

AMERICAN UNIVERSITY OF BEIRUT

ON RECOVERING THE INITIAL STATE OF
THE TRANSPORT EQUATION

by

YOUMNA NADIM LAYOUN

A thesis

submitted in partial fulfillment of the requirements
for the degree of Master of Science
to the Department of Mathematics
of Faculty of Arts and Sciences
at the American University of Beirut

Beirut, Lebanon
September 2022

AMERICAN UNIVERSITY OF BEIRUT

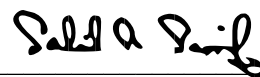
ON RECOVERING THE INITIAL STATE OF THE
TRANSPORT EQUATION

by

YOUMNA NADIM LAYOUN

Approved by:

Dr. Nabil Nassif, Professor
Mathematics



Advisor

Dr. Sophie Moufawad, Assistant Professor
Mathematics



Co-Advisor

Dr. Ahmad Sabra, Assistant Professor
Mathematics



Member of Committee

Dr. Faouzi Triki, Professor
Grenobles Alpes University, France



Member of Committee

Date of thesis defense: September 5, 2022

ACKNOWLEDGEMENTS

First of all, I would like to thank my family and friends who have been my support system throughout this whole process. The presence of some people in my life definitely made it easier and I cannot imagine getting here without them.

I also thank my committee members. My advisor Dr. Nassif and co-advisor Dr. Moufawad for their help and guidance, Dr. Sabra for his remarks and comments and Dr. Triki for proposing the topic.

Last but definitely not least, I thank God for being the light in my darkest days.

ABSTRACT OF THE THESIS OF

Younna Nadim Layoun for Master of Science
Major: Mathematics

Title: On recovering the initial state of the transport equation

Widely used and studied, the transport equation is a partial differential equation that describes how a mass is transported (or translated) through time and space. The goal of this thesis is to recover the initial state of the transport equation (at time $t = 0$) given the measurement of the solution at some end time T .

To this end, we carry a thorough study on the direct problem, both theoretically and numerically, for the linear and non-linear transport equations. This helped us then develop robust numerical schemes to accurately approximate the exact solution of the direct transport problem. In turn these schemes are used to solve the inverse problem and recover the initial state through optimization algorithms provided by the MATLAB platform.

TABLE OF CONTENTS

Acknowledgements	1
Abstract	2
1 Introduction	7
1.1 Linear Transport Equation	7
1.2 Non-linear Transport Equation	8
1.3 Overview of this Thesis	10
2 Theoretical Results for the Transport Equation	11
2.1 Solution of the linear transport equation depending on the initial data	11
2.1.1 Smooth initial data	11
2.1.2 L^p initial data	13
2.2 Non linear case: Burger's equation	15
3 Numerical Approximation of Solutions	25
3.1 Finite Difference Method for Linear Transport Equation	25
3.1.1 Deriving The Schemes	25
3.1.2 Stability Analysis	28
3.2 Finite Volume Method for Non-Linear equation	29
3.2.1 Godunov scheme	30
3.2.2 Roe scheme	32
3.2.3 Rusanov scheme	32
3.2.4 Convergence Analysis (Lax-Wendroff Theorem)	33
4 Implementations: Direct Solvers	37
4.1 Linear Transport Equation	37
4.1.1 Constant velocity	37
4.1.2 Non constant velocity	39
4.2 Non Linear Transport equation: Burgers' equation	39
4.2.1 Case 1: $u_l > u_r$	40
4.2.2 Case 2: $u_l < u_r$	41

5	On the Inverse Problem	43
5.1	Formulation	43
5.2	Implementation	44
5.2.1	Linear transport equation with constant velocity	44
5.2.2	Non linear transport equation	49
5.3	Concluding Remarks on the Inverse Problem	53
6	Conclusion	54
	Bibliography	55

ILLUSTRATIONS

2.1	Characteristics lines for Burger's equation with discontinuous initial data	16
2.2	shock wave as a line in (x,t) plane	20
2.3	rarefaction wave as a line in (x,t) plane	21
4.1	Solutions using finite difference schemes with $a = 1$ and $\frac{\Delta t}{\Delta x} = 0.9$	38
4.2	Finite difference schemes for other cases	39
4.3	Finite difference upwind scheme solution for non constant velocity, $\Delta x = 1/64$, end time $T = 5$	40
4.4	Finite volume schemes for $u_l > u_r$	41
4.5	Finite volume schemes for $u_l < u_r$	42
5.1	u_0 exact and approximated using "interior-point" and $tol = 10^{-6}$.	46
5.2	u_0 exact and approximated using "sqp" and $tol = 10^{-6}$	48
5.3	u_0 exact and approximated using "interior-point" and $tol = 10^{-6}$.	50
5.4	u_0 exact and approximated using "sqp" and $tol = 10^{-6}$	52

TABLES

4.1	Relative error between exact and approximated solutions for end time $T=5$	38
4.2	Relative error between exact and approximated solutions for end time $T=10$	38
4.3	Relative error between exact and approximated solutions for end time $T=5$	40
4.4	Relative error between exact and approximated solutions for end time $T=10$	41
4.5	Relative error between exact and approximated solutions for end time $T=5$	41
4.6	Relative error between exact and approximated solutions for end time $T=10$	42
5.1	Relative error between exact and approximated u_0 and value of objective function using “interior-point” algorithm and unconstrained solution	45
5.2	Relative error between exact and approximated u_0 and value of objective function using “interior-point” algorithm and constrained solution	45
5.3	Relative error between exact and approximated u_0 and value of objective function using “sqp” algorithm and unconstrained solution	47
5.4	Relative error between exact and approximated u_0 and value of objective function using “sqp” algorithm and constrained solution	47
5.5	Relative error between exact and approximated u_0 and value of objective function using “interior-point” algorithm and unconstrained solution	49
5.6	Relative error between exact and approximated u_0 and value of objective function using “interior-point” algorithm and constrained solution	49
5.7	Relative error between exact and approximated u_0 and value of objective function using “sqp” algorithm and unconstrained solution	51
5.8	Relative error between exact and approximated u_0 and value of objective function using “sqp” algorithm and constrained solution	51

CHAPTER 1

INTRODUCTION

The transport equation is a partial differential equation which is generally used to describe how a quantity is transported through space and time. It is the basis for the formulation of many physical problems which we will mention in the following sections.

1.1 Linear Transport Equation

Starting with the linear transport equation we focus in this thesis on the initial value problem:

$$\begin{cases} u_t + \vec{a}(x) \cdot \nabla_x u = 0, & x \in \mathbb{R}^d, t \in \mathbb{R}^+ \\ u(x, 0) = u_0(x), & x \in \mathbb{R}^d \end{cases} \quad (1.1)$$

We are interested in finding the solution $u : \mathbb{R}^d \times [0, T] \rightarrow \mathbb{R}$, taking $\vec{a}(x)$ to be $C^1(\mathbb{R}^d)$. In this thesis, we will only deal with the one dimensional case $d = 1$. Also, $a(x)$ could depend on t which we will see in the examples.

The method of characteristics is widely practised in the field of PDEs. However, for the transport equation, it is only applicable with smooth initial data.

One important theory concerning the transport equation is the theory of renormalization developed by DiPerna, R. J., and Lions, P.-L in [1] in 1989. It revolves around proving the existence and uniqueness of a solution to the transport equation (1.1) with initial data in $L^p(\mathbb{R}^d)$ for $1 \leq p < \infty$ and the velocity field $a(x) \in C^1(\mathbb{R}^d) \cap W^{1,\infty}(\mathbb{R}^d)$.

This theory was extended by Ambrosio, L. in [2] in 2004 to BV vector fields with initial data in $L^p(\mathbb{R}^d)$, where existence and uniqueness of a solution are also proved.

Since the right hand side is zero, the equation in (1.1) is homogeneous. This simple form of the transport equation enables us to build models to describe interesting physical problems discussed in [3]. We state two of these in what follows.

Demographic model/Cell renewal

A linear transport equation can be used to represent the process of cell renewal. The main idea is that the death of a cell gives birth to two new cells.

Let $u(x, t)$ be the density of people aged $x \geq 0$ at a given time t . Fix some age x^* and let

1. $d(x)$ be the mortality rate defined as $d(x) = \mathcal{I}_{x > x^*}$
2. $b(x)$ be the fertility rate defined as $b(x) = 2\mathcal{I}_{x < x^*}$ since cell mitosis results in the birth of two identical cells.

Then the transport problem can be written as such

$$\begin{cases} \frac{\partial}{\partial t} u(x, t) + \frac{\partial}{\partial x} u(x, t) + d(x)u(x, t) = 0 \\ u(x = 0, t) = \int_{x=0}^{\infty} b(x)u(x, t)dx \end{cases} \quad (1.2)$$

Multiple TCP Connections through a Buffer Implementing RED

Transmission control protocol (TCP) is a communication protocol that facilitates internet transmission of messages called windows which are sent by users to a central server. These windows have sizes which can grow continuously creating a queue and eventually a bottleneck. Baccelli, McDonald and Reynier [4] studied the congestion of these queues using the random early detection (RED) management scheme. When the length of the queue reaches a threshold, the RED scheme drops incoming windows with an increasing probability as the queue length is increasing.

Let $u(x, t)$ be the density of windows of size x and $q(t)$ the length of the queue at time t , then the limiting system is described by a linear transport equation which we present below,

$$\begin{cases} \frac{\partial}{\partial t} u(x, t) + (1 - k(t)) \frac{\partial}{\partial x} u(x, t) = k(t)(4u(2x, t) - u(x, t)) \\ k(t) = F(q(t)) \\ \frac{\partial}{\partial t} q(t) = \int_{x=0}^{\infty} xu(x, t)dx \end{cases} \quad (1.3)$$

where F is an increasing function with $F(0) = 0$ and $F(q_{max}) = 1$.

1.2 Non-linear Transport Equation

Non-linearity could be introduced to the transport equation in (1.1). One simple way is to replace the velocity $a(x, t)$ by a velocity that depends on the solution of the equation. For example, if $a = u$ we get Burgers' equation

$$u_t + uu_x = 0 \quad (1.4)$$

which we will study in the coming chapters using the book by Godlewski and Raviart [5] as a main reference.

In fact, Burgers' equation is part of a bigger family of equations called scalar conservation law which takes the form

$$u_t + (f(u))_x = 0 \quad (1.5)$$

where f is called flux.

Therefore Burger's equation can be written in the following conservative form

$$u_t + \left(\frac{u^2}{2}\right)_x = 0 \quad (1.6)$$

The scalar conservation law is used to describe multiple problems, let's consider the examples below ([5]).

Traffic flow model

The Lighthill-Whitham-Richards (LWR) model is one of the very well-known models to describe traffic flow. Let u be the density of cars on a highway, u_{max} the maximum density and v_{max} the speed limit. Then the model is written as

$$u_t + [v_{max}u(1 - \frac{u}{u_{max}})]_x = 0 \quad (1.7)$$

Note that the velocity is inversely proportional to the density of cars, that is when the numbers of cars increases, their speed decreases.

Gas dynamics equations

Gas dynamics equations describe fluids in rapid flow. The system, often used in plasma physics and astrophysics, consists of three equations, namely the laws of conservation of mass, momentum and total energy of the fluid.

Let ρ be the density of the fluid, u the velocity, p the pressure, ϵ the internal energy and $e = \epsilon + \frac{|u|^2}{2}$ the total energy. Then we have the following system

$$\begin{cases} \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho u) = 0 \\ \frac{\partial}{\partial t}(\rho u) + \nabla \cdot (\rho u \otimes u + pI) = 0 \\ \frac{\partial}{\partial t}(\rho e) + \nabla \cdot ((\rho e + p)u) = 0 \end{cases} \quad (1.8)$$

where p does not depend on u . For a polytropic ideal gas, $p = (\gamma - 1)\rho\epsilon$, $\gamma > 1$.

1.3 Overview of this Thesis

In chapter 2, we present the theoretical frame of the transport equation starting with the linear case. We then proceed to the non-linear case, specifically the conservation law and analyse the Riemann problem which we will define later on.

In chapter 3, we present different schemes to approximate the solutions of the transport equation and conduct some stability analysis. For the linear case we use finite difference schemes, as for the conservation law (non-linear equation) we used finite volume schemes.

Implementations of the schemes are done in chapter 4 using different mesh sizes and end times. Alongside the graphs, we present tables showing the relative errors between the approximated and exact solutions in order to compare the accuracy of the schemes.

In chapter 5, we formulate the inverse problem which aims at recovering the initial data from the data given at some time T . We will implement the inverse problem using the built-in MATLAB function `fmincon` and present the results in tables and graphs.

Concluding remarks are given in the last chapter.

CHAPTER 2

THEORETICAL RESULTS FOR THE TRANSPORT EQUATION

In this chapter we discuss the theoretical existence of solutions $u : \Omega \times [0, T) \rightarrow \mathbb{R}$ to the homogeneous transport equation $u_t + a(x)u_x = 0$ depending on the initial state and taking $a(x)$ to be $C^1(\mathbb{R})$. Note that the velocity a is only dependent on the space domain here, but the results could be extended to the case where it also depends on the time t .

We then move to another form of the transport equation, the conservation law, thus changing the nature of the velocity in the original transport equation, which becomes dependent on the solution itself.

2.1 Solution of the linear transport equation depending on the initial data

Relying on the work of Diperna and Lions [1], we divide our work into two cases; the case where the initial condition is smooth and the other more general case where u_0 is $L^p(\mathbb{R})$. We will see that in the first scenario, a classical solution exists whereas in the latter one, we will seek weak solutions.

2.1.1 Smooth initial data

Given the following Cauchy problem

$$\begin{cases} u_t + a(x)u_x = 0 & x \in \mathbb{R}, t > 0 \\ u(x, 0) = u_0(x) \end{cases} \quad (2.1)$$

where $u_0 \in C^1(\mathbb{R})$.

We will derive the solution $u(x, t)$ to problem (2.1) using the method of characteristics below.

Method of characteristics

Let $x(t)$ be a given curve along which the solution u is constant, in other words, $\frac{d}{dt}u(x(t), t) = 0$. This implies

$$u_x(x(t), t)x'(t) + u_t(x(t), t) = 0 \quad (2.2)$$

Comparing equation (2.2) with our initial equation in (2.1) we get

$$\begin{cases} x'(t) = a(x(t)) \\ x(0) = x_0 \end{cases} \quad (2.3)$$

The problem reduces to solving the above ordinary differential equation with initial data. We know that a solution for (2.3) exists and is unique provided a is Lipschitz.

