[0][0]

Cand. Scient. Thesis in Applied and Industrial Mathematics (AIM), August 1994

ON THE ACCURACY OF A DIMENSIONAL SPLITTING METHOD FOR SCALAR CONSERVATION LAWS

by

Kenneth A. Hvistendahl Karlsen

Department of Mathematics, Mathematics Division,
University of Oslo, Norway.

Preface

Denne hovedoppgaven ble innlevert for å oppfylle det skriftlige kravet til cand. scient. graden i Anvendt og Industriell matematikk (AIM). Selve oppgaven består av to deler; en teoretisk og en eksperimentell del. Den teoretiske delen ble til i perioden April – Desember, 1993. Den eksperimentelle delen, som besto av programmering og gjennomføring av nummeriske eksperimenter, ble til i perioden Januar – Mai, 1994.

Et av resultatene fra den teoretiske delen av hovedoppgaven er et eksplisitt feilestimat (Teorem 3.2) for en dimensjonssplittingsalgoritme for å løse skalare konserveringslover i flere romdimensjoner. Dimensjonsplitting er en teknikk som ble innført på 50-tallet av Godunov i forbindelse med gass-dynamikk. I 1980 viste Crandall og Majda [3] at dimensjonsplitting kan brukes til å generere tilnærmede løsninger til skalare konserveringslover i flere dimensjoner som, i grensen når diskretiseringsparameteren går mot null, konvergerer mot riktig løsning (Teorem 3.1). Beviset av dette resultatet er basert på et kompakthetsargument og gir dermed ikke noe feilestimat. For å kunne fullføre min egen oppgave var et slikt feilestimat helt avgjørende. Den eneste løsningen var dermed å produsere et estimat selv. Dette var et problem som til å begynne med virket uoverkommelig. Etter en del arbeid ble nå likevel dette estimatet ferdig i månedskifte August/September, 1993. Så det var med blandede følser at jeg mottok en artikkel skrevet av Zhen-Huan Teng [21], utgitt i Oslo Mars/April 1994, hvor det ønskede feilestimatet ble vist. Beviset til Teng er essensielt det samme som det jeg selv kom frem til, men det understrekes at mitt eget arbeid ble gjort fullstendig uavhengig av Teng's arbeid.

Det er mange personer jeg burde takke i forbindelse med tilblivelsen av denne hovedoppgaven. Først og fremst vil jeg få takke min veileder dr. Nils Henrik Risebro for å ha introdusert meg for det fasinerende feltet bestående av hyperbolske konserveringslover, og for hele tiden under arbeidet med oppgaven å ha vært en konstant støtte og inspirasjon. Jeg vil også benytte anledningen til å takke Matematisk Institutt, Avd. A. for økonomisk støtte som gjorde det mulig for meg å delta på: i) Workshop on Systems of Hyperbolic Conservation Laws, Oslo (Moss) August 23 – 25, 1993. ii) Summer School in Lie groups and differential equations, Nordfjordeid 27/6 – 9/7, 1994. Spesielt gav deltagelse på (i) stor inspirasjon til videre arbeid med hyperbolske konserveringslover.

En rekke personer har også lest gjennom oppgaven med henblikk på forbedring av syntaktiske, pragmatiske og semantiske aspekter ved teksten. I denne forbindelse vil jeg gjerne få takke

Kristine Osnes (som jeg for første gang ble kjent med under oppholdet på Nordfjordeid), Anne Skappel og Nils Henrik Risebro.

Til slutt en takk til alle venner og kjente som har vært med på gjøre studietiden til noe mye mere enn bare forelesninger, lesesalsbesøk, eksamener, osv. Jeg tror helt og holdent at uten muligheten til å koble av med ikke faglig relaterte begivenheter ville denne oppgaven ikke ha vært ferdig i dag.

Kenneth Hvistendahl Karlsen Blindern, August 1994.

Abstract

The purpose with this paper is to investigate, both mathematically and experimentally, the properties of one particular numerical method for (non-linear) scalar conservation laws in two space dimensions. The numerical method under consideration is based on the use of dimensional splitting and front tracking to solve the one-dimensional equations. The method of dimensional splitting is a popular technique used to extend one-dimensional numerical methods to multidimensional problems. The one-dimensional method we will apply is front tracking as first introduced by Dafermos [4], and later extended and proved well-defined by Holden, Holden and Høegh-Krohn [5]. Dimensional splitting together with front tracking was first proposed as a numerical method by Holden and Risebro in [6]. In that paper they showed that the method generates approximate solutions which, in the limit as the discretization parameters tend to zero, converge to the unique physical solution of the problem. Our paper is a direct continuation of theirs and can be divided into two main parts; i) a theoretical part and ii) an experimental part.

i) In Holden and Risebro [6] no estimate of the convergence rate is provided. We obtain an explicit estimate showing how "fast" the numerical solution tends to the exact solution, i.e., an error estimate. More precisely, if $u = u(x, y, t)$ denotes the exact solution and $u_\eta^n = u_\eta^n(x, y)$ the solution generated by method of dimensional splitting with the exact one-dimensional solutions operators replaced by the front tracking scheme, then we prove the following theorem:

Theorem 1 *If* $u^0 \in L^1(\mathbf{R}^2) \cap L^\infty(\mathbf{R}^2) \cap BV(\mathbf{R}^2)$ *and* $n\Delta t = t$, *then for some finite constant* C,

$$\|u(\cdot,\cdot,t) - u_\eta^n(\cdot,\cdot)\|_1 \leq C((\Delta t)^{1/2} + (\Delta x)^{1/2} + \delta),$$

where Δx *is the space step,* Δt *is the time step and* δ *is the parameter measuring the polygonal approximation of the flux functions.*

In addition we also obtain an error estimate for the "semi-discrete" method: if $u_{\Delta t}^n = u_{\Delta t}^n(x, y)$ denotes the solution generated by the method of dimensional splitting with exact one-dimensional solution operators, then the following theorem holds:

Theorem 2 *If* $u^0 \in L^1(\mathbf{R}^2) \cap L^\infty(\mathbf{R}^2) \cap BV(\mathbf{R}^2)$ *and* $n\Delta t = t$, *then for some finite constant* C,

$$\|u(\cdot,\cdot,t) - u_{\Delta t}^n(\cdot,\cdot)\|_1 \leq C(\Delta t)^{1/2},$$

where Δt *is the time step.*

ii) To illustrate the theoretical results we implement the method on a computer and use it on some test problems to compute (numerically) the rate of convergence. The results from the numerical experiments are used to say something about the optimality of error estimates stated above.

Our paper consists of four chapters which may be summarized as follows: In Chapter 1 we give a brief introduction to the mathematical theory involving (one-dimensional) scalar conservation laws. We also list some well known numerical methods and their properties. In writing this chapter, three sources of information have been used extensively; namely; Langtangen and Tveito [11], LeVeque [13] and Smoller [20]. We emphazise that the introduction is very brief, and those interested in details and/or a more comprehensive guide to the subject of hyperbolic conservation laws should confer the references mentioned above and, by all means, almost any of the other references listed in the bibliography. In Chapter 2 we introduce and describe in detail the method of front tracking. Chapter 3 is devoted to a) introducing the method of dimensional splitting for front tracking and b) rigorously proving Theorem 1 and 2. Finally, in Chapter 4 we carry out the numerical experiments and report the results.

Contents

List of Figures

List of Tables

Chapter 1

Hyperbolic conservation laws

Enhver erkjennelse som du søker bare for å berike din viten, bare for å samle opp rikdommer i deg, fører deg bort fra din vei. Enhver erkjennelse derimot som du søker for å bli mere moden på veien til menneskeforedling og verdensutvikling, bringer deg et skritt fremover.

– Rudolf Steiner –

1.1 Introduction

Hyperbolic systems of conservation laws are time-dependent systems of partial differential equation (usually nonlinear) with a particular simple structure[1]

$$u_t + f(u)_x = 0. \tag{1.1}$$

Here $f : \mathbf{R}^n \to \mathbf{R}^n$ is the so-called flux function and $u : \mathbf{R} \times \mathbf{R}^+ \to \mathbf{R}^n$ is a n-dimensional vector of conserved quantities, or state variables, such as mass, momentum, and energy in a fluid dynamics problem. Systems of the form (1.1) arise in the study of nonlinear wave phenomena when dissipation effects, such as viscosity, are neglected. It is typical of such systems that they admit discontinuous solutions, i.e., shock waves, so that the equation must be understood in some generalized sense. We start by giving some examples of systems of conservation laws.

Example 1 (The Euler equations): The equations of gas-dynamics for a non-heat conducting gas in Eulerian coordinates can be written in the form

$$\rho_t + (\rho v)_x = 0 \tag{1.2}$$

$$(\rho v)_t + (\rho v^2 + p)_x = 0 \tag{1.3}$$

$$[\rho(\frac{v^2}{2} + e)]_t + [v(\rho(e + \frac{v^2}{2}) + p)]_x = 0. \tag{1.4}$$

[1] For brevity throughout this paper partial derivatives will be denoted by subscripts.

Here ρ is the density of the gas, v the velocity, p the pressure and the equation of state e is a given function $e = e(v, s)$, where s is the specific entropy. The pressure is obtained from the formula $p = e_v$. If we introduce the vector $u \in \mathbf{R}^3$ as

$$u = \begin{bmatrix} \rho \\ \rho v \\ \rho(\frac{v^2}{2} + e) \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix},$$

then the system of equations (1.2)–(1.4) can be written simply as $u_t + f(u)_x = 0$, where

$$f(u) = \begin{bmatrix} \rho v \\ \rho v^2 + p \\ v(\rho(e + \frac{v^2}{2}) + p) \end{bmatrix} = \begin{bmatrix} u_2 \\ u_2^2/u_1 + p(u) \\ u_2/u_1(u_1(e(u_1/u_2, s) + \frac{1}{2}u_2^2/u_1^2) + p(u)) \end{bmatrix}.$$

Equation (1.2), (1.3), and (1.4) expresses the conservation of mass, momentum and energy, respectively. These equations are used extensively in areodynamics, for example in modeling the flow around an aircraft or other vehicle. It is worth noting that, even though this system has been studied extensively over the years, we have no mathematical theory what conserns the two and three dimensional version of this system (which is the most interesting case seen from a "real-life" point of view).

