERVATION LAWS**

This serves to confirm that I edited substantively the above document. I returned the document to the author with some tracked changes intended to correct errors and clarify meaning. It was the author's responsibility to attend to these changes.

Yours faithfully



Dr. K. Zano

Ph.D. in English

[kufazano@gmail.com](mailto:kufazano@gmail.com)/[kufazano@yahoo.com](mailto:kufazano@yahoo.com)

0631434276