Note that since u is constant along characteristics $x(t)$, we have the following identity

$$u(x(t), t) = u(x(0), 0) = u_0(x_0)$$

Therefore, if the characteristics lines cover the whole plane, a solution for (2.1) is found, namely $u(x, t) = u_0(x_0)$.

Theorem 1. *Let $x(t)$ be the characteristics functions satisfying (2.3) and suppose that given different values of x_0 , they cover the whole plane (x, t) .*

Then the solution to the initial value problem (2.1) with smooth initial data (in $C^1(\mathbb{R})$) is given by

$$u(x, t) = u_0(x_0)$$

Note that the regularity of the solution is that of the initial data, which will be discussed in 1.

Remark 1. *If a is constant, the solution will be $x(t) = at + x_0 \implies x_0 = x(t) - at$. Hence*

$$u(x, t) = u_0(x_0) = u_0(x - at)$$

Similarly, for the case where a is not constant we solve (2.1). Below are two examples.

Example 1.

$$\begin{cases} u_t + tu_x = 0 \\ u(x, 0) = u_0(x) \end{cases}$$

$$x'(t) = t \text{ then } x(t) = \frac{t^2}{2} + x_0 \implies x_0 = x(t) - \frac{t^2}{2} \implies u(x, t) = u_0(x - \frac{t^2}{2}).$$

Example 2.

$$\begin{cases} u_t + (t+x)u_x = 0 \\ u(x, 0) = u_0(x) \end{cases}$$

$$x'(t) = t+x \text{ then } x(t) = (x_0 - 1)e^t - t + 1 \implies x_0 = (x+t-1)e^{-t} + 1 \implies u(x, t) = u_0((x+t-1)e^{-t} + 1).$$

Remark 2. One main difference between the transport equation and other equations such as the heat equation or more generally the parabolic type family is that if the initial condition is compactly supported (decays to 0 at infinity) then transport occurs at finite speed. In other words, the solution to the transport equation satisfies some energy bound. We discuss this feature in the below lemma.

Lemma 1. Let u be a solution to (2.1) and suppose $u_0 \in C_c^1(\mathbb{R})$. Then u satisfies the following energy bound

$$\int_{\mathbb{R}} u^2(x, t) dx \leq e^{\|a\|_{C^1} t} \int_{\mathbb{R}} u_0^2(x) dx \quad (2.4)$$

for all $t > 0$.

Proof. We proceed in these steps

$$\begin{aligned} u_t + a(x, t)u_x &= 0 \\ uu_t + a(x, t)uu_x &= 0 \text{ (multiply by } u\text{)} \\ \left(\frac{u^2}{2}\right)_t + a\left(\frac{u^2}{2}\right)_x &= 0 \\ \left(\frac{u^2}{2}\right)_t + \left(a\frac{u^2}{2}\right)_x &= a_x \frac{u^2}{2} \text{ (applying the product rule)} \\ \int_{\mathbb{R}} \left(\frac{u^2}{2}\right)_t dx + \int_{\mathbb{R}} \left(a\frac{u^2}{2}\right)_x dx &= \int_{\mathbb{R}} a_x \frac{u^2}{2} dx \text{ (integrate over space)} \\ \frac{d}{dt} \int_{\mathbb{R}} \frac{u^2}{2} dx + a \frac{u^2}{2} \Big|_{-\infty}^{\infty} &= \int_{\mathbb{R}} a_x \frac{u^2}{2} dx \text{ (} u \text{ decays to 0 at } \infty \text{ since } u_0 \in C_c^1\text{)} \\ \frac{d}{dt} \int_{\mathbb{R}} \frac{u^2}{2} dx &\leq \|a\|_{C^1} \int_{\mathbb{R}} \frac{u^2}{2} dx \text{ (by regularity of } a\text{)} \\ \int_{\mathbb{R}} \frac{u^2}{2} dx &\leq e^{\int_0^t \|a\|_{C^1}(s) ds} \int_{\mathbb{R}} \frac{u_0^2}{2} dx \text{ (applying Gronwall's inequality)} \\ \int_{\mathbb{R}} u^2 dx &\leq e^{\|a\|_{C^1} t} \int_{\mathbb{R}} u_0^2 dx \end{aligned}$$

□

2.1.2 L^p initial data

We now move to a more general case of initial data. For this purpose, we take $u_0 \in L^p(\mathbb{R})$, $1 \leq p < \infty$. Diperna and Lions developed a theory of renormalization which proves the existence and uniqueness of a solution for this case in [1].

Variational formulation

Given the initial value problem

$$\begin{cases} u_t + a(x)u_x = 0 & x \in \mathbb{R}, t > 0 \\ u(x, 0) = u_0(x) \end{cases} \quad (2.5)$$

where $u_0 \in L^p(\mathbb{R})$.

Let $\varphi(x, t)$ be a test function in $C_c^1(\mathbb{R}, [0, T])$.

Multiply the equation in (2.5) by φ and integrate.

$$\begin{aligned} \int_0^T \int_{\mathbb{R}} \varphi u_t + \varphi a u_x dx dt &= 0 \\ \int_{\mathbb{R}} \int_0^T \varphi u_t dt dx + \int_0^T \int_{\mathbb{R}} \varphi a u_x dx dt &= 0 \\ \int_{\mathbb{R}} \varphi u|_0^T dx - \int_0^T \int_{\mathbb{R}} \varphi_t u dx dt + \int_0^T a \varphi u|_{\partial \mathbb{R}} dt - \int_0^T \int_{\mathbb{R}} (a \varphi)_x u dx dt &= 0 \\ - \int_0^T \int_{\mathbb{R}} \varphi_t u + (a \varphi)_x u dx dt &= - \int_{\mathbb{R}} \varphi(x, T) u(x, T) dx + \int_{\mathbb{R}} \varphi(x, 0) u(x, 0) dx \end{aligned}$$

Since φ is compactly supported in $\mathbb{R} \times [0, T)$ i.e. φ vanishes outside compact subsets of the domain, we get

$$\int_0^T \int_{\mathbb{R}} \varphi_t u + (a \varphi)_x u dx dt = - \int_{\mathbb{R}} \varphi(x, 0) u_0(x) dx \quad (2.6)$$

This last equation can be generalized to higher dimensions and written in operator form as the following

$$\int_0^T \int_{\mathbb{R}} u L[\varphi] dx dt = \int_{\mathbb{R}} \varphi(x, 0) u_0(x) dx \quad (2.7)$$

where $L[\varphi] := -\partial_t \varphi - \text{div}_x(a \varphi)$.

Weak Solution

As a result of the above formulation we can now define weak solutions.

Definition 1. A function $u(x, t)$ is a weak solution for the transport equation in (2.5) with initial data $u_0 \in L^p(\mathbb{R})$, $1 \leq p \leq \infty$, if $u \in L^\infty(L^p(\mathbb{R}), [0, T])$ and satisfies equation (2.6) for all $\varphi \in C_c^1(\mathbb{R}, [0, T])$.

Remark 3. An important observation is that if the solution u is smooth (in $C^1(\mathbb{R}, [0, T])$) then u satisfies the classical transport equation in (2.5). Thus, the family of weak solutions contains, but is not restricted to, classical solutions.

Existence and uniqueness theorems

We now state the theorems from [1] of existence and uniqueness of weak solutions to the transport equation with general initial data in $L^p(\mathbb{R})$.

Theorem 2. (*Existence*) Let $u_0 \in L^p(\mathbb{R})$, $1 \leq p \leq \infty$, then the transport equation (2.5) admits a global weak solution

$$u \in C(L^p(\mathbb{R}), [0, \infty)) \text{ when } 1 \leq p < \infty$$

and

$$u \in C(L^1_{loc}(\mathbb{R}), [0, \infty)) \text{ when } p = \infty$$

Theorem 3. (*Uniqueness*) Let $u_0 \in L^p(\mathbb{R})$, $1 \leq p < \infty$ then the transport equation (2.5) admits a unique weak solution $u \in C(L^p(\mathbb{R}), [0, \infty))$.

The proofs of these theorems require a long analysis using the theory of renormalization developed by Diperna and Lions [1].

2.2 Non linear case: Burger's equation

We now move to another form of the transport equation, the scalar conservation law, where the velocity a depends on the solution itself. We will discuss in detail a special case, namely, Burger's equation. The analysis is based on [5], [6] and [7].

Define the initial value problem

$$\begin{cases} u_t + (f(u))_x = 0, & x \in \mathbb{R}, t > 0 \\ u(x, 0) = u_0(x) \end{cases} \quad (2.8)$$

where $u : \mathbb{R} \times [0, T) \rightarrow \mathbb{R}$ and f is a sufficiently smooth function (at least C^2).

We note that the initial data u_0 could be smooth, continuous or even piecewise continuous which we will consider in the coming analysis.

Remark 4. The equation in (2.8) can be re-written as

$$u_t + f'(u)u_x = 0 \quad (2.9)$$

If $f(u) = \frac{u^2}{2}$ or equivalently $f'(u) = u$ then the equation is called Burgers' equation.

In the coming analysis we consider the Riemann initial value problem which is defined as follows.

Definition 2. The Riemann problem consists in finding the solution to the conservation law with piecewise constant initial data that admits a single discontinuity,

$$\begin{cases} u_t + f(u)_x = 0 \\ u(x, 0) = \begin{cases} u_l & \text{if } x \leq 0 \\ u_r & \text{if } x > 0 \end{cases} \end{cases} \quad (2.10)$$

where both u_l and u_r are constants.

We try to apply the method of characteristics as we did in the previous section.

Method of characteristics

As in the previous section, the characteristics solve the ODE problem

$$\begin{cases} x'(t) = f'(u(x(t), t)) \\ x(0) = x_0 \end{cases} \quad (2.11)$$

and since u is constant along $x(t)$, we get

$$x'(t) = f'(u(x(t), t)) = f'(u(x(0), 0)) = f'(u_0(x_0))$$

hence for Burger's equation we get

$$x(t) = \begin{cases} u_l t + x_0 & \text{if } x_0 \leq 0 \\ u_r t + x_0 & \text{if } x_0 > 0 \end{cases} \quad (2.12)$$

We thus need to divide our analysis to 2 cases:

case 1: $u_l > u_r$ and case 2: $u_r > u_l$

Below are the graphs of the characteristics in both cases.

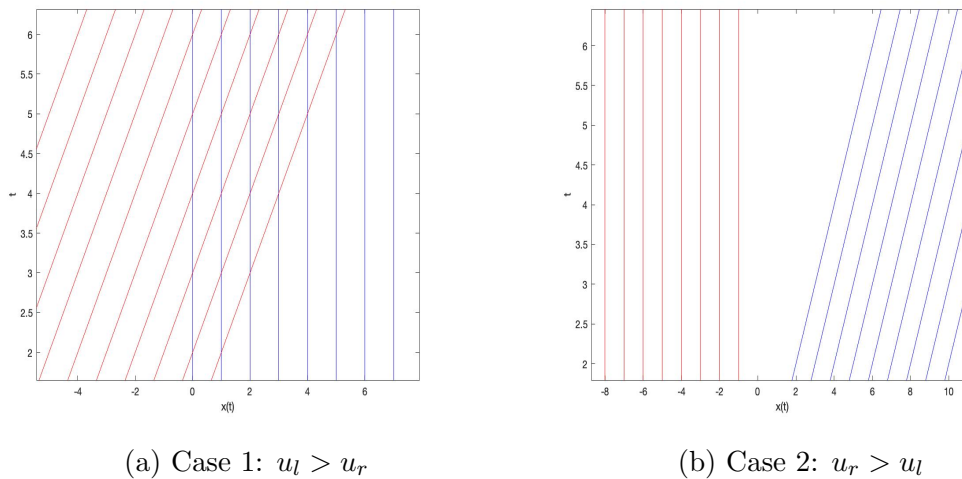


Figure 2.1: Characteristics lines for Burger's equation with discontinuous initial data

We notice that in the first case ($u_l > u_r$), the lines intersect, so we can solve the problem up to a time when the characteristics intersect. Whereas in the second case ($u_r > u_l$), they do not cover the whole plane. Therefore, a smooth enough classical solution can be only constructed in a small time interval. These discontinuities lead us to look into weak solutions.

Weak solutions

Let $u \in C^1(\mathbb{R} \times \mathbb{R}_+)$ be a solution to (2.10) and let $\varphi \in C_c^1(\mathbb{R} \times \mathbb{R}_+)$ be a test function. Then, as we proceeded in the previous section, using Green's formula and the fact that φ is compactly supported we get

$$\begin{aligned} & \int_0^\infty \int_{\mathbb{R}} (u_t + f(u)_x) \varphi dx dt \\ &= - \int_0^\infty \int_{\mathbb{R}} u \varphi_t + f(u) \varphi_x dx dt - \int_{\mathbb{R}} u(x, 0) \varphi(x, 0) dx \end{aligned}$$

Therefore we get

$$\int_0^\infty \int_{\mathbb{R}} u \varphi_t + f(u) \varphi_x dx dt + \int_{\mathbb{R}} u(x, 0) \varphi(x, 0) dx = 0 \quad (2.13)$$

Definition 3. Consider the problem (2.10) and assume $u_0 \in L_{loc}^\infty(\mathbb{R})$ then $u \in L_{loc}^\infty(\mathbb{R} \times \mathbb{R}_+)$ is a weak solution of (2.10) if $u(x, t)$ satisfies equation (2.13) for all $\varphi \in C_c^1(\mathbb{R} \times \mathbb{R}_+)$.

Of particular importance in our case is the family of piece-wise smooth functions which admit jump discontinuities at a finite number of smooth surfaces Σ . Let Σ be a smooth orientable surface. We say $u(x, t)$ is a piece-wise C^1 function if u is C^1 outside Σ and admits a jump discontinuity across Σ (for simplicity, we take u to have only one discontinuity).

Let $\vec{n} = (n_t, n_x)^T$ be a normal vector to Σ and denote by u_+ and u_- the values of u at either sides of Σ .

In the one-dimensional case, Σ could be parametrized with $(t, \xi(t))$ where $\xi : (t_1, t_2) \rightarrow \mathbb{R}$ is a C^1 function. Hence, for $\epsilon > 0$ we have

$$u_\pm(x, t) = \lim_{\epsilon \rightarrow 0} u((x, t) \pm \epsilon \vec{n}), \quad x \in \Sigma, t > 0 \quad (2.14)$$

or

$$u_\pm(\xi(t), t) = \lim_{\epsilon \rightarrow 0} u(\xi(t) \pm \epsilon, t) \quad (2.15)$$

Moreover, n_t and n_x could be seen as the speed s and direction of propagation of the discontinuity Σ , hence we take $\vec{n} = (-s(t), 1)$.