Example 2 (The p-system): The conservation laws

$$\begin{aligned} u_t + p(v)_x &= 0 \\ v_t - u_x &= 0, \end{aligned}$$

are called the p-system. This system arises in gas dynamics and in elasticity theory. It is probably the simplest system of hyperbolic conservation laws and has been thoroughly studied. As for the Euler equations we have no mathematical theory for the two and three dimensional version of this system, but this fact apply to almost any system of hyperbolic conservation laws in several space dimensions. The development of mathematical theory for systems of conservation laws in several space dimensions is one of the exciting challenges for the future in this field.

Example 3 (Polymer flooding): When an underground source of oil is tapped, a certain amount oil flows out on its own due to high pressure in the oil reservoir. After the flow stops, there is typically a large amount of oil still in the ground. One standard method of "secondary recovery" is to pump water into the oil field through some wells, forcing oil out through others. This is an example of a two phase plow, and the flow takes place in a porous medium of rock or sand. A simple model that captures som features of this flow is the scalar Buckley-Leverett equation

$$s_t + f(s)_x = 0,$$

where the flux function is given by

$$f(s) = \frac{k_w(s)/\mu_w}{k_w(s)/\mu_w + k_0(s)/\mu_0}.$$

Here $s = s(x,t)$ denotes the saturation of the aqueous phase, $1-s$ is the oil saturation ($s = 0$ corresponds to pure oil, $s = 1$ to pure water), $k_w = k_w(s)$ is the relative permability for the aqueous phase and μ_w the viscosity which we assume constant. To improve the displacement properties of the water, we may dissolve a polymer in the aqueous phase to increase the viscosity. Displacement of oil by water dissolved a polymer is called polymer flooding. This process is modeled by the system of conservation laws

$$\begin{aligned} s_t + f(s,c)_x &= 0 \\ (sc)_t + (cf(s,c))_x &= 0. \end{aligned}$$

Here $c = c(x,t)$ is the concentration of the polymer in the aqueous phase and s is still the saturation of aqueous phase. The flux function is now given by

$$f(s) = \frac{k_w(s)/\mu_w(c)}{k_w(s)/\mu_w(c)) + k_0(s)/\mu_0},$$

where we observe that the viscosity μ_w now is a function of the polymer concentration c. For a detailed treatment of polymer flooding cf. [18] and references therein.

1.2 Mathematical theory

In this section we will give a brief introduction to the mathematical theory involving hyperbolic conservation laws. Before we do that let us introduce some (standard) notation which will be used extensively throughout this paper. From the language of set theory we will adopt the following terminology:

- $\mathbf{Z}$ = the set of integers
- $\mathbf{Z}^+$ = the set of positive integers (including zero)
- $\mathbf{N}$ = the set of positive integers (not including zero)
- $\mathbf{R}$ = the set of real numbers
- $\mathbf{R}^+$ = the set of positive real numbers (including zero).

If K is a domain in $\mathbf{R}$ (or more generally in $\mathbf{R}^n$), then $L^p(K)$, $1 \leq p \leq \infty$ will denote the classical L^p spaces of real valued functions on K, and the norms on $L^p(K)$ are denoted by $||\cdot||_p$. The localized versions of L^p, consisting of functions on $\mathbf{R}^n$ which are in $L^p(K)$ for

any compact subset K of $\mathbf{R}^n$, are denoted by $L^p_{loc}(\mathbf{R}^n)$. Furthermore, $BV(\mathbf{R}^n)$ denotes the subspace of $L^p_{loc}(\mathbf{R}^n)$ consisting of functions with bounded variation[2], i.e.,

$$BV(\mathbf{R}^n) = \{v \in L^p_{loc}(\mathbf{R}^n) : T.V._x(v) < \infty\},$$

where

$$T.V._x(v) = \sum_{i=0}^{n} \int_{\mathbf{R}^{n-1}} T.V._{x_i}(v(x_1, \ldots, x_{i-1}, \cdot, x_{i+1}, \ldots, x_n, t))\, d^{n-1}x. \tag{1.5}$$

If $v = v(x)$ is a function on $\mathbf{R}$, then $T.V._x(v)$ is defined as

$$T.V._x(v) = \limsup_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_{\mathbf{R}} |h(x) - h(x-\varepsilon)|\, dx.$$

Later we will mainly be interested in working with functions v on $\mathbf{R}^2$, (1.5) then takes the form

$$T.V._{(x,y)}(v) = \int_{\mathbf{R}} T.V._x(v(\cdot, y))\, dy + \int_{\mathbf{R}} T.V._y(v(x, \cdot))\, dx.$$

The class of Lipschitz continuous functions on a domain $K \subset \mathbf{R}$ is denoted by $Lip(K)$. More precisely

$$Lip(K) = \{v \in L^\infty(\mathbf{R}) : \|v\|_{Lip} < \infty\},$$

where

$$\|v\|_{Lip} = \sup_{x \neq y} \frac{|v(x) - v(y)|}{|x - y|}.$$

We denote by $C^k(K)$ the space of all real valued functions on K possessing continuous (partial) derivatives of order $\leq k$, and we set $C^\infty(K) = \bigcap_1^\infty C^k(K)$. The usual $\mathcal{O}$-notation will also be used frequently throughout this paper. We recall that if a function $v = v(x)$ on $\mathbf{R}$ satisfies

$$v(x) \leq Cx^\alpha, \quad x \leq x_0$$

for some constants C, α and x_0, then we say that v is of order x^α as $x \to 0$ and write $v(x) = \mathcal{O}(x^\alpha)$. Left and right-hand limits are denoted by $v(x-)$ and $v(x+)$, respectively.

1.2.1 Weak solutions and entropy conditions

Here and in the rest of this paper we will only consider scalar conservation laws (first in one space dimension (1D) and later in two and higher dimensions), i.e. problems on the form

$$\begin{aligned} u_t + f(u)_x &= 0 \\ u(x, 0) &= u^0(x), \end{aligned} \tag{1.6}$$

[2] If $x = (x_1, \cdots, x_n) \in \mathbf{R}^n$, then $TV_x(v)$ is short for $TV_{(x_1,\ldots,x_n)}(v)$.

where $f : \mathbf{R} \to \mathbf{R}$ and $u : \mathbf{R} \times \mathbf{R}^+ \to \mathbf{R}$. Problems on the form (1.6) are called (general) initial value problems or simply Cauchy problems. We begin with perhaps the simplest example which lead to discontinuous solutions. Consider Burgers' equation (without viscosity),

$$u_t + (\frac{u^2}{2})_x = 0 \tag{1.7}$$

which can be written in the form:

$$u_t + uu_x = 0.$$

The solutions of Burgers' equation may be constructed using the method of characteristics. The characteristics satisfy

$$x'(t) = u(x(t), t)$$

and along each characteristic u is constant, since

$$\frac{d}{dt}u(x(t), t) = u_t + uu_x = 0.$$

Moreover, since u is constant on each characteristic, the slope $x'(t)$ is constant and so the characteristics are straight lines determined by the initial data. If the initial data is smooth, then this can used to determine the solution $u(x, t)$ for small enough t that characteristics do not cross: For each (x, t) we can solve the equation

$$x = \zeta + u(\zeta, 0)t \tag{1.8}$$

for ζ and then

$$u(x, t) = u(\zeta, 0).$$

For larger t the equation (1.7) may not have a unique solution. This happens when characteristics cross, as will eventually happen if $u_x(x, 0) < 0$ at any point. At the time where the characteristics cross, the function $u(x, t)$ has an infinite slope and a shock forms. Beyond this point there is no classical solution of the PDE, and the solution we wish to determine becomes discontinuous. A differentiable function that satisfies the differential equation and the initial data is called a classical solution. A discontinuous function can not be such a solution, so the definition of a (classical) solution of the PDE is too limited to obtain global existence. A natural way to define a generalized solution of (1.6) that does not require differentiability is to consider an integral form of the conservation law. Let $\phi \in C_0^1(\mathbf{R} \times \mathbf{R}^+)$. Here $C_0^1(\mathbf{R} \times \mathbf{R}^+)$ is the space of functions on $\mathbf{R} \times \mathbf{R}^+$ that are continuously differentiable with compact support. If we multiply $u_t + f(u)_x = 0$ by $\phi(x, t)$ and then integrate over space and time, we obtain

$$\int_0^\infty \int_{-\infty}^\infty [\phi u_t + \phi f(u)_x] dx\, dt = 0.$$

Now integrating by parts yields

$$\int_0^\infty \int_{-\infty}^{\infty} [\phi_t u + \phi_x f(u)] dx\, dt + \int_{\infty}^{\infty} \phi(x,0) u(x,0)\, dx = 0. \tag{1.9}$$

We are now in the position of defining a weak (or generalized) solution of (1.6).