The next theorem tells us that not every discontinuity of piece-wise C^1 functions is admissible and provides us with the Rankine-Hugoniot condition.

Theorem 4. A piece-wise C^1 function $u : \mathbb{R} \times \mathbb{R}^+ \rightarrow \mathbb{R}$ is a solution to (2.10) in the sense of (2.13) on $\mathbb{R} \times \mathbb{R}^+$ if and only if the following two conditions hold

- i. u is a classical solution in the domains where $u \in C^1$.

ii. u satisfies the Rankine-Hugoniot condition, i.e. for $x \in \Sigma, t > 0$:

$$-s.(u_+(x, t) - u_-(x, t)) + f(u_+(x, t)) - f(u_-(x, t)) = 0 \quad (2.16)$$

Proof. Let M be a point on Σ and D a ball centered at M . Denote by D_+ and D_- the open subsets of D on each side of Σ .

Let $\varphi \in C_c^1(D)$, then we have

$$0 = \int_D u\varphi_t + f(u)\varphi_x dxdt = \int_{D_+} + \int_{D_-}$$

Now choosing the normal vector \vec{n} in the direction of D_+ , we split the integrals as follows

$$\begin{aligned} 0 &= - \int_{D_+} (u_t + (f(u))_x)\varphi dxdt \\ &\quad - \int_{\Sigma \cap D} (n_t u_+ + n_x f(u_+))\varphi dS \\ &\quad - \int_{D_-} (u_t + (f(u))_x)\varphi dxdt \\ &\quad + \int_{\Sigma \cap D} (n_t u_- + n_x f(u_-))\varphi dS \end{aligned}$$

The first and third terms cancel out since u is a classical solution in D_{\pm} , hence we get

$$\int_{\Sigma \cap D} (n_t(u_+ - u_-) + n_x(f(u_+) - f(u_-))\varphi dS = 0$$

which gives us the condition we want since φ is arbitrary and $n_t = -s$ and $n_x = 1$. \square

Remark 5. *If the function $u(x, t)$ is continuous then the Rankine-Hugoniot condition (2.16) is already satisfied, so it is enough to check that it is a classical solution in the domains where it is C^1 .*

Non-uniqueness of weak solutions

In this part we present explicit forms of different weak solutions for the special conservation law, namely Burger's equation.

Suppose again we have the Riemann problem

$$\begin{cases} u_t + (\frac{u^2}{2})_x = 0 \\ u(x, 0) = \begin{cases} u_l & \text{if } x \leq 0 \\ u_r & \text{if } x > 0 \end{cases} \end{cases} \quad (2.17)$$

where $u_l \neq u_r$.

Below are examples of weak solutions satisfying the Rankine-Hugoniot condition.

- For $s = \frac{u_l + u_r}{2}$

$$u(x, t) = \begin{cases} u_l & \text{if } x < st \\ u_r & \text{if } x > st \end{cases} \quad (2.18)$$

- For $a \geq \max(u_l, u_r)$, $s_1 = \frac{u_l - a}{2}$ and $s_2 = \frac{a + u_r}{2}$

$$u(x, t) = \begin{cases} u_l & \text{if } x < s_1 t \\ -a & \text{if } s_1 t < x < 0 \\ a & \text{if } 0 < x < s_2 t \\ u_r & \text{if } x > s_2 t \end{cases} \quad (2.19)$$

- For $u_l \leq u_r$ specifically

$$u(x, t) = \begin{cases} u_l & \text{if } x \leq u_l t \\ \frac{x}{t} & \text{if } u_l t \leq x \leq u_r t \\ u_r & \text{if } x \geq u_r t \end{cases} \quad (2.20)$$

This is a self-similar solution which is meant to fill the empty space between the characteristics in figure 2.1b with a rarefaction wave.

Lax-Entropy condition

In order to solve this non-uniqueness and eliminate non-physical solutions we provide an additional condition, the Lax Entropy condition below

$$f'(u_-(x, t)) > s(t) > f'(u_+(x, t)) \quad (2.21)$$

where u_- and u_+ are the values of the solution on both sides of the shock and $s(t)$ is the speed of the shock.

Condition (2.21) excludes solution (2.19) for $u_l > u_r$ and both (2.18) and (2.19) for $u_l < u_r$. On the other hand we have the following two propositions that provide us with weak solutions satisfying the Lax entropy condition for both cases.

Proposition 1. *A weak solution for (2.17) where $u_l > u_r$ that satisfies the Rankine-Hugoniot is given by*

$$u(x, t) = \begin{cases} u_l & \text{if } x < st \\ u_r & \text{if } x > st \end{cases} \quad (2.22)$$

where $s = \frac{u_l^2 + u_r^2}{2}$

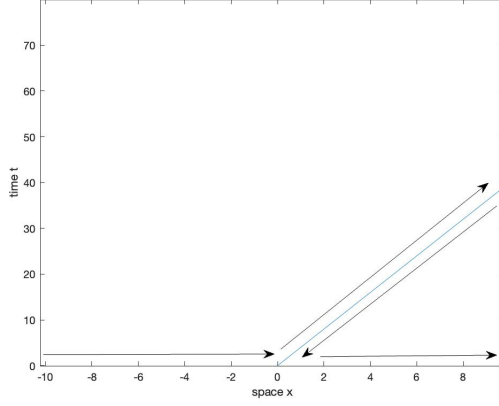


Figure 2.2: shock wave as a line in (x,t) plane

Proof.

$$\begin{aligned} \int_0^\infty \int_{-\infty}^\infty \varphi_t u + \varphi_x \frac{u^2}{2} dx dt &= \int_0^\infty \int_{x < st} \varphi_t u_l + \varphi_x \frac{u_l^2}{2} dx dt + \int_0^\infty \int_{x > st} \varphi_t u_r + \varphi_x \frac{u_r^2}{2} dx dt \\ &= \int_0^\infty \int_{x < st} (\varphi u_l)_t + (\varphi \frac{u_l^2}{2})_x dx dt + \int_0^\infty \int_{x > st} (\varphi u_r)_t + (\varphi \frac{u_r^2}{2})_x dx dt \end{aligned}$$

Now using Green's theorem we get

$$\begin{aligned} &= \int_{\partial(x < st)} -\varphi u_l dx + \varphi \frac{u_l^2}{2} dt + \int_{\partial(x > st)} -\varphi u_r dx + \varphi \frac{u_r^2}{2} dt \\ &= \int_{-\infty}^0 -\varphi u_l dx + \int_{x=st} -\varphi u_l + \frac{u_l^2}{2s} dx + \int_0^\infty -\varphi u_r dx - \int_{x=st} -\varphi u_r + \frac{u_r^2}{2s} dx \\ &= \int_{x=st} \varphi \left(\frac{u_l^2 - u_r^2}{2s} - (u_l - u_r) \right) dx - \int_{-\infty}^\infty \varphi(x, 0) u(x, 0) dx \end{aligned}$$

Take $s = \frac{u_r + u_l}{2}$

$$= \int_{x=st} \varphi ((u_l - u_r) - (u_l - u_r)) dx - \int_{-\infty}^\infty \varphi(x, 0) u(x, 0) dx = - \int_{-\infty}^\infty \varphi(x, 0) u(x, 0) dx$$

□

Proposition 2. A weak solution for (2.17) where $u_r > u_l$ that satisfies the Rankine-Hugoniot and Lax-Entropy conditions is given by

$$u(x, t) = \begin{cases} u_l & \text{if } x < u_l t \\ \frac{x}{t} & \text{if } u_l t < x < u_r t \\ u_r & \text{if } x > u_r t \end{cases} \quad (2.23)$$

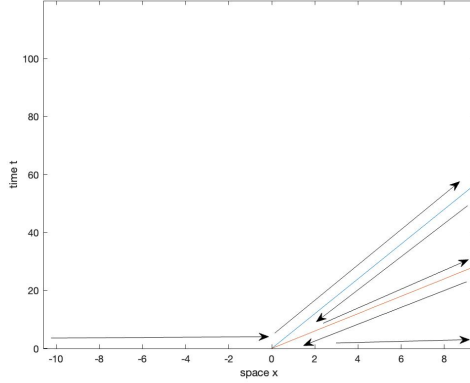


Figure 2.3: rarefaction wave as a line in (x,t) plane

Proof. The proof is somewhat similar to the previous one.

$$\begin{aligned}
\int_0^\infty \int_{-\infty}^\infty \varphi_t u + \varphi_x \frac{u^2}{2} dx dt &= \int_0^\infty \int_{x < u_l t} \varphi_t u_l + \varphi_x \frac{u_l^2}{2} dx dt \\
&\quad + \int_0^\infty \int_{u_l t < x < u_r t} \varphi_t \frac{x}{t} + \varphi_x \frac{x^2}{2t^2} dx dt \\
&\quad + \int_0^\infty \int_{x > u_r t} \varphi_t u_r + \varphi_x \frac{u_r^2}{2} dx dt
\end{aligned}$$

Re-arranging the terms we get

$$\begin{aligned}
&= \int_0^\infty \int_{x < u_l t} (\varphi u_l)_t + \left(\varphi \frac{u_l^2}{2}\right)_x dx dt \\
&\quad + \int_0^\infty \int_{u_l t < x < u_r t} \left(\varphi \frac{x}{t}\right)_t + \left(\varphi \frac{x^2}{2t^2}\right)_x dx dt \\
&\quad + \int_0^\infty \int_{x > u_r t} (\varphi u_r)_t + \left(\varphi \frac{u_r^2}{2}\right)_x dx dt
\end{aligned}$$

Using Green's theorem we get

$$\begin{aligned}
&= \int_{\partial(x < u_l t)} -\varphi u_l dx + \varphi \frac{u_l^2}{2} dt \\
&\quad + \int_{\partial(u_l t < x < u_r t)} -\varphi \frac{x}{t} dx + \varphi \frac{x^2}{2t^2} dt \\
&\quad + \int_{\partial(x > u_r t)} -\varphi u_r dx + \varphi \frac{u_r^2}{2} dt
\end{aligned}$$

In this next step, terms will cancel out except the first and last one

$$\begin{aligned}
&= \int_{-\infty}^0 -\varphi u_l dx + \int_{x=u_l t} -\varphi u_l + \varphi \frac{u_l}{2} dx \\
&- \int_{x=u_l t} -\varphi \frac{x}{t} + \varphi \frac{x}{2t^2} dx + \int_{x=u_r t} -\varphi \frac{x}{t} + \varphi \frac{x}{2t^2} dx \\
&- \int_{x=u_r t} -\varphi u_r + \varphi \frac{u_r}{2} dx + \int_0^{\infty} -\varphi u_r dx \\
&= - \int_{-\infty}^{\infty} \varphi(x, 0) u(x, 0) dx
\end{aligned}$$

□

Entropy solutions

The concept of entropy solutions is another rigorous way to identify the physically relevant solutions. We present the following analysis leading to the main theorem at the end of the chapter.

Let Ω be an open subset of \mathbb{R} and let $u : \mathbb{R} \times [0, \infty) \rightarrow \Omega$.

Let $U(u)$ and $F(u)$ be sufficiently smooth functions from Ω to \mathbb{R} and suppose that $u(x, t)$ satisfies another conservation law

$$(U(u))_t + (F(u))_x = 0 \tag{2.24}$$

we proceed with the following steps

$$\begin{aligned}
&u_t U'(u) + u_x F'(u) = 0 \\
&-u_x f'(u) U'(u) + u_x F'(u) = 0 \text{ (since } u_t + f'(u)u_x = 0\text{)} \\
&f'(u) U'(u) = F'(u)
\end{aligned}$$

Definition 4. *Suppose Ω is convex. If $U(u)$ and $F(u)$ are smooth functions from Ω to \mathbb{R} such that*

$$f'(u) U'(u) = F'(u) \tag{2.25}$$

then U is called an entropy for the conservation law and F an entropy flux.

Remark 6. *Any convex function is an entropy for the one dimensional case and we take the entropy flux F to be the primitive of $U' f'$.*

Let $\epsilon > 0$ be small enough and consider the viscous perturbation of the equation of the conservation law

$$(u_\epsilon)_t + (f(u_\epsilon))_x = \epsilon (u_\epsilon)_{xx} \tag{2.26}$$

where the right hand side is called the viscosity term.

Assume that $u_\epsilon(x, 0) \rightarrow u(x, 0) = u_0(x)$ as $\epsilon \rightarrow 0$, then the paper by Goodman and Xin [8], insures the existence of such a solution u_ϵ to (2.26).

This method will help us recover physical solutions u from u_ϵ by taking the limit as ϵ goes to zero.

Theorem 5. *Assume that the problem (2.10) admits an entropy U with entropy flux F . Let $(u_\epsilon)_\epsilon$ be a sequence of sufficiently smooth solutions of the viscous conservation law (2.26) that satisfy the following*

- i. $\|u_\epsilon\|_{L^\infty(\mathbb{R} \times \mathbb{R}_+)} \leq C$ for some constant $C > 0$
- ii. $u_\epsilon \rightarrow u$ as $\epsilon \rightarrow 0$ a.e. in $\mathbb{R} \times \mathbb{R}_+$

then u is a weak solution of (2.7) and satisfies the entropy condition

$$(U(u))_t + (F(u))_x \leq 0 \quad (2.27)$$

in the weak sense.

Remark 7. *Weak form of the entropy condition (2.27) is as follows*

$$\int_0^\infty \int_{\mathbb{R}} U(u) \varphi_t + F(u) \varphi_x dx dt \geq 0 \quad (2.28)$$

for all $\varphi \in C_c^1(\mathbb{R} \times \mathbb{R}_+)$, $\varphi \geq 0$.

Definition 5. *A weak solution of (2.10) is called an entropy solution if $u \in L^\infty(\mathbb{R} \times \mathbb{R}_+)$ satisfies the entropy condition (2.28).*

Existence and uniqueness of entropy solutions

Definition 6. *Let g be a function on $[a, b]$. Then define the total variation of g by*

$$TV(g) = \sup_{P \in \mathcal{P}} |g(x_{i+1}) - g(x_i)| \quad (2.29)$$

where the sup is taken over the set \mathcal{P} of all partitions of the interval $[a, b]$.

Remark 8.

- i. *If g is differentiable and its derivative is Riemann integrable then its Total variation becomes*

$$TV(g) = \int_a^b |g'(x)| dx$$

- ii. *The total variation is only a semi-norm since the total variation of a constant function is zero.*

Definition 7. Define the norm $BV(g) = \|g\|_{L^1} + TV(g)$.

Moreover, the set of functions of bounded variation is defined as

$$BV(\mathbb{R}) = \{g \in L^1(\mathbb{R}) \text{ such that } BV(g) < \infty\}$$

We now state the main theorem of existence and uniqueness of an entropy solution ([5]).

Theorem 6. Let $u_0 \in L^\infty(\mathbb{R})$. Then problem (2.10) admits a unique entropy solution $u \in L^\infty(\mathbb{R} \times \mathbb{R}_+)$ such that

$$\|u(\cdot, t)\|_{L^\infty(\mathbb{R})} \leq \|u_0\|_{L^\infty(\mathbb{R})} \quad (2.30)$$

for all $t \geq 0$.

Moreover, if u and v are two entropy solutions of (2.10) with initial data u_0 and v_0 respectively then

$$u_0 \geq v_0 \implies u(\cdot, t) \geq v(\cdot, t) \text{ a.e.} \quad (2.31)$$

Finally, if $u_0 \in L^\infty(\mathbb{R}) \cap BV(\mathbb{R})$ then $u(\cdot, t) \in BV(\mathbb{R})$ such that

$$TV(u(\cdot, t)) \leq TV(u_0). \quad (2.32)$$

Weak entropy solutions for the Riemann problem

To summarize, the weak entropy solution for Burger's equation with initial data

$$u_0(x) = \begin{cases} u_l & \text{if } x \leq 0 \\ u_r & \text{if } x > 0 \end{cases}$$

is as follows

- if $u_l < u_r$

$$u(x, t) = \begin{cases} u_l & \text{if } x < u_l t \\ \frac{x}{t} & \text{if } u_l t < x < u_r t \\ u_r & \text{if } x > u_r t \end{cases} \quad (2.33)$$

- if $u_l = u_r$

$$u(x, t) = u_l \quad (2.34)$$

- if $u_l > u_r$

$$u(x, t) = \begin{cases} u_l & \text{if } x < st \\ u_r & \text{if } x > st \end{cases} \quad (2.35)$$

where $s = \frac{u_l + u_r}{2}$.

CHAPTER 3

NUMERICAL APPROXIMATION OF SOLUTIONS

In this chapter we present numerical schemes in order to approximate solutions to the linear and non-linear transport equations mentioned in the previous chapter. We use finite difference methods for the linear case then proceed with finite volume for the non-linear case.

3.1 Finite Difference Method for Linear Transport Equation

In order to approximate solutions to the transport equation, we will use finite difference schemes using [6] as a reference. The discussion in this section will focus on equations with constant velocity but will be generalized in the next chapter through numerical examples.

3.1.1 Deriving The Schemes

We will start with the Taylor expansion of the solution u in order to approximate the partial derivatives in our PDE.

Taylor expansion of a function $f \in C^{(n+1)}$ about x_0 :

$$f(x_0 + h) = f(x_0) + f'(x_0)h + \dots + \frac{1}{n!}f^{(n)}(x_0)h^n + \frac{1}{(n+1)!}f^{(n+1)}(\xi)h^{n+1} \quad (3.1)$$

where $\xi \in (x_0, x_0 + h)$.

Therefore we have the following formulae for approximating derivatives:

- Forward difference: for $f \in C^k$, $k \geq 2$,

$$f'(x_0) = \frac{f(x_0 + h) - f(x_0)}{h} + O(h) \quad (3.2)$$

- Backward difference: for $f \in C^k$, $k \geq 2$,

$$f'(x_0) = \frac{f(x_0) - f(x_0 - h)}{h} + O(h) \quad (3.3)$$

- Central difference: for $f \in C^k$, $k \geq 4$,

$$f'(x_0) = \frac{f(x_0 + h) - f(x_0 - h)}{2h} + O(h^2) \quad (3.4)$$

Now let's apply formulae (3.2), (3.3) and (3.4) to discretize the transport equation $u_t + au_x = 0$, replacing derivatives by differences.

Discretizing the domain

In order to implement the schemes numerically we will need to discretize the space and time domains. To this end, let

- $\Omega_{\Delta x} \in \Omega$ such that $\Omega_{\Delta x} = \{x_j = j\Delta x | j = -N, \dots, 0, \dots, N\}$ where $N\Delta x = L$ and $\lim_{\Delta x \rightarrow 0} N\Delta x = \infty$ i.e., $N = O(\frac{1}{\Delta x^\alpha})$ and $\alpha > 1$.
- $\mathcal{T}_{\Delta t} = \{t^n = (n-1)\Delta t | n = 1, 2, \dots, M+1 \text{ and } M\Delta t = T\}$

Hence the numerical solution U at x_j and t^n will be denoted as U_j^n .

Central scheme

The central scheme is built using forward difference in time and central difference in space, so using formulas (3.2) and (3.4) we get

$$\frac{U_j^{n+1} - U_j^n}{\Delta t} + a \frac{U_{j+1}^n - U_{j-1}^n}{2\Delta x} = 0$$

Definition 8. Let U_j^n be the numerical approximation of the solution u to the Transport equation at time n and mesh point j , then the Central scheme is defined as follows

$$U_j^{n+1} = U_j^n - a \frac{\Delta t}{2\Delta x} (U_{j+1}^n - U_{j-1}^n) \quad (3.5)$$

Upwind scheme

The upwind scheme is build using forward difference in time and, depending on the sign of a , forward or backward difference in space, so again using formulas (3.2), (3.3) and (3.4) we get

- Forward in time, backward in space

$$\frac{U_j^{n+1} - U_j^n}{\Delta t} + a \frac{U_j^n - U_{j-1}^n}{\Delta x} = 0$$

- Forward in time, forward in space

$$\frac{U_j^{n+1} - U_j^n}{\Delta t} + a \frac{U_{j+1}^n - U_j^n}{\Delta x} = 0$$

Definition 9. Let U_j^n be the numerical approximation of the solution u to the Transport equation at time t^n and mesh point x_j , then the Upwind scheme is defined as follows

$$U_j^{n+1} = \begin{cases} U_j^n - a \frac{\Delta t}{\Delta x} (U_{j+1}^n - U_j^n), & \text{if } a < 0 \\ U_j^n - a \frac{\Delta t}{\Delta x} (U_j^n - U_{j-1}^n), & \text{if } a > 0 \end{cases} \quad (3.6)$$

The below theorem will provide a global formula for the Upwind scheme that will make it simpler to implement.

Theorem 7. The Upwind scheme (3.6) can be written as

$$U_j^{n+1} = U_j^n + a \frac{\Delta t}{2\Delta x} (U_{j+1}^n - U_{j-1}^n) + |a| \frac{\Delta t}{2\Delta x} (U_{j+1}^n - 2U_j^n + U_{j-1}^n) \quad (3.7)$$

Proof. Let $a^+ = \max\{a, 0\}$, $a^- = \min\{a, 0\}$, we get $|a| = a^+ - a^-$ and $a = a^+ + a^-$. If $a > 0$, $a^+ = a$ and $a^- = 0$, and if $a < 0$, $a^- = 0$ and $a^+ = a$. We can now combine the two formulae of (3.6) as follows:

$$\frac{u(x_j, t^{n+1}) - u(x_j, t^n)}{\Delta t} + a^+ \frac{u(x_j, t^n) - u(x_{j-1}, t^n)}{\Delta x} + a^- \frac{u(x_{j+1}, t^n) - u(x_j, t^n)}{\Delta x} = 0$$

$$\text{But } a^+ = \frac{a^+ + a^+}{2} = \frac{|a| + a}{2} \text{ and } a^- = \frac{a^- + a^-}{2} = \frac{a - |a|}{2}.$$

The above formula becomes

$$\begin{aligned} & \frac{u(x_j, t^{n+1}) - u(x_j, t^n)}{\Delta t} + |a| \frac{u(x_j, t^n)}{\Delta x} - |a| \frac{u(x_{j-1}, t^n)}{2\Delta x} \\ & - a \frac{u(x_{j-1}, t^n)}{2\Delta x} + a \frac{u(x_{j+1}, t^n)}{2\Delta x} - |a| \frac{u(x_{j+1}, t^n)}{2\Delta x} = 0 \end{aligned}$$

Which implies the global scheme above. □

3.1.2 Stability Analysis

Definition 10. The discrete energy of the solution U_j^n is defined as follows

$$E^n = \frac{1}{2} \Delta x \sum_j (U_j^n)^2 \quad (3.8)$$

Since the exact solution has a bounded energy as we saw in lemma 1 of chapter 2, we will show that the central scheme is unstable because the energy is unbounded whereas the upwind scheme will turn out to be conditionally stable.

Lemma 2. Let U_j^n be the solutions computed with the Central scheme (3.5) then

$$E^{n+1} = E^n + \frac{1}{8} \frac{\Delta t^2}{\Delta x} a^2 \sum_j (U_{j+1}^n - U_{j-1}^n)^2 \quad (3.9)$$

Proof.

$$\begin{aligned} E^{n+1} &= \frac{1}{2} \Delta x \sum_j (U_j^{n+1})^2 \\ &= \frac{1}{2} \Delta x \sum_j \left(U_j^n - a \frac{\Delta t}{2\Delta x} (U_{j+1}^n - U_{j-1}^n) \right)^2 \\ &= \frac{1}{2} \Delta x \sum_j \left((U_j^n)^2 + \left(a \frac{\Delta t}{2\Delta x} \right)^2 (U_{j+1}^n - U_{j-1}^n)^2 - a \frac{\Delta t}{\Delta x} (U_{j+1}^n - U_{j-1}^n) U_j^n \right) \\ &= E^n + \frac{1}{8} \frac{a^2 \Delta t^2}{\Delta x} \sum_j (U_{j+1}^n - U_{j-1}^n)^2 \end{aligned}$$

The last line was possible assuming zero or periodic boundary conditions. \square

Lemma 3. If the condition

$$\left| a \frac{\Delta t}{\Delta x} \right| \leq 1 \quad (\text{CFL condition}) \quad (3.10)$$

is satisfied then the solutions computed using the Upwind scheme (3.6) satisfies the energy bound

$$E^{n+1} \leq E^n \quad (3.11)$$

Proof.

$$\begin{aligned} E^{n+1} &= \frac{1}{2} \Delta x \sum_j (U_j^{n+1})^2 \\ &= \frac{1}{2} \Delta x \sum_j \left(\left| a \frac{\Delta t}{2\Delta x} (U_{j+1}^n - 2U_j^n + U_{j-1}^n) - a \frac{\Delta t}{2\Delta x} (U_{j+1}^n - U_{j-1}^n) + U_j^n \right|^2 \right) \end{aligned}$$

in the following we will assume $a > 0$, same result follows for $a < 0$ by replacing j by $j + 1$

$$\begin{aligned}
E^{n+1} &= \frac{1}{2}\Delta x \sum_j (U_j^n + a\frac{\Delta t}{\Delta x}(U_{j-1}^n - U_j^n))^2 \\
&= \frac{1}{2}\Delta x \sum_j ((U_j^n)^2 + a^2\frac{\Delta t^2}{\Delta x^2}(U_{j-1}^n - U_j^n)^2 + 2a\frac{\Delta t}{\Delta x}(U_{j-1}^n U_j^n - (U_j^n)^2)) \\
&= E^n + (\frac{1}{2}a^2\frac{\Delta t^2}{\Delta x} - \frac{1}{2}a\Delta t) \sum_j (U_{j-1}^n - U_j^n)^2 + \frac{a\Delta t}{4} \sum_j ((U_{j-1}^n)^2 - (U_j^n)^2)
\end{aligned} \tag{3.12}$$

assuming zero or periodic boundary conditions we can ignore the last term

$$= E^n + (\frac{1}{2}a^2\frac{\Delta t^2}{\Delta x} - \frac{1}{2}a\Delta t) \sum_j (U_{j-1}^n - U_j^n)^2$$

So we need

$$\begin{aligned}
\frac{1}{2}a^2\frac{\Delta t^2}{\Delta x} - \frac{1}{2}a\Delta t &\leq 0 \\
a\frac{\Delta t}{\Delta x} &\leq 1
\end{aligned}$$

Note that equation (3.12) was possible since

$$\begin{aligned}
\frac{(U_{j-1}^n - U_j^n)^2}{2} &= \frac{(U_{j-1}^n)^2}{2} + \frac{(U_j^n)^2}{2} - U_{j-1}^n U_j^n + (U_j^n)^2 - (U_j^n)^2 \\
&= \frac{(U_{j-1}^n)^2}{2} - \frac{(U_j^n)^2}{2} - (U_{j-1}^n U_j^n - (U_j^n)^2)
\end{aligned}$$

□

Thus the central scheme is unstable however the upwind scheme is conditionally stable.

3.2 Finite Volume Method for Non-Linear equation

In order to approximate solutions to the non linear Transport equation, namely the conservation law $u_t + (f(u))_x = 0$, we will use finite volume schemes. We will also limit the discussion to equations where the function f is convex and piecewise constant initial data as in chapter 2 (specifically Burger's equation). We discretize our time-space domain into mesh points x_j and t^n with distance Δt and Δx between them respectively. Define the control volumes as the intervals

$$\{[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}], \quad j = 0, 1, \dots, N-1\}$$

We will approximate the solution U_j^n on these control volumes using the cell averages below

$$U_j^n = \frac{1}{\Delta x} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} U(x, t^n) dx + O(h^2)$$

In order to get the cell average at the next time step ($n+1$) we proceed as follows

$$\begin{aligned} U_t + (f(U))_x &= 0 \\ \int_{t^n}^{t^{n+1}} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} U_t dx dt + \int_{t^n}^{t^{n+1}} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} (f(U))_x dx dt &= 0 \\ \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} U(x, t^{n+1}) - U(x, t^n) dx + \int_{t^n}^{t^{n+1}} f(U(x_{j+\frac{1}{2}}, t)) - f(U(x_{j-\frac{1}{2}}, t)) dt &= 0 \\ U_j^{n+1} = U_j^n + \frac{1}{\Delta x} \int_{t^n}^{t^{n+1}} f(U(x_{j+\frac{1}{2}}, t)) - f(U(x_{j-\frac{1}{2}}, t)) dt \end{aligned}$$

Hence

$$U_j^{n+1} = U_j^n + \frac{\Delta t}{\Delta x} (F_{j+\frac{1}{2}}^n - F_{j-\frac{1}{2}}^n) \quad (3.13)$$

where F_s^n is the approximation of $\int_{t^n}^{t^{n+1}} f(U(x_s, t)) dt$ for $s = j \pm \frac{1}{2}$ and H is the solution operator.