Definition 1.1 (Weak solution) *The function* $u(x,t)$ *is called a weak solution of the conservation law (1.6) if for all* $\phi \in C_0^1(\mathbf{R} \times \mathbf{R}^+)$ *the equality (1.9) holds.*

Notice that the definition above does not require that u is differentiable or continuous. Furthermore, one can show if u is smooth then Definition 1.1 and the definition of a classical solution are equivalent. We have now introduced a new "solution concept" which allows discontinuous solutions. But can all kinds of discontinuities be accepted ? The answer to that is no. One can show from the weak formulation of the problem that any discontinuity must satisfy what is called the Rankine-Hugoniot (R-H) condition. This condition reads as follows: let $x(t)$ be a curve of discontinuity and let

$$\begin{aligned} u_L &= \lim_{\delta \to 0+} u(x - \delta, t) \\ u_R &= \lim_{\delta \to 0+} u(x + \delta, t), \end{aligned}$$

i.e., u_L and u_R are values of u, at time t, immediately to the left and the right of the discontinuity. Define the shock speed by $s = \frac{dx(t)}{dt}$, then s must satisfy the (R-H) condition

$$s(u_R - u_L) = f(u_R) - f(u_L). \tag{1.10}$$

For a proof of this see Smoller [20]. Rather than checking Definition 1.1 directly, we can instead use the R-H condition to decide whether a function u is a weak solution, or not. If u satisfies the following conditions

1. $u(x,0) = u^0(x)$
2. $u_t + f(u)_x = 0$ when u is smooth
3. $s(u_R - u_L) = f(u_R) - f(u_L)$ when u is discontinuous,

then $u(x,t)$ is a weak solution of problem (1.6). Let us now use this to find the weak solution of Burgers' equation with initial data

$$u(x,0) = \begin{cases} 0 & x \le 0 \\ 1 & x > 0. \end{cases}$$

We calculate $s = \frac{u_L + u_R}{2} = \frac{1}{2}$, and the function

$$u(x,t) = \begin{cases} 0 & x \leq t/2 \\ 1 & x > t/2. \end{cases} \tag{1.11}$$

is therefore a weak solution to the problem. Simple differentiating show that also the function

$$v(x,t) = \begin{cases} 0 & x \leq 0 \\ x/t & 0 < x \leq t \\ 1 & x > t \end{cases} \tag{1.12}$$

is a weak solution to the problem. For this problem we have to weak solutions; u and v. The point here is that the conditions mentioned above (1–3) are not strong enough to secure uniqueness.

As demonstrated above, there are situations in which the weak solution is not unique and an additional condition is required to pick out the physically relevant solution. Hyperbolic conservation laws often arise in models of physical processes which ignore the effects of various dissipative (and dispersive) mechanisms. In models at the next level of exactness these mechanisms make their appearance felt by the presence of higher derivatives in the equations multiplied by small coefficients called viscosity coefficients. For example in the 1D scalar case, the parabolic equation

$$u_t + f(u)_x = \varepsilon u_{xx}, \quad \varepsilon > 0 \tag{1.13}$$

would be a more exact description of the physics than the inviscid equation (1.6). The consistency of the models would then demand that solutions of the two equations be "close" in some sense, and in the limit, as the viscosity coefficient ε tend to zero, the solution of (1.13) should converge to the solution of (1.6). These ideas can be carried one step further; namely; we can try to obtain the existence of a solution of the hyperbolic equation (1.6) as the limit of the solution of the parabolic equation (1.13) as the viscosity coefficient tend to zero. This technique is usually refered to as the viscosity method, and the limit solution as the vanishing viscosity solution. It can be shown that the vanishing viscosity solution is in fact a weak solution of (1.6). The parabolic equation presumably has smooth solutions so that these should be easier to construct, but the more difficult problem is to obtain estimates independent of ε, so as to allow passage to the limit. The viscosity method has been shown to work only in the case of scalar equations; no such theorem has been proved for general systems. The requirement that the solution of (1.6) should be the limit solution of (1.13) is called the entropy condition. A weak solution satisfying the entropy condition is called the entropy (or entropy weak) solution. It can be shown that the entropy solution is unique. The entropy condition can be stated in

various forms. One version, due to Oleinik [16], is the following:

Entropy condition (Oleinik:) $u(x,t)$ is the entropy solution if all discontinuities (with left value u_L and right value u_R) have the property that

$$\frac{f(u)-f(u_L)}{u-u_L} \geq s \geq \frac{f(u)-f(u_R)}{u-u_R}$$

for all u between u_L and u_R.

For convex f (e.g. Burgers' equation) this requirement reduces to

$$f'(u_L) > s > f'(u_R). \tag{1.14}$$

We recall that we found two weak solutions of Burgers' equation with initial data

$$u(x,0) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0. \end{cases}$$

Using condition (1.14) we see that the entropy solution to this problem is given by (1.12). The discontinuous solution (1.11) does not satisfy the entropy condition.

We have now seen that a function u is the unique physical solution of (1.6) if it satisfies two requirements; i) the definition of weak a solution and ii) the entropy condition. In some situations it is preferable to have just one condition instead of the two conditions listed above. Kružkov [8] has come up with a definition of the entropy solution which, in some sense, combine or unite the two conditions into one:

Definition 1.2 (Kružkov) *u is the entropy weak solution of (3.1) if for all constants a, all $\phi \in C_0^1(\mathbf{R}\times\mathbf{R}^+)$, $\phi \geq 0$, the inequality*

$$\int_{\mathbf{R}}\int_{t>0} [\phi_t|u-a| + \operatorname{sign}(u-a)(f(u)-f(a))\phi_x]\,dtdx + \int_{\mathbf{R}} |u^0-a|\phi(x,0)\,dxdy \geq 0$$

holds.

This definition (or more precisely a two-dimensional version of it) will be used extensively in Chapter 3.

1.2.2 Riemann problems

Before we say a few words about the general initial value problem, let us look at a very special class of initial value problems; namely; the Riemann problems. The Riemann problem for a 1D scalar hyperbolic conservation law is the initial value problem with initial data of the form

$$u(x,0) = \begin{cases} u_L & x \leq 0 \\ u_R & x > 0, \end{cases}$$

where $u_L, u_R \in \mathbf{R}$. The name is due to a problem Riemann studied in gas dynamics. The physical setup for the problem (also refered to as the "shock tube problem") is a tube filled with gas, initially divided by a membrane into two sections. The gas has higher density and pressure in one half of the tube than in the other, with zero velocity everywhere. The question that Riemann wished to answer was what happens if, at time $t = 0$, the membrane is suddenly removed or broken, and the gas allowed to flow. This situation is modeled by the one-dimensional Euler equations, and the answer lies in the solution of the Riemann problem for this system of conservation laws, cf. [20] for details.

For 1D scalar conservation laws the solution of the Riemann problem is relatively easy to construct. The reason for this is that the solution is a function of only one variable; $\xi = x/t$. The solution $u = u(\xi)$ satisfies $u(-\infty) = u_L$ and $u(\infty) = u_R$. Between u_L and u_R, u is a composition of two types of waves; i) shock waves and ii) rarefaction waves.

Shock waves: A shock wave is an elementary wave consisting of a simple discontinuity which satisfies both the R-H condition and the entropy condition, i.e., a wave of the form

$$u(x,t) = \begin{cases} u_L & x \le st \\ u_R & x > st, \end{cases} \tag{1.15}$$

where

$$s(u_R - u_L) = f(u_R) - f(u_L).$$

The shock wave (1.15) is the entropy solution of the Riemann problem depending on the relationship between u_L and u_R. Two seperate cases have to be considered:

- $u_L < u_R$: The shock wave (1.15) is the solution if

  $$f(u) \ge f(u_L) + s(u - u_L)$$

  for all $u \in (u_L, u_R)$.

- $u_L > u_R$: The shock wave (1.15) is the solution if

  $$f(u) \le f(u_L) + s(u - u_L)$$

  for all $u \in (u_R, u_L)$.

Rarefaction waves: A rarefaction wave is a smooth part of the solution, and is of the form

$$u(x,t) = \begin{cases} u_L & \xi \le f'(u_L) \\ v(\xi) & f'(u_L) < \xi \le f'(u_R) \\ u_R & \xi > f'(u_R), \end{cases} \tag{1.16}$$

where $v(\xi)$ is given by the equation

$$f'(v(\xi)) = \xi. \tag{1.17}$$

This easily follows by assuming that $v = v(\xi)$ is differentiable and then substituting $v = v(\xi)$ into $u_t + f(u)_x = 0$. The speed of a rarefaction wave is given by $f'(v(\xi))$, and since the speed must be an increasing function of ξ, $f'(v(\xi))$ has to be increasing in a rarefaction wave. By differentiating (1.17) with respect to ξ, we see that

$$f''(v(\xi))v'(\xi) = 1.$$

In a rarefaction wave f'' and v' therefore have the same sign. Similarly to the shock wave solution, we have two seperate cases to consider:

- $u_L < u_R$: The rarefaction wave (1.16) is then the solution of the Riemann problem if $f''(v) > 0$ for all $v \in (u_L, u_R)$.

- $u_L > u_R$: The rarefaction wave $u(x,t)$ (1.16) is then the solution of the Riemann problem if $f''(v) < 0$ for all $v \in (u_R, u_L)$.

The fact that the solution of a Riemann problem can be constructed relatively easy is exploited by many numerical methods. In Chapter 2 we will consider a particular method that is based on solving Riemann problems.

1.2.3 The general initial value problem

Before we look at some of the numerical methods available for conservation laws, let us first state some of the well-known facts about the true solution of the general initial value problem (1.6). The reason for this is that many of the properties we would like the numerical solution to have is motivated by corresponding properties of the true solution. If $u = u(x,t)$ is the unique entropy solution of (1.6), then the following results are well-known (cf. Smoller [20]):

1. Maximum principle: $\|u(\cdot,t)\|_\infty \leq \|u^0\|_\infty$ for all $t \in \mathbf{R}^+$.

2. Stability: Let $v = v(x,t)$ be the entropy solution of $v_t + f(v)_x = 0$ with initial data v^0. If $u^0, v^0 \in L^1(\mathbf{R})$, then $\|u(\cdot,t) - v(\cdot,t)\|_1 \leq \|u^0 - v^0\|_1$ for all $t \in \mathbf{R}^+$.

3. TVD property: $T.V._x(u(\cdot,t_2)) \leq T.V._x(u(\cdot,t_1))$ for $t_2 > t_1$.

1.3 Numerical methods

When we attempt to solve nonlinear conservation laws numerically we run into additional difficulties not seen in linear equations. Moreover, the nonlinearity makes everything harder to analyze. For smooth solutions to nonlinear problems, the numerical method can often be linearized and results from the theory of linear finite difference methods applied to obtain convergence results for nonlinear problems. For discontinuous solutions very reasonable looking finite difference methods can easily give disastrous results that obviously (or sometimes not so obviously) are incorrect. There are number of difficulties that can arise. A few examples are:

- The method might be "nonlinearly unstable", i.e., unstable on the nonlinear problem even though linearized versions appear to be stable. Often oscillations will trigger nonlinear instabilities.
- The method might converge to a function that is not a weak solution of our original equation (or that is the wrong weak solution, i.e., does not satisfy the entropy condition).

Luckily, there turns out to be a simple and natural requirement we can impose on our numerical method which will guarantee that we do not converge to non-solutions. This is the requirement that the method be in conservative form, which means it has the form

$$u_j^{n+1} = u_j^n - \frac{\Delta t}{\Delta x}[F(u_{j-p}^n, u_{j-p+1}^n, \ldots, u_{j+q}^n) - F(u_{j-p-1}^n, u_{j-p}^n, \ldots, u_{j+q-1}^n)], \qquad (1.18)$$

for some function F of $p+q+1$ arguments. We have here discretized the $x-t$ plane by choosing a mesh width Δx, a time step Δt, and defined the discrete mesh points (x_j, t_n) by

$$x_j = j\Delta x, \quad j \in \mathbf{Z}$$
$$t_n = n\Delta t, \quad n \in \mathbf{Z}^+.$$

F is called the numerical flux function. In the simplest case, $p = 0$ and $q = 1$ so that F is a function of only two variables and (1.18) becomes

$$u_j^{n+1} = u_j^n - \frac{\Delta t}{\Delta x}[F(u_j^n, u_{j+1}^n) - F(u_{j-1}^n, u_j^n)]. \qquad (1.19)$$

We shall also require that the method (1.19) is consistent with the original conservation law, i.e., the numerical flux F reduces to the true flux for the case of constant flow: If $u(x,t) \equiv \bar{u}$, say, then $F(\bar{u}, \bar{u}) = f(\bar{u})$ for all $\bar{u} \in \mathbf{R}$.

Let us now state some properties we would like the numerical solution to have:

- The exact solution of (1.6) satisfies a maximum principle. It is therefore natural that the numerical solution satisfies a corresponding maximum principle, i.e.,

$$\min_{i\in\mathbf{Z}} u_i^0 \le u_j^n \le \max_{i\in\mathbf{Z}} u_i^0 \quad \forall (j,n) \in \mathbf{Z}\times\mathbf{Z}^+. \tag{1.20}$$

- The exact solution is also stable with respect to perturbations of the initial data. It is therefore desirable that the numerical solution has this property, i.e.,

$$\|u^n - v^n\|_1 \le \|u^0 - v^0\|_1, \quad \forall n \in \mathbf{Z}^+. \tag{1.21}$$

  Here $\{u_j^n\}$ and $\{v_j^n\}$ are two numerical solutions generated by the same numerical method, but with initial data; u^0 and v^0. Since $\{v_j^n\}$ and $\{v_j^n\}$ are discrete functions,

$$\|u^n - v^n\|_1 = \Delta x \sum_{j\in\mathbf{Z}} |u_j^n - v_j^n|, \tag{1.22}$$

  and (1.21) takes the form

$$\sum_{j\in\mathbf{Z}} |u_j^n - v_j^n| \le \sum_{j\in\mathbf{Z}} |u_j^0 - v_j^0|. \tag{1.23}$$

- Furthermore, we saw that the true solution u of (1.6) satisfied the TVD property. We will therefore require that the numerical solution satisfies

$$TV(u^{n+1}) \le TV(u^n), \tag{1.24}$$

  where

$$TV(u^n) = \sum_{j\in\mathbf{Z}} |u_j^n - u_{j-1}^n|. \tag{1.25}$$

  It can easily be seen that this condition is necessary for the stability condition (1.21) to be fulfilled. Numerical methods that satisfies (1.24) are called total variation diminishing (TVD) methods.

- The last, and the most severe condition we will impose on a numerical method, is that the numerical solution should converge to the true solution of (1.6) as the discretization parameters $\Delta x, \Delta t \to 0$ (in this case we say for short that the numerical method converges). More precisely, if $u = u(x,t)$ is the entropy solution of (1.6), and $u_\Delta = u_\Delta(x,t)$ is the solution generated by some numerical method[3], then

$$\|u(\cdot,t) - u_\Delta(\cdot,t)\|_1 \to 0$$

[3]The numerical solution u_Δ is defined by $u_\Delta(x,t) = u_j^n$ for all $(x,t) \in [x_j, x_{j+1}) \times [t_n, t_{n+1})$.

as $\Delta x, \Delta t \to 0$. For some methods this has been shown, and for some not. Both the methods we will consider (front tracking (1D) and dimensional splitting (2D)) has been proven to converge. Actually, these two methods have all the properties listed above.

Let us state a condition that the discretization parameters Δx and Δt must obey for any conservative method to converge to the entropy solution (though this is a necessary, but not sufficient condition for a method to converge). We are of course refering to the CFL condition (named after Courant, Friedrich and Lewy) which in the case of a 1D scalar conservation law takes the form

$$\frac{\Delta t}{\Delta x} \max_u |f'(u)| \leq 1.$$

This is a classical condition. It implies that the domain of dependence of the difference equation (1.18) contains the domain of dependence of the differetial equation - the difference approximants "have all information".

We end this chapter by listing some well known ("first order") numerical methods. We will not go into details, instead we refer the reader to the books of LeVeque [13] and Langtangen and Tveito [11]. Those interested in "more sophisticated" methods should confer the same references.

Lax-Friedrichs' scheme: This scheme is a classical one and has been used both for practical calculations and mathematical analysis. Olenik [16] used this it to prove existence and stability of a solution of the scalar conservation law (1.6). For computional purposes this scheme is not particularly good. It is known to smear out discontinuous parts of the solution (but, of course, this fact apply more or less to any first order numerical method) and the rate of convergence is slow. But on the other hand, the scheme is very reliable since it satisfies all the requirements listed above and the scheme is very easy to implement. It takes the form

$$u_j^{n+1} = \frac{1}{2}(u_{j+1}^n - u_{j-1}^n) - \frac{\Delta t}{\Delta x}[f(u_{j+1}^n) - f(u_{j-1}^n)].$$

By defining the function

$$F(a, b) = -\frac{\Delta x}{2\Delta t}(b - a) + \frac{1}{2}(f(b) + f(a))$$

we see that Lax-Friedrich's scheme can be written on the form

$$u_j^{n+1} = u_j^n - \frac{\Delta t}{\Delta x}[F(u_j^n, u_{j+1}^n) - F(u_{j-1}^n, u_j^n)].$$

Since $F(\bar{u}, \bar{u}) = f(\bar{u})$, we have that the scheme is conservative. This method has the following properties:

- Maximum principle: Yes.
- Stability: Yes.
- TVD property : Yes.
- Convergence: Yes, it has been proven that $\|u(\cdot,t) - u_\Delta(\cdot,t)\|_1 = \mathcal{O}((\Delta x)^{1/2})$.
- Local truncation error[4]: $\mathcal{O}(\Delta x)$ for smooth solutions.

The upwind scheme: This scheme is well suited for problems where the speeds are either positive or negative. If we assume that they are positive, i.e., $f'(u) \geq 0$ for all u, then the scheme takes the conservative form:

$$u_j^{n+1} = u_j^n - \frac{\Delta t}{\Delta x}[f(u_j^n) - f(u_{j-1}^n)].$$

This scheme has the following properties:

- Maximum principle: Yes.
- Stability: Yes.
- TVD property : Yes.
- Convergence: Yes, it has been proven that $\|u(\cdot,t) - u_\Delta(\cdot,t)\|_1 = \mathcal{O}((\Delta x)^{1/2})$.
- Local truncation error: $\mathcal{O}(\Delta x)$ for smooth solutions.

Lax-Wendroff's scheme: This is another classical scheme for hyperbolic conservation laws. It is based on a second order Taylor expansion, and takes the form

$$\begin{aligned} u_j^{n+1} &= u_j^n - \frac{\Delta t}{2\Delta x}(f(u_{j+1}^n) - f(u_{j-1}^n)) \\ &+ \frac{(\Delta t)^2}{2(\Delta x)^2}((f(u_{j+1}^n) - f(u_j^n))f'(u_{j+\frac{1}{2}}^n) - (f(u_j^n) - f(u_{j-1}^n))f'(u_{j-\frac{1}{2}}^n)), \end{aligned}$$

where

$$u_{j+\frac{1}{2}}^n = \frac{1}{2}(u_{j+1}^n + u_j^n).$$

[4]The local truncation error $L_\Delta(x,t)$ is a measure of how well the difference equation models the differential equation locally. It is defined by replacing the approximate solution u_j^n in the difference equation by the true solution $u(x_j, t_n)$. Of course the true solution of the PDE is only a approximate solution of the difference equation, and how well it satisfies the difference equation gives an indication of how well the exact solution of the difference equation satisfies the differential equation. We say that a method is (formally) of order p if for all sufficiently smooth initial data $\|L_\Delta(\cdot,t)\|_\infty = \mathcal{O}((\Delta x)^p)$. This is the local order of of the method, but it turns out that for smooth solutions, the global error will be of the same order (provided the method is stable).