In the following sections we will discuss schemes to approximate the flux F_s^n and thus the solution to the problem. The first one, being the Godunov scheme, uses the exact solution of the Riemann problem whereas the other two use approximations of the solution and thus are called Approximate Riemann Solvers.

3.2.1 Godunov scheme

The first scheme to be discussed is the Godunov scheme. The idea behind it is to solve the Riemann problem at each $x_{j+\frac{1}{2}}$.

$$\begin{cases} U_t + (f(U))_x = 0 \\ U(x, t^n) = \begin{cases} U_j^n & \text{if } x < x_{j+\frac{1}{2}} \\ U_{j+1}^n & \text{if } x > x_{j+\frac{1}{2}} \end{cases} \end{cases} \quad (3.14)$$

where $s = \frac{U_j^n + U_{j+1}^n}{2}$ in the case of Burger's equation.

Thus we will end up solving a sequence of Riemann problems, and need to impose the CFL condition

$$|\max_j f'(U_j^n)| \frac{\Delta t}{\Delta x} \leq \frac{1}{2} \quad (3.15)$$

in order for waves not to intersect between each problem.

The solution of (3.14) depends on whether U_j^n is greater or less than U_{j+1}^n , which

was discussed in the previous chapter.

An explicit formula for the Godunov flux could be obtained and it is stated in the definition below.

Definition 11. *The Godunov flux is defined as follows*

$$F_{j+\frac{1}{2}}^n = \begin{cases} \min_{U_j^n \leq \theta \leq U_{j+1}^n} f(\theta) & \text{if } U_j^n \leq U_{j+1}^n \\ \max_{U_{j+1}^n \leq \theta \leq U_j^n} f(\theta) & \text{if } U_{j+1}^n \leq U_j^n \end{cases} \quad (3.16)$$

Since equation (3.16) requires the solution of an optimization problem, the following theorem will provide a simpler form of (3.16) but under conditions.

Theorem 8. *Let f be a function with a single minimum at w , then the Godunov flux (3.16) can be written as*

$$F_{j+\frac{1}{2}}^n = \max(f(\max(U_j^n, w)), f(\min(U_{j+1}^n, w))) \quad (3.17)$$

Proof. 1. If $U_j^n \leq U_{j+1}^n$

- (a) If $w \leq U_j^n \leq U_{j+1}^n$
 $\min f(\theta) = f(U_j^n)$
 $\max(f(U_j^n), f(w)) = f(U_j^n)$
- (b) If $U_j^n \leq U_{j+1}^n \leq w$
 $\min f(\theta) = f(U_{j+1}^n)$
 $\max(f(w), f(U_{j+1}^n)) = f(U_{j+1}^n)$
- (c) If $U_j^n \leq w \leq U_{j+1}^n$ or $U_{j+1}^n \leq w \leq U_j^n$
 $\min f(\theta) = f(w)$
 $\max(f(w), f(w)) = f(w)$

2. If $U_{j+1}^n \leq U_j^n$

Similar cases can be taken here. □

Remark 9. *Since $f(u) = \frac{u^2}{2}$ admits only one minimum at 0, then formula (3.17) can be used for Burger's equation.*

Remark 10. *Since the Godunov flux requires an optimization problem, it may not be the most efficient scheme to use as it will be costly. To this end, we look into other methods.*

3.2.2 Roe scheme

The Roe scheme is the first approximate Riemann solver we're presenting which uses a linearizing technique to define a numerical flux.

The idea behind it is to replace the derivative of the function f in the non-linear equations such as the conservation law

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

or

$$u_t + f'(u)u_x = 0$$

with a linear form.

We use the Roe average below to approximate the derivative

$$D_{j+\frac{1}{2}}^n = \begin{cases} \frac{f(U_{j+1}^n) - f(U_j^n)}{U_{j+1}^n - U_j^n} & \text{if } U_{j+1}^n \neq U_j^n \\ U_j^n & \text{if } U_{j+1}^n = U_j^n \end{cases} \quad (3.18)$$

Note that for Burger's equation, the Roe average becomes

$$D_{j+\frac{1}{2}}^n = \frac{U_{j+1}^n + U_j^n}{2}.$$

After linearizing the equation we can get an explicit formula for the solution of the Riemann problem and hence we can define the Roe flux.

Definition 12. *The Roe flux is defined as follows*

$$F_{j+\frac{1}{2}}^n = \begin{cases} f(U_j^n) & \text{if } D_{j+\frac{1}{2}}^n \geq 0 \\ f(U_{j+1}^n) & \text{if } D_{j+\frac{1}{2}}^n < 0 \end{cases} \quad (3.19)$$

3.2.3 Rusanov scheme

The second approximate Riemann solver is a central scheme, which we define below.

Definition 13. *An approximate Riemann solver is a central scheme if it defines the numerical flux as follows*

$$F_{j+\frac{1}{2}}^n = \frac{f(U_j^n) + f(U_{j+1}^n)}{2} - \frac{s_{j+\frac{1}{2}}}{2}(U_{j+1}^n - U_j^n) \quad (3.20)$$

Different choices of $s_{j+\frac{1}{2}}$ lead to different central schemes, one of them being the Rusanov scheme where

$$s_{j+\frac{1}{2}} = \max(|f'(U_j^n)|, |f'(U_{j+1}^n)|)$$

Definition 14. *The Rusanov flux is defined as follows*

$$F_{j+\frac{1}{2}}^n = \frac{f(U_j^n) + f(U_{j+1}^n)}{2} - \frac{\max(|f'(U_j^n)|, |f'(U_{j+1}^n)|)}{2}(U_{j+1}^n - U_j^n) \quad (3.21)$$

3.2.4 Convergence Analysis (Lax-Wendroff Theorem)

We now conduct some convergence analysis on the finite volume method in order to understand the numerical results given in the next chapter. To this end, we will define some concepts such as conservation, consistency and monotonicity of the schemes. These concepts will provide us with sufficient conditions for the convergence of the schemes to the weak entropy solution.

Let us first recall the main form of the finite volume scheme (3.13).

$$U_j^{n+1} = U_j^n + \frac{\Delta t}{\Delta x} (F_{j+\frac{1}{2}}^n - F_{j-\frac{1}{2}}^n)$$

Let H denote the discrete solution operator then we have

$$U_j^{n+1} = H(U_{j-1}^n, U_j^n, U_{j+1}^n) \quad (3.22)$$

Definition 15. A finite volume scheme of the form (3.13) is called a conservative scheme if we have, for all times t^n

$$\sum_j U_j^{n+1} = \sum_j U_j^n \quad (3.23)$$

Theorem 9. Let U_j be a sequence in $L^1(\mathbb{Z})$ (i.e. $\sum_j |U_j| < \infty$). Then the scheme (3.22) is conservative if and only if it can be put in the form (3.13).

Proof. \Leftarrow : Suppose we have the form (3.13) then

$$\sum_j U_j^{n+1} = \sum_j U_j^n + \frac{\Delta t}{\Delta x} (F_{j+\frac{1}{2}}^n - F_{j-\frac{1}{2}}^n) = \sum_j U_j^n$$

Since the second term is a telescoping series.

$$\Rightarrow: \text{Define } G(U_{j-1}, U_j, U_{j+1}) = \frac{\Delta t}{\Delta x} (U_j - H(U_{j-1}, U_j, U_{j+1})).$$

We know that $\sum_j G = 0$ and we want to show that there exist $F_{j+\frac{1}{2}}$ and $F_{j-\frac{1}{2}}$ such that $F_{j+\frac{1}{2}} - F_{j-\frac{1}{2}} = G(U_{j-1}, U_j, U_{j+1})$.

Let's prove it for $j = 1$.

Let $U_j = 0$ for $j \leq 0$ and $j > 2$ then

$$\begin{aligned} 0 &= \sum_j G(U_{j-1}, U_j, U_{j+1}) \\ &= G(0, 0, U_1) + G(0, U_1, U_2) + G(U_1, U_2, 0) + G(U_2, 0, 0) \\ &= F_{\frac{3}{2}} + G(U_1, U_2, 0) + G(U_2, 0, 0) \end{aligned}$$

Now let $U_j = 0$ for $j < 0$ and $j > 2$ then

$$\begin{aligned} 0 &= \sum_j G(U_{j-1}, U_j, U_{j+1}) \\ &= G(0, U_0, U_1) + G(U_0, U_1, U_2) + G(U_1, U_2, 0) + G(U_2, 0, 0) \\ &= F_{\frac{1}{2}} + G(U_0, U_1, U_2) + G(U_1, U_2, 0) + G(U_2, 0, 0) \end{aligned}$$

Hence $G(U_0, U_1, U_2) = F_{\frac{3}{2}} - F_{\frac{1}{2}}$. \square

Definition 16. A finite volume scheme is said to be consistent if the numerical flux F satisfies the following

$$F(U, U) = f(u) \quad (3.24)$$

We have seen in Theorem 6 that weak entropy solutions satisfy monotone conditions, more precisely we have that if $u_0 \geq v_0$ then $u(\cdot, t) \geq v(\cdot, t)$ almost everywhere. The below definition gives the discrete form of the monotonicity of the scheme.

Definition 17. A finite volume scheme is said to be monotone if the discrete solution operator H is non-decreasing in each of its arguments.

Theorem 10. Let (3.13) be a finite volume scheme. If the numerical flux $F(a, b)$ is Liptchiz non-decreasing in the first argument, non-increasing in the second argument and satisfies the CFL condition

$$\left| \frac{\partial F}{\partial a} \right| + \left| \frac{\partial F}{\partial b} \right| \leq \frac{\Delta x}{\Delta t} \quad (3.25)$$

then the scheme is monotone.

Proof. The proof is rather simple. We find the derivative of H with respect of each of its arguments and show they are non-negative.

We have

$$\begin{aligned} U_j^{n+1} &= U_j^n + \frac{\Delta t}{\Delta x} (F(U_j^n, U_{j+1}^n) - F(U_{j-1}^n, U_j^n)) \\ &= H(U_{j-1}^n, U_j^n, U_{j+1}^n) \end{aligned}$$

- $\frac{\partial H}{\partial U_{j-1}^n} = \frac{\Delta t}{\Delta x} \left(-\frac{\partial F}{\partial b} \right) \geq 0$
- $\frac{\partial H}{\partial U_j^n} = 1 + \frac{\Delta t}{\Delta x} \left(\frac{\partial F}{\partial b} - \frac{\partial F}{\partial a} \right) = 1 - \frac{\Delta t}{\Delta x} \left(\left| \frac{\partial F}{\partial b} \right| + \left| \frac{\partial F}{\partial a} \right| \right) \geq 0$ (using (3.22)).
- $\frac{\partial H}{\partial U_{j+1}^n} = \frac{\Delta t}{\Delta x} \left(\frac{\partial F}{\partial a} \right) \geq 0$

Hence H is non-decreasing in each of its arguments. \square

We now state propositions that follow from monotonicity and provide us with bounds on the numerical solutions of the schemes.

Proposition 3. Suppose U_j^n is the solution of a consistent monotone scheme, then it satisfies the following maximum principle

$$\min(U_{j-1}^n, U_j^n, U_{j+1}^n) \leq U_j^{n+1} \leq \max(U_{j-1}^n, U_j^n, U_{j+1}^n). \quad (3.26)$$

Iterating over all times t^n we get the L^∞ bound

$$\min_k U_k^0 \leq U_j^n \leq \max_k U_k^0 \quad (3.27)$$

for all n and j .

The next two propositions revolve around the total variation diminishing property of some schemes.

Remark 11. *A scheme is called total variation diminishing if*

$$\sum_j |U_{j+1}^{n+1} - U_j^{n+1}| \leq \sum_j |U_{j+1}^n - U_j^n|$$

First, let's introduce the incremental form of the finite volume scheme (3.13). For this purpose, let

- $C_{j+\frac{1}{2}}^n = \frac{\Delta t}{\Delta x} \frac{f(U_j^n) - F_{j+\frac{1}{2}}}{U_{j+1}^n - U_j^n}$
- $D_{j+\frac{1}{2}}^n = \frac{\Delta t}{\Delta x} \frac{f(U_{j+1}^n) - F_{j+\frac{1}{2}}}{U_{j+1}^n - U_j^n}$

then (3.13) can be written in the incremental form below

$$U_{j+1}^n = U_j^n + C_{j+\frac{1}{2}}^n (U_{j+1}^n - U_j^n) - D_{j-\frac{1}{2}}^n (U_j^n - U_{j-1}^n). \quad (3.28)$$

Proposition 4. *Suppose we have a finite volume scheme written in the incremental form (3.28). Then the scheme is total variation diminishing if $C_{j+\frac{1}{2}}^n \geq 0$, $D_{j+\frac{1}{2}}^n \geq 0$ and $C_{j+\frac{1}{2}}^n + D_{j+\frac{1}{2}}^n \leq 1$ for all n and j .*

Proposition 5. *If a finite volume scheme is consistent, monotone and satisfies the CFL condition (3.22) then it is total variation diminishing.*

Let $U^\Delta(x, t) = U_j^n$ for $x \in [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}})$ and $t \in [t^n, t^{n+1})$, we can now state the Lax-Wendroff theorem.

Theorem 11. (Lax-Wendroff) *Let (3.13) be a conservative and consistent finite volume scheme where the numerical flux F is Lipschitz. Suppose $u_0 \in L^\infty(\mathbb{R})$, $\|U^\Delta\|_{L^\infty} \leq C$, for some constant $C > 0$ and U^Δ converges to u in $L^1(\mathbb{R} \times \mathbb{R}_+)$ almost everywhere.*

Then u is a weak solution to the problem.

Since convergence of the numerical solution to the weak solution is insured by the bounds we discussed before, we conclude that a scheme has to be monotone, conservative and consistent in order to approximate the weak entropy solution accurately.

Application of Lax-Wendroff Theorem

We discuss successively the implications of Lax-Wendroff on the three schemes: (i) Godunov's (3.16), (ii) Roe's (3.19) and (iii) Rusanov's (3.21).

Conservation and consistency are shared by all three schemes based on their definition. It remains to check monotonicity.