The scheme is conservative, but it has some well-known problems. It does not satisfy a maximum principle or the TVD property, and therefore not the stability condition (the scheme is known to introduce spurious oscillations). In some situations the scheme can converge to the wrong weak solution. Since the scheme is based on a second order Taylor expansion we would expect, tough, that the scheme will do better on smooth solutions than the above schemes. This is also the case. While the schemes listed above are formally first order, Lax-Wendroff's method is second order. This method is therefore often used together with some other method (which works "well" on discontinuous solutions) to yield a high resolution method [5]; Lax-Wendroff's method is applied to smooth parts of the solution and the other method is used near discontinuities. We summarize by listing the properties of the scheme:

- Maximum principle: No.
- Stability: No.
- TVD property : No.
- Convergence: No.
- Local truncation error: $\mathcal{O}((\Delta x)^2)$ for smooth solutions.

Godunov's scheme: This (conservative) scheme is one of many that are based on solving Riemann problems. We simply state the properties of the method and say that it is the same as the upwind scheme in the case of positive speeds (Godunov's method will be explained in greater detail in Chapter 3). The scheme has the following properties:

- Maximum principle: Yes.
- Stability: Yes.
- TVD property : Yes.
- Convergence: Yes, it has been proven that $\|u(\cdot,t) - u_\Delta(\cdot,t)\|_1 = \mathcal{O}((\Delta x)^{1/2})$.
- Local truncation error: $\mathcal{O}(\Delta x)$ for smooth solutions.

[5] For a precise definition of a high resolution method cf. LeVeque [13]. Some keywords are: i) at least second order accuracy on smooth solutions. ii) Sharp resolution of discontinuities without excessive smearing. iii) The absence of spurious oscillations in the computed solution. iv) The method should converge.

Chapter 2

Front tracking

På erkjennelsens vei kan man ikke gå i
flokk. Den har bare plass til en om gangen.

– Johannes Hohlenberg –

2.1 Introduction

In this chapter we are going to study a different sort of numerical method for one dimensional scalar conservation laws. The problem under consideration is still the general intial value problem

$$\begin{aligned} u_t + f(u)_x &= 0 \\ u(x,0) &= u^0(x), \end{aligned} \tag{2.1}$$

where f is absolutely continuous, f' bounded, and u^0 is assumed to be bounded and of locally bounded variation in $\mathbf{R}$. The standard numerical approach to conservation laws are by finite difference schemes. The derivatives are approximated by finite difference quotients, and the equation may be solved for one time-step at the time.

The front tracking approach is somewhat different. The idea is to replace the flux function with a piecewise linear and continuous approximation, the initial data with a piecewise constant approximation, and then to solve this perturbed problem exactly which is essentially to solve a series of Riemann problems. When the flux function is piecewise linear it is possible to find the exact solution of the Riemann problem. Loosely speaking one can say that the front tracking method is, instead of finding an approximate solution to the exact problem, to find the exact solution to some approximate problem. This approach is based upon an idea by Dafermos [4]. He used the idea as a new method to study the initial value problem and using this method he was able to prove some properties about the solution of the initial value problem.

In [5] Holden, Holden and Høegh-Krohn proved that the technique briefly introduced above used as a numerical method is well–defined, and it is used to give an alternative proof of existence, uniqueness and stability of the solution of the differential equation. They also provide us with explicit error estimates which show that the error in the method is smaller than in any other first order method. Front tracking as introduced in this chapter will be the main building block for the method of dimensional splitting. A reasonable understanding of this method is therefore required before reading the rest of the paper.

2.2 The numerical method

We start by giving a procedure for how to construct the exact solution to the problem

$$\begin{aligned} v_t + g(v)_x &= 0 \\ v(x,0) &= v^0(x), \end{aligned} \tag{2.2}$$

where g is some piecewise linear and continuous function, and v^0 is some piecewise constant function (taking a finite number values). If $g'(v)$ does not exist then $(v, g(v))$ is called a breakpoint. Let $\{(v_i, g_i)\}_{i=0}^N$ be the breakpoints of the piecewise linear g, i.e., g is linear on each interval $[v_i, v_{i+1}]$ for $i = 0, \ldots, N-1$, and $g(v_i) = g_i$ for $i = 0, \ldots, N$. We assume that $v < v_0 \Rightarrow g(v) = g_0$ and $v > v_N \Rightarrow g(v) = g_N$. The simplest problem of the form (2.2) is the Riemann problem, where

$$v(x,0) = \begin{cases} v_L & x \leq 0 \\ v_R & x > 0. \end{cases}$$

Let us for a moment assume that $v_L < v_R$. This means that $v_0 = v_L$ and $v_N = v_R$. Furthermore, let g_c denote the lower convex envelope of g. One may imagine constructing the lower convex envelope in the following way. Fix one end of a rubber band at the point $(v_L, g(v_L))$. While keeping the rubber band below the graph of g fix the other end at the point $(v_R, g(v_R))$. The rubber band is now the graph of the convex envelope of g, for an example see the dotted line in Figure 2.1. [1] Since g is piecewise linear so is g_c. Let $\bar{v}_0 < \bar{v}_1 < \cdots < \bar{v}_M$ be such that

$$\bar{v}_0 = v_0, \quad \bar{v}_M = v_N \quad \{\bar{v}_0, \ldots, \bar{v}_M\} \subseteq \{v_0, \ldots, v_N\},$$

[1] More precisely: the lower convex envelope of a function g is the lower boundary of the convex hull of the set $\{(x,y) : v_L \leq x \leq v_R, y \geq g(x)\}$. The convex hull is the smallest convex set containing the original set.

and such that g_c is linear in each interval $[\bar{v}_i, \bar{v}_{i+1}]$. The solution of the Riemann problem with left state $v_L = v_0$ and right state $v_R = v_N$ is then given by

$$v(x,t) = \begin{cases} v_L & x \le \bar{s}_0 t \\ \bar{v}_i & \bar{s}_i t < x \le \bar{s}_{i+1} t,\ i = 1, \dots, N-1 \\ v_R & x > \bar{s}_{N-1} t. \end{cases} \tag{2.3}$$

where

$$\bar{s}_i = \frac{\bar{g}_{i+1} - \bar{g}_i}{\bar{v}_{i+1} - \bar{v}_i} = g_c'(v_i+), \quad i = 0, \dots, N-1$$

and $\bar{g}_i = g(\bar{v}_i)$. When $v_L > v_R$ there is an similar formula involving the upper convex envelope[2]. For examples of convex envelopes and solutions in the $x-t$ plane see Figure 2.1–2.4. Observe that as a result of g being piecewise linear, the solution contains no rarefaction waves, only shocks. We see that the solution $v(x,t)$ consists of N constant states separated by $N-1$ discontinuities (also called shock fronts). Since the shock speeds are $\bar{s}_i = \frac{\bar{g}_{i+1} - \bar{g}_i}{\bar{v}_{i+1} - \bar{v}_i}$, $i = 0, \dots, N-1$, the Rankine–Hugoniot condition is satisfied. Furthermore, the entropy condition for the $[\bar{v}_i, \bar{v}_{i+1}]$ front reads:

$$g(v) \ge g(\bar{v}_i) + \bar{s}_i (v - \bar{v}_i), \quad v \in [\bar{v}_i, \bar{v}_{i+1}],$$

i.e. g should lie above the straight line from $(\bar{v}_i, g(\bar{v}_i))$ to $(\bar{v}_{i+1}, g(\bar{v}_{i+1}))$. By construction of the convex envelope this requirement is always fulfilled. This is also seen to be true in Figure 2.1, where $g(v)$ lies above the dotted line. The unique entropy solution of the Riemann problem is therefore given by (2.3)

Next, we show how to construct the solution of (2.2) with v^0 piecewise constant on a finite number of intervals. Observe that each jump in v^0 defines a Riemann problem, see Figure 2.5. The solution of these problems leads to a series of discontinuities propagating in the $x-t$ plane. "Gluing" together the solutions of the Riemann problems we have the solution of (2.2) until, at some point, two or more of these discontinuities interact, and we have what is called a shock collision, see Figure 2.6. When two or more neighboring discontinuities collide they define a new Riemann problem with left and right state the values immediately to the left and the right of the collision. This Riemann problem is solved and we have the solution of (2.2) until the next shock collision occur. This collision is of course handled in the same way, and so on. The method for constructing the global solution to (2.2) may briefly be summarized as follows:

1. Solve the Riemann problems defined by the piecewise constant initial data.

[2] If $v_L > v_R$, then construct the convex hull of the set $\{(x,y) : v_R \le x \le v_L, y \le g(x)\}$. The upper convex envelope is then the upper boundary of the convex hull.

2. Keep track of the shock collisions, and solve the Riemann problems arising at the collision points.

From the construction it is easily seen that the solution v will be piecewise constant, there are no rarefaction waves only shocks. For this construction-procedure to be well-defined it is necessary to prove that if v^0 has a finite number of jumps, then $v(x,t)$ is constant on a finite number domains for $t < T_0$, for some fixed T_0. In other words the procedure presented above must consists of a finite number of steps, i.e., the number of shock collisions must be finite in finite time. This was shown in Holden, Holden and Høegh-Krohn [5], where they in fact proved that v is constant on finite number domains for all t. See Figure 2.7 for an example of a typical solution in the $x-t$ plane. Note that the construction above will yield the exact solution of (2.2), and not just some approximation.

Let us now return to the original problem (2.1). The numerical method for solving this problem is to replace the flux function f with some piecewise linear and continuous approximation and the initial data with some piecewise constant approximation, and then to solve the perturbed problem exactly using the procedure described above. Solving the perturbed problem, we get an approximation to the solution of (2.1). The approximate solution will as mentioned before, be piecewise constant and rarefaction waves will be approximated by a series of weak shocks, i.e. $|v_L - v_R|$ "small". This numerical method has several features that stand out from other methods:

- Since the perturbed system is solved exactly, the numerical method is completely stable (no CFL condition) and introduces no artificial diffusion (shocks will not be smeared out).

- When it comes to accuracy, the front tracking scheme compares favourably to all known methods. If we choose the piecewise constant approximation v^0 and the piecewise linear approximation g properly, then one can show that

$$\|u(\cdot,t) - v(\cdot,t)\|_1 \leq C(\Delta x + \delta),$$

  for some finite constant C and where Δx is the length of the intervals where v^0 is piecewise constant, δ the length of the intervals where g is piecewise linear, cf. [5] for details. The front tracking method therefore has an error which decreases linearly (in Δx and δ) to zero. Most other (finite difference) methods has an error which decreases to zero like $(\Delta x)^{1/2}$.

- An essential part of the front tracking method is the "Riemann solver". To solve a Riemann problem we need to determine the convex envelope of a piecewise linear function which numerically is "easy". As a result the method becomes very fast.

Numerical experiments suggest that front tracking is superior to other methods when it comes to achieving a certain degree of accuracy in shortest possible CPU-time.

- It is possible with the front tracking scheme to approximate the solution $u(\cdot, t)$, as the time t tends to infinity, in finite time (since the number of shock collisions is finite even in infinite time). This makes the method well-suited to study asymptotic behaviour for large time of the solution.

Finally, we end this chapter by giving a precise (algorithmic) presentation of the front tracking scheme. In order do so, we need some data structure. We first define a front

front: $v_L, v_R, s, (x_0, t_0), (x_1, t_1), Nx, Px, Nt$.

Thus a front is an object containing the left state v_L and the right state v_R of a discontinuity. The discontinuity starts in (x_0, t_0) and its speed is given by s. The time and place of interaction of the discontinuity with its right-hand side neighbour is given by (x_1, t_1). The fronts are stored in two lists, one X-list and one T-list. In the X-list the fronts are placed from left to right (with respect to x_0) using the Nx (next front in the X-list) and Px (previous front in the X-list) pointers. To slightly simplify the implementation, let the last front in the X-list have Nx pointing to the first front, and let the first front's Px pointer point to the last front. This construction makes the X-list a two-way double linked circular list. In the T-list fronts are organized with respect to collision times. The front with smallest collision time t_1, is always the first element of the T-list. This front has the Nt (next front in the T-list) pointer pointing to the front with the second smallest collision time, and so on. Suppose that P is a pointer to one of objects in the X-list. Then $P.v_L$ denotes the value of v_L in the object at which P is pointing; similarly for the rest of the real variables in the front object. Furthermore, P.Nx points to the next front in the X-list; similarly for P.Px and P.Nt. All the front objects are stored only once, hence the X-list and T-list use the same objects.

Algorithm 2.1 (Front tracking) *The function rightmost(*L*) used in the algorithm below, returns a pointer to the front on the right-hand side of* L*. Usually it will return the* L.Nx*, but if three or more fronts collide, the function rightmost(*L*) will return the rightmost front involved the collision. We approximate the solution to the Cauchy problem from* $t = 0$ *to* $t = T_0$.

1. *If the initial data is not piecewise constant and/or if the flux function is not piecewise linear, then construct piecewise constant initial data and/or piecewise linear flux function;*

2. *Solve the Riemann problems defined by the piecewise constant initial data;*

3. *Organize the fronts in the X-list. If a discontinuity will collide with its right-hand side neighbour, then the collision time and place* (x_1, t_1) *must be computed. If not let* $t_1 = \infty$*;*

4. *Organize the fronts in the T-list according to increasing values of collision time, using the* Nt *pointer;*

5. *While T-list not empty and* $T.t_1 < T_0$

   5.1. $L = T$*;* $R =$*Rightmost(*L*);*

   5.2. *Compute the solution to the Riemann problem with left state* $L.v_L$ *and right state* $R.v_R$*;*

   5.3. *Remove the fronts* L *and* R *from the X and T-list, and include the fronts generated by the solution of the Riemann problem in the X-list. Let* $\bar{L}$ *and* $\bar{R}$ *be pointers to the first and the last front of the solution of the Riemann problem.*

   5.4. *If* $\bar{L}.Px.s > \bar{L}.s$

      5.4.1. *Compute* $\bar{L}.Px.x_1$, $\bar{L}.Px.t_1$, *and update the T-list;*

   5.5. *Else*

      5.5.1. $L.Px.t_1 = \infty$*;*

   5.6. *If* $\bar{R}.s > \bar{R}.Nx.s$

      5.6.1. *Compute* $\bar{R}.x_1$, $\bar{R}.t_1$, *and update the T-list;*

   5.7. *Else*

      5.7.1. $R.t_1 = \infty$*;*

The data structure and the algorithm presented above were first introduced by Risebro and Tveito in [18]. The front tracking scheme will be used as a part of the dimensional splitting method studied in the next chapter. Numerical experiments and aspects concerning the implementation will be reported in Chapter 4.

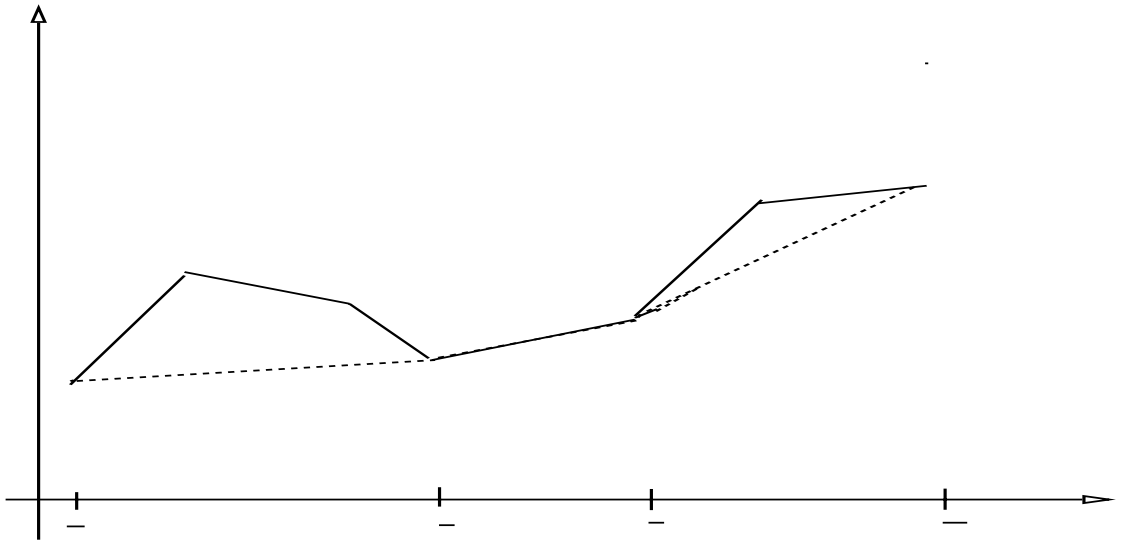

Figure 2.1: The piecewise linear g and the lower convex envelope g_c, $v_L = v_0 < v_R = v_3$.

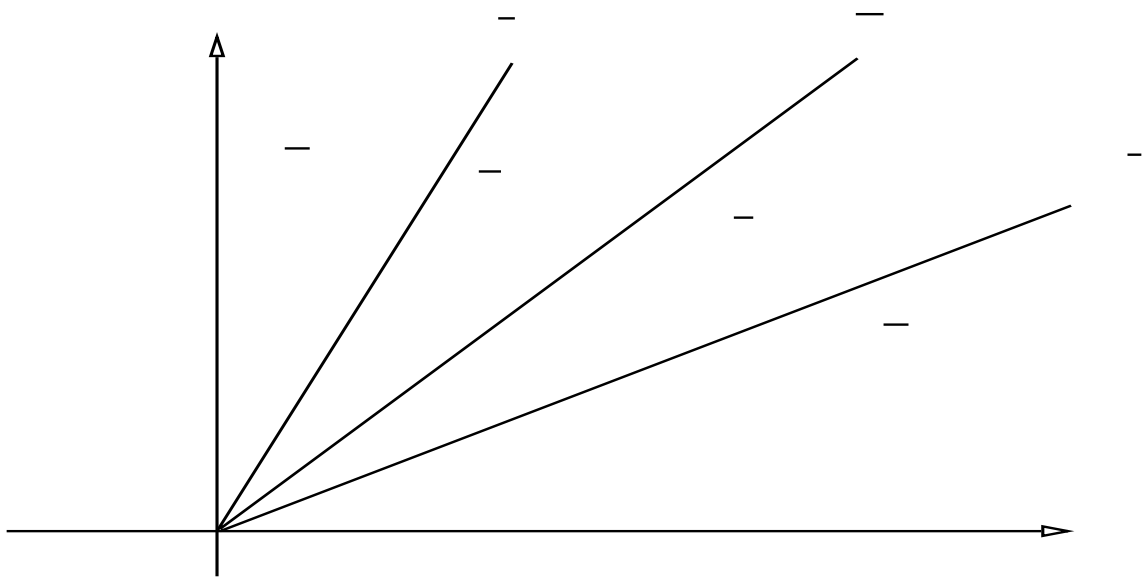

Figure 2.2: The corresponding solution to the lower convex envelope above, $v_L = v_0 < v_R = v_3$.