- Godunov scheme: Computations to check the monotonicity of the Godunov scheme are long because they require a case-by-case analysis. However, since the scheme is derived directly from the Riemann problem's weak entropy solution, monotonicity follows.
- Roe scheme: For the Roe scheme, it is clear that the conditions are not satisfied since, for example, if $D_{\frac{1}{2}} \geq 0$, $\frac{\partial F}{\partial U_j^n} = U_j^n$, which is not necessarily of a particular sign.
- Rusanov scheme: Finally, the Rusanov scheme is indeed monotone since

$$\frac{\partial F}{\partial a} = \frac{a}{2} + \frac{\max(a, b)}{2} \geq 0$$

$$\frac{\partial F}{\partial b} = \frac{b}{2} - \frac{\max(a, b)}{2} \leq 0$$

and

$$\left| \frac{\partial F}{\partial U_j^n} \right| + \left| \frac{\partial F}{\partial U_{j+1}^n} \right| = \frac{|U_j^n| + |U_{j+1}^n|}{2} \leq \max_j |U_j^n| \leq \frac{\Delta x}{\Delta t}$$

where the last inequality is due to the CFL condition (3.15).

CHAPTER 4

IMPLEMENTATIONS: DIRECT SOLVERS

We now test the schemes discussed in chapter 3 using MATLAB taking different cases and compare them with the theoretical results.

4.1 Linear Transport Equation

We test the finite difference schemes on the linear transport equation (2.1) and present tables with relative errors between approximated and exact solutions along with graphs.

4.1.1 Constant velocity

Starting with constant velocity a , we test the central and upwind finite difference schemes taking the following settings:

- $u_0(x) = \sin(2\pi x)$
- $a = 1$
- $\frac{\Delta t}{\Delta x} = 0.9$ (applying CFL condition (3.10))
- $\Delta x = 1/16, 1/32, 1/64$ and $1/128$
- end times $T = 5$ and $T = 10$

Tables 4.1 and 4.2 show the relative errors $\frac{\|u_{approx} - u_{exact}\|}{\|u_{exact}\|}$ between the exact and approximated solutions using the central and upwind schemes for end times $T = 5$ and $T = 10$ respectively. Figure 4.1 shows their graphs with $\Delta x = 1/128$ and $T = 5$ and $\Delta x = 1/32$ and $T = 10$.

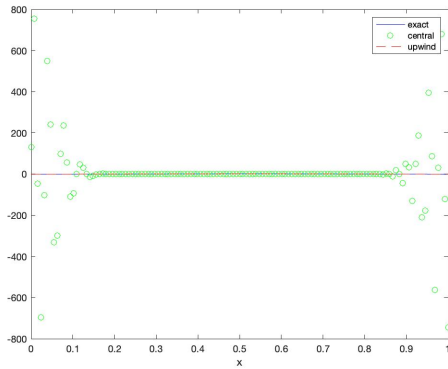
The graphs and tables agree with the analysis done in chapter 3 about the instability of the central scheme as the relative error between the approximated

Δ_x	Central	Upwind
1/16	4.3131e-01	1.7858e-01
1/32	4.1628e-01	9.4858e-02
1/64	2.0682e+00	4.9241e-02
1/128	2.2811e+02	2.5192e-02

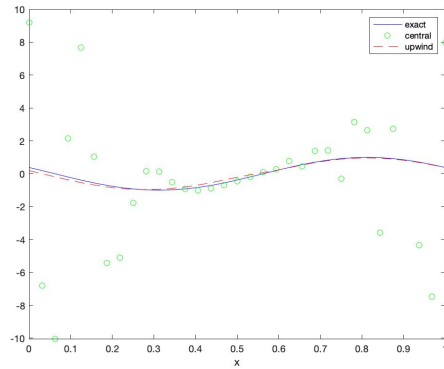
Table 4.1: Relative error between exact and approximated solutions for end time $T=5$

Δ_x	Central	Upwind
1/16	1.7488e+01	4.3059e-01
1/32	1.8349e+03	2.3584e-01
1/64	7.6742e+07	1.2395e-01
1/128	4.6142e+17	6.3662e-02

Table 4.2: Relative error between exact and approximated solutions for end time $T=10$



(a) $\Delta x = 1/128, T = 5$



(b) $\Delta x = 1/32, T = 10$

Figure 4.1: Solutions using finite difference schemes with $a = 1$ and $\frac{\Delta t}{\Delta x} = 0.9$

and exact solution are very large.

The upwind scheme appeared to be stable, approximating the solution accurately and the relative error was decreasing as we decreased the mesh size.

Comparing the two tables with the two end times 5 and 10, we notice that as we increase the time, the instability of the central scheme was more clear and the upwind scheme was less accurate.

Necessity of the CFL condition

In the next two graphs 4.2, we show the necessity of the CFL condition (3.10). In 4.2a, we set $a = 5$ and consequently $\frac{\Delta t}{\Delta x} = \frac{0.9}{5}$, keeping the CFL condition satisfied. As a result, the upwind scheme was still stable. On the other hand, in 4.2b, we kept $a = 1$ but took $\frac{\Delta t}{\Delta x} = 1.3$. We see clearly that some unstability appeared in the upwind scheme.

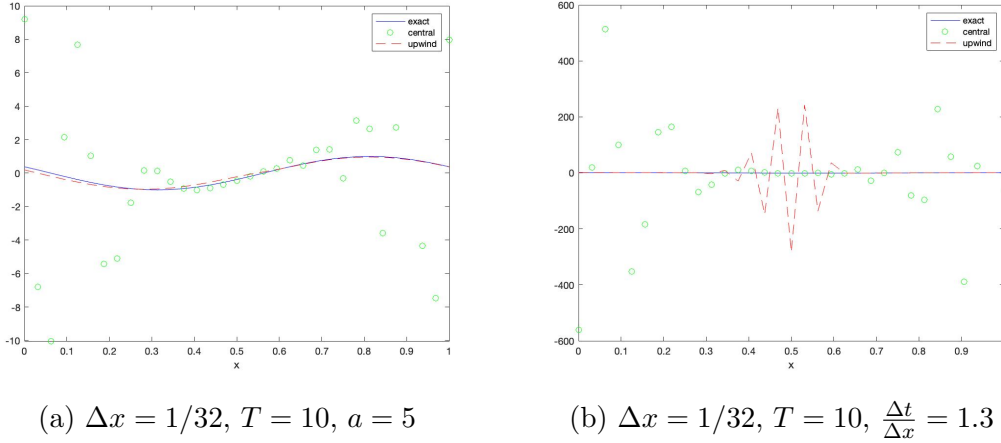


Figure 4.2: Finite difference schemes for other cases

4.1.2 Non constant velocity

In the non constant velocity case, we take the CFL condition to be

$$|\max_j(a(x_j, t^n))| \frac{\Delta t}{\Delta x} \leq 1.$$

The graphs in Figure 4.3 present the two examples $a(x, t) = t$ and $a(x, t) = x$.

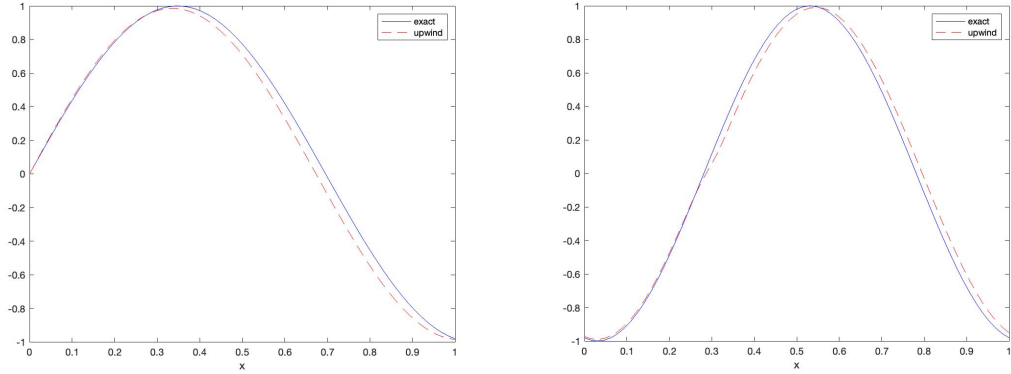
The graphs show that the upwind scheme was still stable under the CFL condition even for non constant velocities.

4.2 Non Linear Transport equation: Burgers' equation

We now test the three finite volume schemes (3.16), (3.21) and (3.19) presented in chapter 3 on Burgers' equation $u_t + (\frac{u^2}{2})_x = 0$ with discontinuous initial data

$$u_0(x) = \begin{cases} u_l & \text{if } x \leq 0 \\ u_r & \text{if } x > 0 \end{cases}$$

and we consider two cases: $u_l > u_r$ and $u_l < u_r$.



(a) $a(x, t) = x$, $\Delta t = 0.9\Delta x$

(b) $a(x, t) = t$, $\Delta t = 0.9\frac{\Delta x}{T}$

Figure 4.3: Finite difference upwind scheme solution for non constant velocity, $\Delta x = 1/64$, end time $T = 5$

While testing we will take the following settings

- $\Delta x_1 = 1/16, 1/32, 1/64$ and $1/128$
- end times $T = 5$ and $T = 10$
- $\frac{\Delta t}{\Delta x} = 0.3$ (applying the CFL condition (3.15))

4.2.1 Case 1: $u_l > u_r$

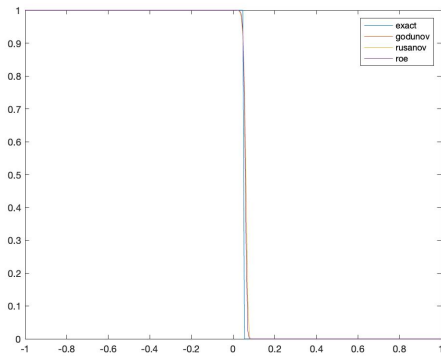
In this first case we took $u_l = 1$ and $u_r = 0$. Figures 4.4 and tables 4.3 and 4.4 show the graphs of the three schemes with the exact solution and the relative errors $\frac{\|u_{approx} - u_{exact}\|}{\|u_{exact}\|}$.

Δx	Godunov	Rusanov	Roe
1/16	1.6417e-01	1.6680e-01	1.6417e-01
1/32	1.0752e-01	1.1517e-01	1.0752e-01
1/64	6.4422e-02	7.6398e-02	6.4422e-02
1/128	6.7850e-02	7.8513e-02	6.7850e-02

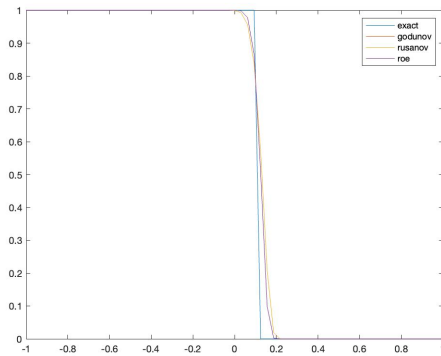
Table 4.3: Relative error between exact and approximated solutions for end time $T=5$

Δ_x	Godunov	Rusanov	Roe
1/16	1.4778e-01	1.5829e-01	1.4778e-01
1/32	8.8540e-02	1.0500e-01	8.8540e-02
1/64	9.3560e-02	1.0826e-01	9.3560e-02
1/128	1.0503e-01	1.2071e-01	1.0503e-01

Table 4.4: Relative error between exact and approximated solutions for end time $T=10$



(a) $\Delta x = 1/128, T = 5$



(b) $\Delta x = 1/32, T = 10$

Figure 4.4: Finite volume schemes for $u_l > u_r$

Concerning this case, the tables and graphs show that all three schemes approximated the exact solution accurately. The relative errors between the schemes are also comparable.

Decreasing the mesh size did decrease the relative error for end time $T = 5$. However, as in the linear case, increasing the end time to $T = 10$ lowered the accuracy of the schemes.

4.2.2 Case 2: $u_l < u_r$

Here we will take $u_l = -1$ and $u_r = 1$ and perform the same tests as before.

Δ_x	Godunov	Rusanov	Roe
1/16	7.5156e-02	1.0553e-01	1.9892e-01
1/32	6.8700e-02	8.5587e-02	1.8586e-01
1/64	5.3559e-02	6.2844e-02	1.8303e-01
1/128	4.3950e-02	4.9120e-02	1.8250e-01

Table 4.5: Relative error between exact and approximated solutions for end time $T=5$

Δx	Godunov	Rusanov	Roe
1/16	9.9538e-02	1.2400e-01	2.6929e-01
1/32	7.8015e-02	9.1539e-02	2.6661e-01
1/64	6.4180e-02	7.1729e-02	2.6650e-01
1/128	5.5430e-02	5.9766e-02	2.6677e-01

Table 4.6: Relative error between exact and approximated solutions for end time $T=10$

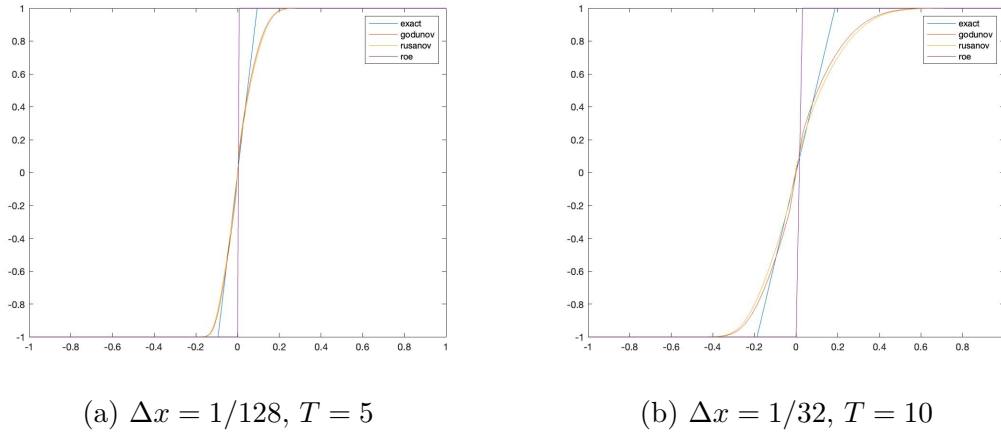


Figure 4.5: Finite volume schemes for $u_l < u_r$

The first thing to notice in the graphs is that the Roe scheme failed completely at approximating the exact solution since its result was a step function instead of a continuous one like the weak entropy solution we found in chapter 2. However, the Godunov and Rusanov schemes were much more accurate. These implementations reinforce the results we got on the convergence of the schemes in chapter 3.

The table also shows that the errors between the Godunov and Rusanov schemes were comparable, however, the Roe scheme was far from them.

Lastly, decreasing the mesh size had a positive effect on the Godunov and Rusanov schemes for both end times $T = 5$ and $T = 10$.

CHAPTER 5

ON THE INVERSE PROBLEM

5.1 Formulation

Our main goal is to recover the initial data u_0 knowing a measured solution $u_{meas}(T) \rightarrow \mathbb{R}$ at some time T of the transport phenomenon. Usually, $u_{meas}(T)$ is defined on a discrete subset of $(-\infty, \infty)$, but we assume it can simply be extended to a function on $(-\infty, \infty)$, to become $\mathcal{E}u_{meas}(T) : (-\infty, \infty) \rightarrow \mathbb{R}$.

We now let \mathcal{I} be the set of initial conditions. In our case, we will take:

$$\text{either } \mathcal{I} = C(-\infty, \infty) \text{ or } \mathcal{I} = \{v \in C(-\infty, \infty) \mid v > 0\},$$

For a function $v \in \mathcal{I}$, let $u_D(v; x, t)$ be the unique direct solution of the problem modeled by the transport equation. The continuous inverse problem is equivalent to an optimization problem with the following objective function:

$$v \in \mathcal{I} : G(v) = \|u_{meas}(T) - u_D(v; \cdot, T)\|_2^2. \quad (5.1)$$

We then seek $u_0 \in \mathcal{I}$, such that:

$$G(u_0) = \min_{v \in \mathcal{I}} \{G(v)\} \quad (5.2)$$

To construct an algorithm for solving (5.2), the strategy is to discretize \mathcal{I} and $f(\cdot)$ and then obtain a discrete optimization problem on subsets of finite-dimensional vector spaces. Specifically, we let $\mathcal{I}_{\Delta x}$, be a finite dimensional subset of \mathcal{I} , i.e., $\mathcal{I}_{\Delta x} \subset \mathcal{I}$.

Furthermore, for $v \in \mathcal{I}_{\Delta x}$, $u_{\{D, \Delta x, \Delta t\}}(v; \cdot, t)$ is the discrete solution obtained using a solver for the direct problem. This allows us defining:

$$v \in \mathcal{I}_{\Delta x} : G_{\Delta x, \Delta t}(v) = \|u_{meas}(T) - u_{\{D, \Delta x, \Delta t\}}(v; \cdot, T)\|_2^2. \quad (5.3)$$

We then seek $u_{\{0, \Delta x, \Delta t\}} \in \mathcal{I}_{\Delta x}$, such that:

$$G_{\Delta x, \Delta t}(u_{\{0, \Delta x, \Delta t\}}) = \min_{v \in \mathcal{I}_{\Delta x}} \{G_{\Delta x, \Delta t}(v)\} \quad (5.4)$$

To solve (5.4), we use a built-in MATLAB function `fmincon`, which finds the value that minimizes $G_{\Delta x, \Delta t}$ by taking as input: $G_{\Delta x, \Delta t}$ itself along with a guess of an initial data for $v_0 \in \mathcal{I}_{\Delta x}$.

5.2 Implementation

A lot of variables come into the strategy mentioned above, namely, the mesh size, the end time T , the tolerance, the algorithm used in the function `fmincon`, and whether the optimal u_0 is constrained or not.

To this end, we run tests taking into consideration all the variables above, on both the linear and non linear transport equations and present the results in tables alongside the graphs of some cases.

5.2.1 *Linear transport equation with constant velocity*

Starting with the linear transport equation, we test the inverse problem with the following settings

- upwind finite difference scheme (3.6)
- $u_{0,exact} = \sin(2\pi x)$ (which we aim to recover)
- $a = 1$
- $\Delta x = 1/16, 1/32, 1/64$ and $1/128$
- $\frac{\Delta t}{\Delta x} = 0.9$
- tolerance for `fmincon` $tol = 10^{-6}$ and 10^{-8}
- constrains of the solution of `fmincon`: upper bound $ub = 1$ and lower bound $lb = -1$
- algorithms for `fmincon`: “interior-point” and “sqp”
- initial guess for `fmincon`: random vector containing numbers drawn from a standard normal distribution generated by built-in MATLAB function `randn`
- end time $T = 5$

Tables 5.1 and 5.2 show the relative error between the approximated and exact initial data $\frac{\|u_{0,approx} - u_{0,exact}\|}{\|u_{0,exact}\|}$, the value of the minimized objective function (5.3) and the number of iterations that the algorithm took to get to the result.

Δx	$tol = 10^{-6}$			$tol = 10^{-8}$		
	rel. error	G(v)	iterations	rel. error	G(v)	iterations
1/16	1.7320e-07	5.9064e-14	16	1.3277e-08	3.4137e-16	17
1/32	1.2611e-06	3.7366e-13	48	3.8897e-08	9.4361e-16	50
1/64	3.2522e-01	1.2176e-03	45	3.2522e-01	1.2176e-03	45
1/128	9.7430e-01	1.0197e-02	23	9.7430e-01	1.0197e-02	23

Table 5.1: Relative error between exact and approximated u_0 and value of objective function using “interior-point” algorithm and unconstrained solution

Δx	$tol = 10^{-6}$			$tol = 10^{-8}$		
	rel. error	G(v)	iterations	rel. error	G(v)	iterations
1/16	4.8423e-04	3.1693e-07	35	4.3661e-05	2.5579e-09	53
1/32	8.8087e-04	2.2573e-07	72	4.1402e-04	4.8479e-08	88
1/64	7.6949e-03	4.6840e-05	45	7.6949e-03	4.6840e-05	45
1/128	2.8311e-02	7.9408e-04	23	2.8311e-02	7.9408e-04	23

Table 5.2: Relative error between exact and approximated u_0 and value of objective function using “interior-point” algorithm and constrained solution

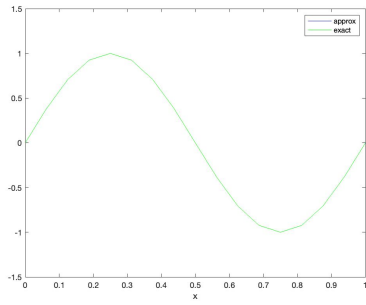
For $\Delta x = 1/16$ and $1/32$ the results were better for the unconstrained solution than the case of constrained solution. However for $\Delta x = 1/64$ and $1/128$ the problem did not converge for the unconstrained solution.

Looking at the results of the constrained solution, we see that the problem converged for all cases of Δx , even though for the first two cases of Δx , the results were better with unconstrained solution.

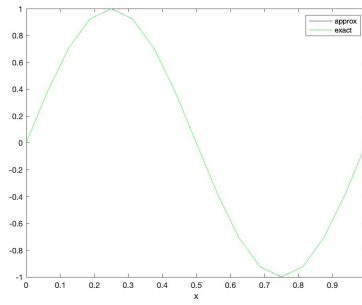
We also notice that decreasing the mesh size made the results less accurate, which is the opposite case for the direct problem.

The graphs in 5.1 show the approximated initial data (in blue) and exact initial data (in green) for all cases of Δx , constrained and unconstrained solution and $tol = 10^{-6}$.

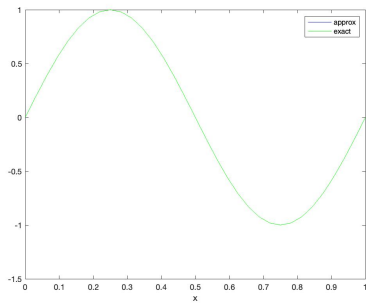
We can visually see where the problem did not converge by the unstable plots of the approximated initial data.



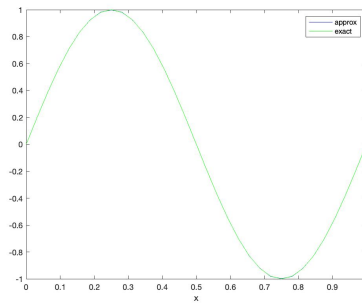
(a) $\Delta x = 1/16$, unconstrained



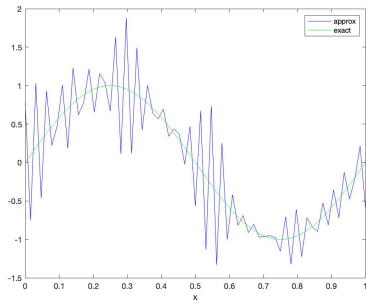
(b) $\Delta x = 1/16$, constrained



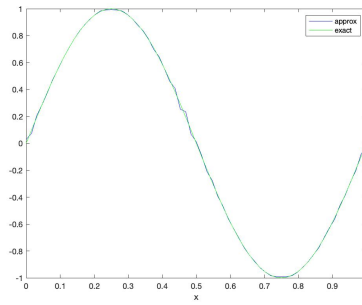
(c) $\Delta x = 1/32$, unconstrained



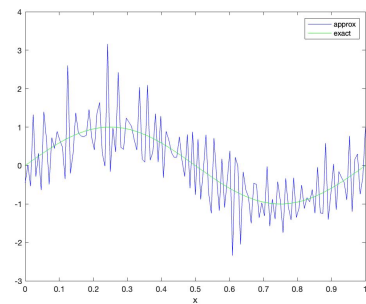
(d) $\Delta x = 1/32$, constrained



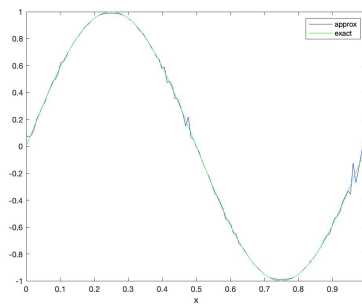
(e) $\Delta x = 1/64$, unconstrained



(f) $\Delta x = 1/64$, constrained



(g) $\Delta x = 1/128$, unconstrained



(h) $\Delta x = 1/128$, constrained

Figure 5.1: u_0 exact and approximated using “interior-point” and $tol = 10^{-6}$

We now test the problem using the “sqp” algorithm for `fmincon`. The results are shown in tables 5.3 and 5.4.

Δx	$tol = 10^{-6}$			$tol = 10^{-8}$		
	rel. error	G(v)	iterations	rel. error	G(v)	iterations
1/16	2.5916e-07	9.5787e-14	16	1.4915e-08	6.6915e-16	17
1/32	3.8311e-06	4.7605e-12	47	5.5071e-08	9.6235e-16	49
1/64	7.1830e-02	4.0946e-05	98	7.1830e-02	4.0946e-05	98
1/128	9.1832e-01	3.1038e-04	99	9.1832e-01	3.1038e-04	99

Table 5.3: Relative error between exact and approximated u_0 and value of objective function using “sqp” algorithm and unconstrained solution

Δx	$tol = 10^{-6}$			$tol = 10^{-8}$		
	rel. error	G(v)	iterations	rel. error	G(v)	iterations
1/16	3.3941e-07	3.0489e-13	19	3.3941e-07	3.0489e-13	20
1/32	9.7605e-06	2.4758e-11	52	3.0428e-06	2.5955e-12	59
1/64	5.4124e-02	2.3358e-05	98	5.4124e-02	2.3358e-05	98
1/128	3.1227e-01	6.3740e-05	99	3.1227e-01	6.3740e-05	99

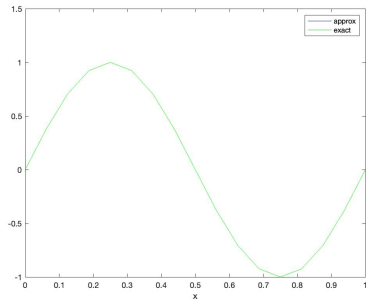
Table 5.4: Relative error between exact and approximated u_0 and value of objective function using “sqp” algorithm and constrained solution

The “sqp” algorithm made a slight improvement in the case of unconstrained solution compared to “interior-point”, however, it did not converge for the case of constrained solution when $\Delta x = 1/128$.

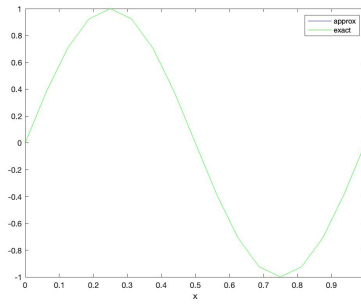
The results for unconstrained and constrained were comparable using “sqp”. Hence constraining the solution did not make a noticeable difference.

Even on the level of the number of iterations, we conclude that “interior-point” was more efficient in general.

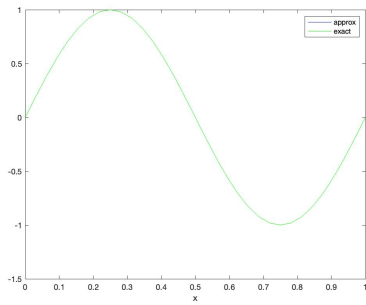
The graphs in Figure 5.2 show the results using “sqp” for $tol = 10^{-6}$.