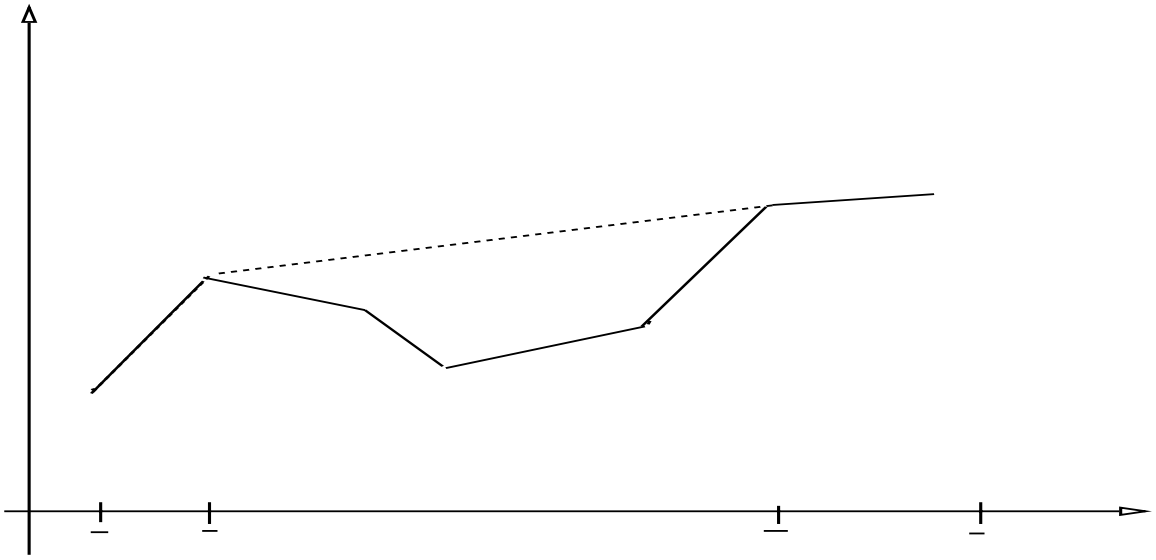

Figure 2.3: The piecewise linear g and the upper convex envelope g_c, $v_L = v_3 > v_R = v_0$.

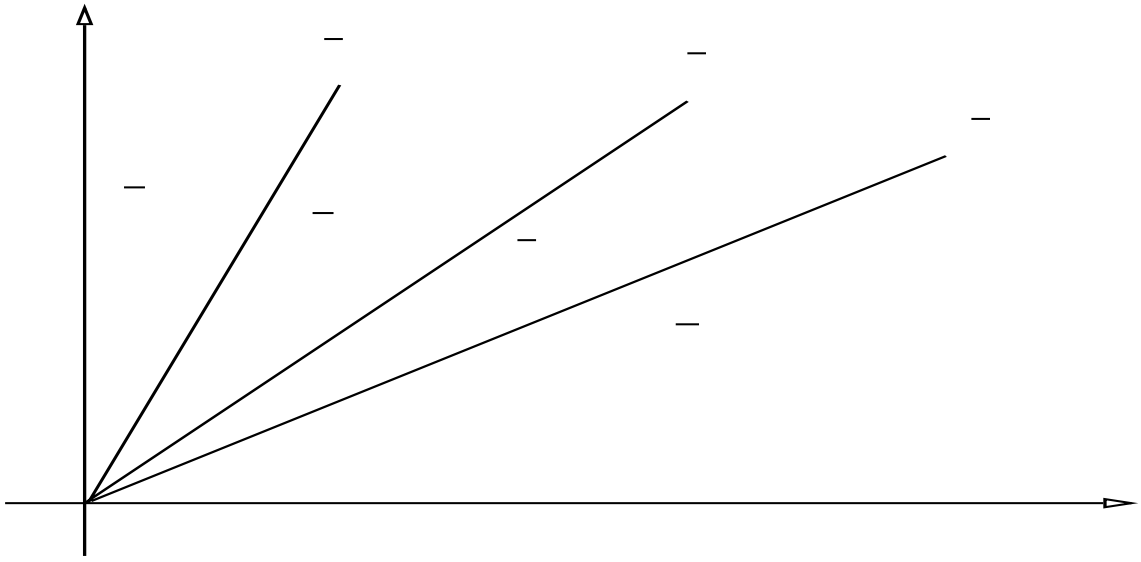

Figure 2.4: The corresponding solution to the upper convex envelope above, $v_L = v_3 > v_R = v_0$.

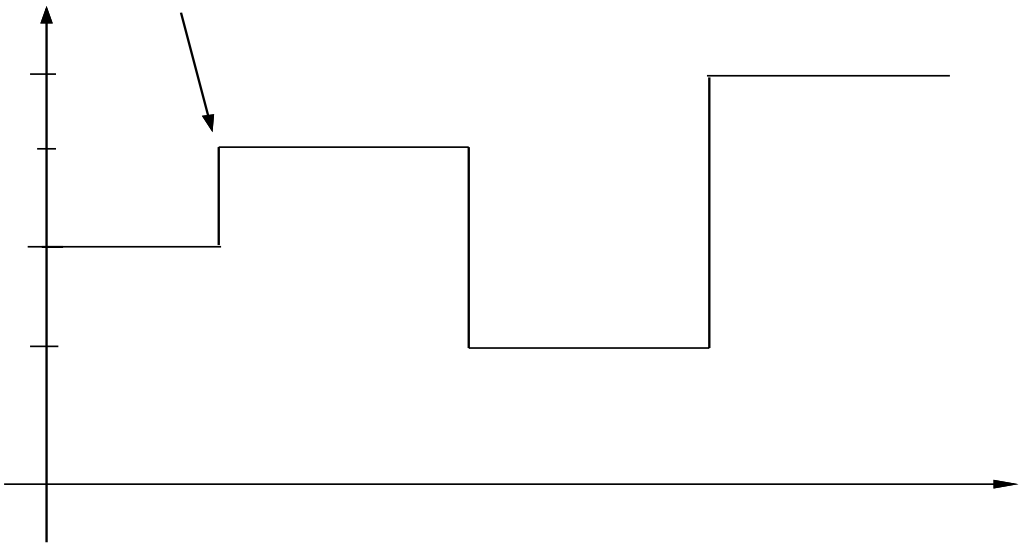

Figure 2.5: Initial data piecewise constant. Each jump defines a Riemann problem. The first jump defines a Riemann problem with left state $v_L = v_1$ and right state $v_R = v_2$.

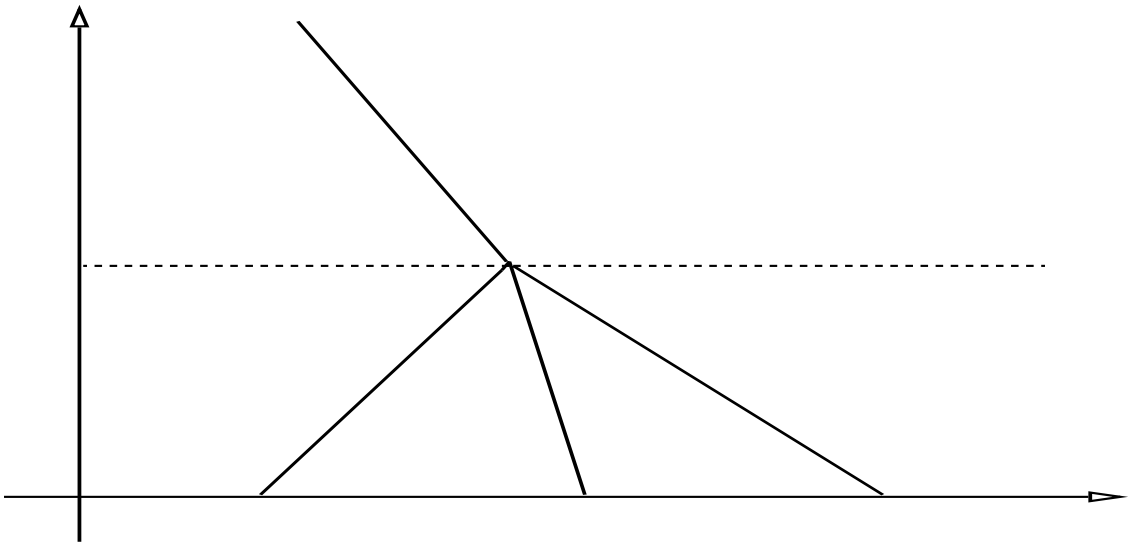

Figure 2.6: A Shock collision involving three fronts. At the shock collision a new Riemann problem is solved with left state v_1 and right state v_4.

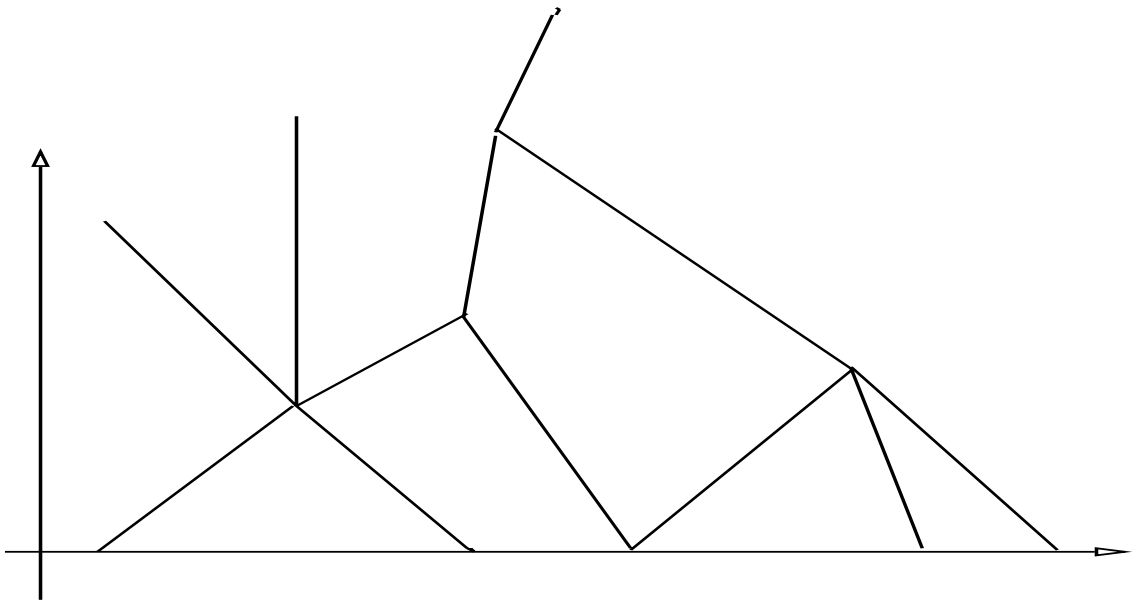

Figure 2.7: A typical solution; $\nu(x,t)$ is constant on a finite number of domains.