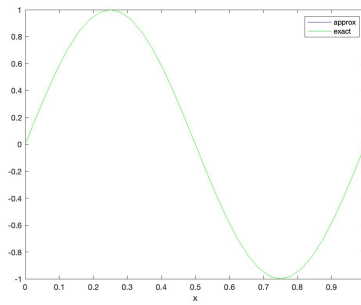
(a) $\Delta x = 1/16$, unconstrained



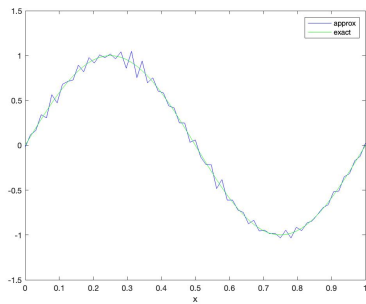
(b) $\Delta x = 1/16$, constrained



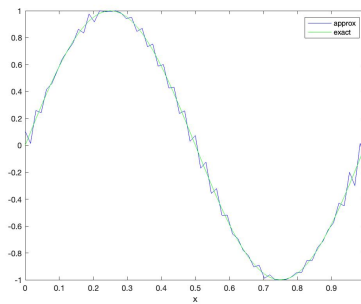
(c) $\Delta x = 1/32$, unconstrained



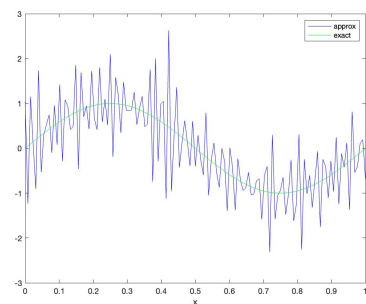
(d) $\Delta x = 1/32$, constrained



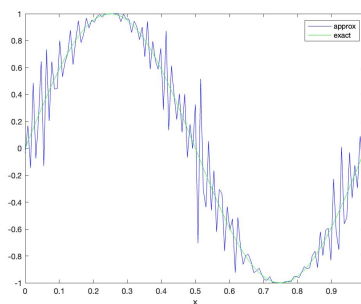
(e) $\Delta x = 1/64$, unconstrained



(f) $\Delta x = 1/64$, constrained



(g) $\Delta x = 1/128$, unconstrained



(h) $\Delta x = 1/128$, constrained

Figure 5.2: u_0 exact and approximated using “sqp” and $tol = 10^{-6}$

5.2.2 Non linear transport equation

We now test the inverse problem on Burgers' equation with the following settings

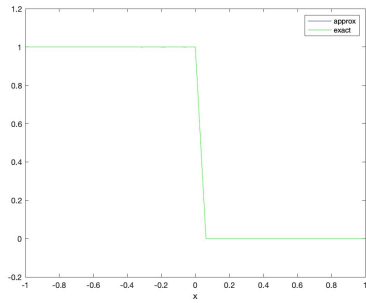
- Godunov finite volume scheme
- $u_{0,exact} = \begin{cases} 1 & \text{if } x \leq 0 \\ 0 & \text{if } x > 0 \end{cases}$
- $\Delta x = 1/16, 1/32, 1/64$ and $1/128$
- $\frac{\Delta t}{\Delta x} = 0.3$
- tolerance for `fmincon` $tol = 10^{-6}$ and 10^{-8}
- constraints of the solution of `fmincon` : upper bound $ub = 1$ and lower bound $lb = 0$
- algorithms for `fmincon`: “interior-point” and “sqp”
- initial guess for `fmincon`: random vector between 0 and 1
- end time $T = 5$

Δx	$tol = 10^{-6}$			$tol = 10^{-8}$		
	rel. error	G(v)	iterations	rel. error	G(v)	iterations
1/16	8.3587e-05	3.5380e-10	68	2.4888e-08	3.2002e-16	88
1/32	1.4007e-01	7.0663e-05	45	1.4007e-01	7.0663e-05	45
1/64	2.3311e-01	2.1057e-03	22	2.3311e-01	2.1057e-03	22
1/128	3.6001e-01	7.6187e-01	11	3.6001e-01	7.6187e-01	11

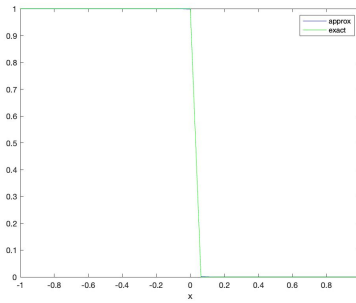
Table 5.5: Relative error between exact and approximated u_0 and value of objective function using “interior-point” algorithm and unconstrained solution

Δx	$tol = 10^{-6}$			$tol = 10^{-8}$		
	rel. error	G(v)	iterations	rel. error	G(v)	iterations
1/16	7.1632e-04	4.2220e-06	48	6.4162e-05	3.3787e-08	70
1/32	2.3160e-03	4.1528e-05	45	2.3160e-03	4.1528e-05	45
1/64	6.2079e-02	4.8933e-02	23	6.2079e-02	4.8933e-02	23
1/128	1.6676e-01	2.2477e+00	11	1.6676e-01	2.2477e+00	11

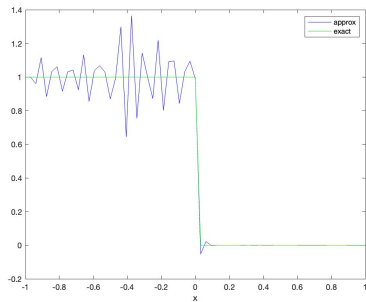
Table 5.6: Relative error between exact and approximated u_0 and value of objective function using “interior-point” algorithm and constrained solution



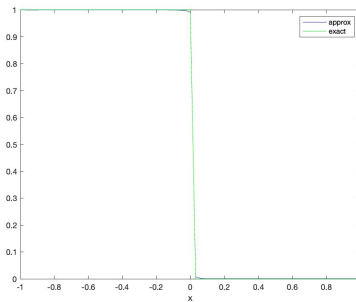
(a) $\Delta x = 1/16$, unconstrained



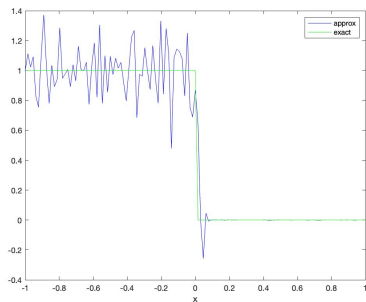
(b) $\Delta x = 1/16$, constrained



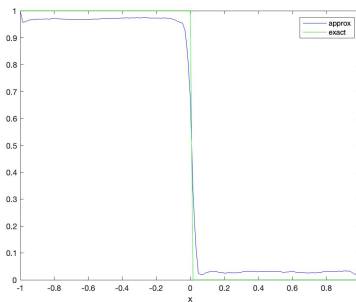
(c) $\Delta x = 1/32$, unconstrained



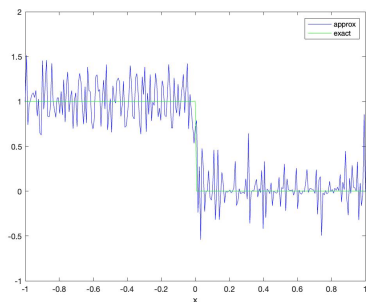
(d) $\Delta x = 1/32$, constrained



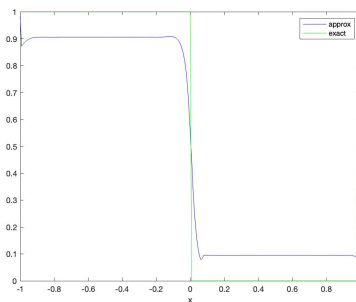
(e) $\Delta x = 1/64$, unconstrained



(f) $\Delta x = 1/64$, constrained



(g) $\Delta x = 1/128$, unconstrained



(h) $\Delta x = 1/128$, constrained

Figure 5.3: u_0 exact and approximated using “interior-point” and $tol = 10^{-6}$

For the non linear case, we see clearly in tables 5.5 and 5.6 and figure 5.3 that unconstraining the solution of `fmincon` was highly inefficient since the problem did not converge for $tol = 10^{-6}$ in all cases except for $\Delta x = 1/16$. Constraining the solution made a much bigger impact, since the problem converged in all cases except for $\Delta x = 1/128$.

We now test using the “sqp” algorithm.

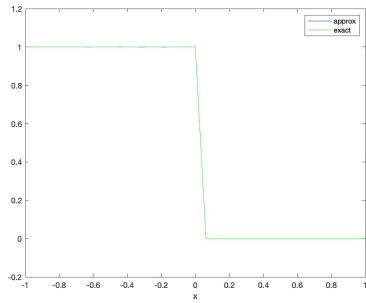
Δx	$tol = 10^{-6}$			$tol = 10^{-8}$		
	rel. error	G(v)	iterations	rel. error	G(v)	iterations
1/16	6.1785e-05	1.6091e-10	69	6.7757e-06	1.7569e-12	77
1/32	8.8941e-02	2.8699e-06	98	8.8941e-02	2.8699e-06	98
1/64	2.2049e-01	4.8363e-06	99	2.2049e-01	4.8363e-06	99
1/128	3.0951e-01	2.1038e-02	99	3.0951e-01	2.1038e-02	99

Table 5.7: Relative error between exact and approximated u_0 and value of objective function using “sqp” algorithm and unconstrained solution

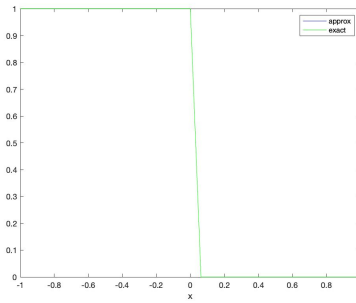
Δx	$tol = 10^{-6}$			$tol = 10^{-8}$		
	rel. error	G(v)	iterations	rel. error	G(v)	iterations
1/16	3.3005e-08	1.5425e-15	17	3.3005e-08	1.5425e-15	18
1/32	7.9792e-08	2.1439e-14	23	7.9792e-08	2.1439e-14	23
1/64	2.6875e-07	3.8795e-13	41	4.4435e-09	1.0605e-16	43
1/128	2.0340e-02	7.9649e-08	99	2.0340e-02	7.9649e-08	99

Table 5.8: Relative error between exact and approximated u_0 and value of objective function using “sqp” algorithm and constrained solution

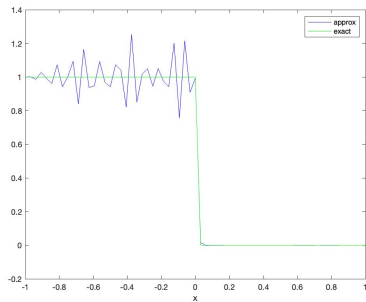
The “sqp” algorithm was more accurate for the non linear case, a slight improvement was noticeable in the unconstrained solution (Table 5.7). However, we see better results in the constrained case (Table 5.8) compared to the “interior-point” algorithm. Here, the problem converged for all mesh sizes that we tested. The graphs in 5.4 show the high level of accuracy of the “sqp” algorithm when constraining the solution since the plots of the exact and approximated initial data were overlapping.



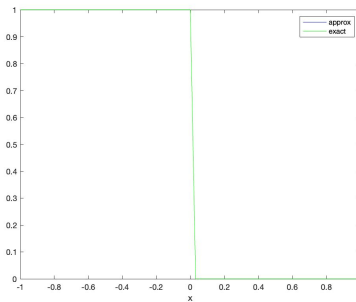
(a) $\Delta x = 1/16$, unconstrained



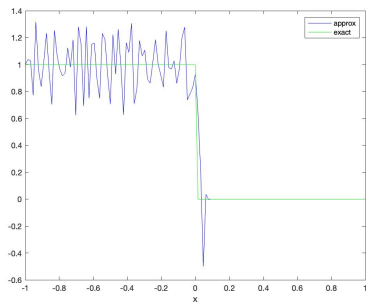
(b) $\Delta x = 1/16$, constrained



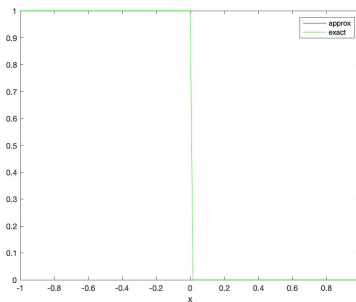
(c) $\Delta x = 1/32$, unconstrained



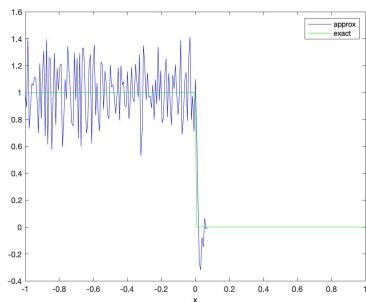
(d) $\Delta x = 1/32$, constrained



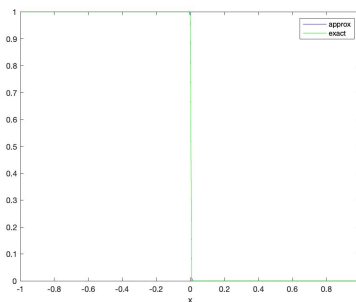
(e) $\Delta x = 1/64$, unconstrained



(f) $\Delta x = 1/64$, constrained



(g) $\Delta x = 1/128$, unconstrained



(h) $\Delta x = 1/128$, constrained

Figure 5.4: u_0 exact and approximated using “sqp” and $tol = 10^{-6}$

5.3 Concluding Remarks on the Inverse Problem

First, it is worth mentioning that decreasing the mesh size in the inverse problem had a negative effect on the level of accuracy of the algorithms, which was not the case for the direct problem.

Regarding the choice of algorithms, we would have to constrain the solution of `fmincon` for more accurate results. For the linear case, “interior-point” is a better a choice, however, for the non linear case, “sqp” is the recommended algorithm.

CHAPTER 6

CONCLUSION

A thorough analysis of the transport equation was conducted in this thesis. Aside from the theoretical part, a numerical analysis was also conducted and implemented showing different schemes and their efficiencies.

Based on these results, the inverse problem was formulated and implemented, allowing us to recover the initial data from a measured data at some time T .

We were able to test different algorithms and conclude which combinations were the best for both the linear and non linear transport equations.

For future work, one could implement the inverse problem on the velocity of the equation and recover it as well. A theoretical study of the inverse problem could be done too.

BIBLIOGRAPHY

- [1] R. J. DIPERNA and P. L. LIONS, “Ordinary differential equations, transport theory and sobolev spaces,” English, *Inventiones mathematicae*, vol. 98, no. 3, pp. 511–547, 1989.
- [2] L. Ambrosio, “Transport equation and cauchy problem for bv vector fields,” English, *Inventiones mathematicae*, vol. 158, no. 2, pp. 227–260, 2004.
- [3] B. Perthame, “Equations de transport non lineaires et systemes hyperboliques, theorie et methodes numeriques,” French, 2003-2004.
- [4] F. Baccelli, D. McDonald, and J. Reynier, “A mean-field model for multiple tcp connections through a buffer implementing red,” *Performance Evaluation*, vol. 49, no. 1, pp. 77–97, 2002, Performance 2002, ISSN: 0166-5316. DOI: [https://doi.org/10.1016/S0166-5316\(02\)00136-0](https://doi.org/10.1016/S0166-5316(02)00136-0). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0166531602001360>.
- [5] E. Godlewski and P.-A. Raviart, *Numerical Approximation of Hyperbolic Systems of Conservation Laws*, English, 2nd 2021. New York, NY: Springer New York, vol. 118, ISBN: 0066-5452.
- [6] S. Mishra, U. Skre Fjordholm, and R. Abgrall, *Numerical methods for conservation laws and related equations*, English.
- [7] E. Godlewski and P.-A. Raviart, *Hyperbolic Systems of Conservation Laws*, English. Ellipes, 1991.
- [8] J. GOODMAN and Z. XIN, “Viscous limits for piecewise smooth solutions to systems of conservation laws,” English, *Archive for rational mechanics and analysis*, vol. 121, no. 3, pp. 235–265, 1992.