Chapter 3

The method of dimensional splitting

Ofte er det jo slik at når man kikker det nye
etter i sømmene, er det bare sømmene som er nye.

– Kaj Munk –

3.1 Introduction

In this chapter we will study one particular numerical method for the n-dimensional Cauchy problem

$$\begin{aligned} u_t + \sum_{i=1}^{n} f_i(u)_{x_i} = 0, \quad & x \in \mathbf{R}^n \\ u(x,0) = u^0(x,) \quad & x \in \mathbf{R}^n. \end{aligned} \tag{3.1}$$

With a solution to (3.1) we mean a bounded, measurable function which satisfies Kružkov definition of the entropy weak solution.

Definition 3.1 (Kružkov) u *is the entropy weak solution of (3.1) if for all constants* a*, all* $\phi \in C_0^1(\mathbf{R}^2 \times \mathbf{R}^+)$*,* $\phi \geq 0$*, the inequality*

$$\begin{aligned} \int_{\mathbf{R}^n} \int_{t>0} [\phi_t |u-a| + \operatorname{sign}(u-a) \sum_{i=1}^{n} (f_i(u) - f_i(a))\phi_{x_i}] \, dt d^n x + \\ \int_{\mathbf{R}^n} |u^0 - a| \phi(x_1, \ldots, x_n, 0) \, dx dy \geq 0 \end{aligned} \tag{3.2}$$

holds.

The advantage with this formulation is the embedding of the entropy condition. Kružkov's definition of the entropy weak solution is a mechanism to identify the unique physical solution of (3.1).

In this and the forthcoming chapters we will for simplicity of notation consider the two-dimensional problem which reads

$$\begin{aligned} u_t + f(u)_x + g(u)_y = 0, \quad & x, y \in \mathbf{R} \\ u(x, y, 0) = u^0(x, y), \quad & x, y \in \mathbf{R}. \end{aligned} \tag{3.3}$$

For the two-dimensional problem, inequality (3.2) takes the form

$$\int_{\mathbf{R}^2} \int_0^T [\phi_t |u - a| + \operatorname{sign}(u - a)\{(f(u) - f(a))\phi_x + (g(u) - g(a))\phi_y\}] \, dt \, dx \, dy + \int_{\mathbf{R}^2} |u^0 - a| \phi(x, y, 0) \, dx \, dy \geq 0. \tag{3.4}$$

We remark that all the results derived in this chapter extend easily to the n-dimensional case. Most practical problems are, indeed, in two or three space dimensions. So far we have only considered the one-dimensional problem. To some extent the 1D methods can be applied to problems in more than one space dimension. There are two approaches:

1. semi-discrete methods
2. dimensional splitting.

Our main concern in this paper will be on the splitting methods. For a brief introduction to semi-discrete methods we refer the reader to LeVeque's book [13]. But first, what about an extension of the front tracking method from the previous chapter to two or three dimensions ? Recall that the idea was to replace the flux function with a piecewise linear function and the initial data with a piecewise constant function, and then to solve this new problem exactly. Finding the exact solution to this new "perturbed" problem is essentially to solve a series of Riemann problems and to "track" the shock collisions (the points where two or more discontinuities interact). There is one difficulty that immediately comes to mind. In one space dimension it is always possible to sort points (fronts) from left to right. In two and three dimensions this becomes much harder. Algorithms for finding and sorting collision points increase dramatically in terms of complexity.

The Riemann problem in two dimensions is the conservation law together with initial data that is constant on a finite number of sectors which meet in origo. In principle we could solve the Riemann problem approximatively by replacing the flux functions with piecewise linear functions. We remember that the one-dimensional Riemann problem with a piecewise linear flux function had a solution with a particular simple structure. This not at all the case in two and three dimensions. It turns out that the solution is of a very complicated nature, and finding an explicit expression representing the solution is in most cases not possible. Nevertheless, some attempts have been made on extending

the front tracking technique to two and higher space dimensions, cf. for instance [17], but these attempts only allow a very limited movement of the fronts, essentially in one direction. The conclusion is therefore that the development of a front tracking method for two or higher space dimensions is highly non-trivial and one may be tempted to begin the search for other alternatives.

Let us instead look at the idea behind the method of dimensional splitting (or fractional steps as it is sometimes called) which was first introduced by Godunov (around 1950) in connection with gas dynamics and has since been extended by various authors. Loosely speaking, we may say that the idea is simply to alternate between solving one-dimensional problems in x and y directions. Before we become formalistic let us consider an example which might serve as an illustration. The advection equation in two dimensions reads

$$\begin{aligned} u_t + a u_x + b u_y &= 0 \\ u(x,y,0) &= u^0(x,y), \end{aligned} \tag{3.5}$$

where $a, b \in \mathbf{R}$. The exact solution, at some fixed time T, is $u(x,y,T) = u^0(x - aT, y - bT)$. We now wish to solve this equation by instead solving a pair of 1D problems. We first solve

$$\begin{aligned} v_t + a v_x &= 0 \\ v(x,y,0) &= u^0(x,y). \end{aligned} \tag{3.6}$$

The solution $v = v(x,y,T)$ is then used as initial data to solve the 1D problem

$$\begin{aligned} w_t + b w_y &= 0 \\ w(x,y,0) &= v(x,y,T). \end{aligned} \tag{3.7}$$

How well does $w(x,y,T)$ approximate $u(x,y,T)$? It is easily seen that $v(x,y,T) = u^0(x - aT, y)$ is the exact solution of (3.6), and solving (3.7) gives

$$\begin{aligned} w(x,y,T) &= w(x, y - bT, 0) \\ &= v(x, y - bT, T) \\ &= u^0(x - aT, y - bT) \\ &= u(x,y,T). \end{aligned}$$

We see that w is in fact the exact solution in the linear case. Let us now develop a more general theory. From now on, $S(t)$ will denote the solution operator which takes an initial function u^0 on $\mathbf{R}^2$ to the entropy weak solution of (3.3) at time t, i.e., $u(x,y,t) = S(t)u^0(x,y)$. Note that $S(t)$ maps $L^1(\mathbf{R}^2)$ into $BV(\mathbf{R}^2)$. Furthermore, let $S^{f,x}(t)$ and $S^{g,y}(t)$ denote the solution operators which take initial functions v^0 and w^0

on $\mathbf{R}$ to the entropy weak solutions of

$$v_t + f(v)_x = 0$$
$$v(x,0) = v^0(x)$$

and

$$w_t + g(w)_y = 0$$
$$w(y,0) = w^0(y)$$

at time t, respectively, i.e., $v(x,t) = S^{f,x}(t)v^0(x)$ and $w(y,t) = S^{g,y}(t)w^0(y)$. The idea behind dimensional splitting is then to alternately apply the 1D solution operators $S^{f,x}$ and $S^{g,y}$ for small time steps Δt (as in the linear example, only now we alternate more than two times and with smaller time steps) to approximate $S(t)u^0$. The dimensional splitting solution $\{u^n_{\Delta t}\}_{n=1}^N$, where $n\Delta t = t_n$, $N\Delta t = T < T_0$, for some fixed T_0, is defined by

$$u^n_{\Delta t}(x,y) = [S^{g,y}(\Delta t)S^{f,x}(\Delta t)]^n u^0(x,y). \tag{3.8}$$

Observe that the solution is only defined for discrete t-values, $t = t_n$. The hope is that $S(T)u^0 \approx u^N_{\Delta t}$ in some appropriate norm. This was in fact shown by Crandall and Majda in 1980 [3] (see also LeVeque's book [13], Chapter 18), more precisely they proved the following theorem:

Theorem 3.1 (Crandall and Majda) *Assume that* $u^0 \in L^1(\mathbf{R}^2) \cap L^\infty(\mathbf{R}^2)$*, then the dimensional splitting solution always converges to the entropy weak solution. More precisely, if* $N\Delta t = T$*, then as* $\Delta t \to 0$ *and* $N \to \infty$*,*

$$\max_{T \geq 0} \|S(T)u^0(\cdot,\cdot) - u^N_{\Delta t}(\cdot,\cdot)\|_1 \to 0.$$

A detailed proof of this theorem can be found in [3].

In order to use the splitting (3.8) in a numerical method we simply replace the exact 1D solution operators $S^{f,x}$ and $S^{g,y}$ by numerical methods (approximate 1D solution operators). When solving 1D problems we may choose form a variety of methods available. Crandall and Majda [3] analyze the method of dimensional splitting for monotone schemes, the Glimm method and the Lax-Wendroff scheme. In particular, they show that Theorem 3.1 remains true if the exact solution operators are replaced by monotone schemes. In this paper we will analyze the method of dimensional splitting for the front tracking scheme. This scheme, first proposed by Holden and Risebro in [6], will have the advantage of yielding an unconditionaly stable approximation in the sense that the time

step is not limited by the space step, i.e., one does not need the CFL condition.

Let $\delta > 0$ denote the parameter measuring the polygonal approximation of the flux functions (i.e., the piecewise linear approximations will have breakpoints at $i\delta$, $i \in \mathbf{Z}$), and fix a grid $\{(i\Delta x, j\Delta y)\}_{i,j\in\mathbf{Z}}$ in the x-y plane. The method of dimensional splitting is then to use the front tracking scheme in the x-direction for a small time step Δt. The front tracking solution will not necessarily be piecewise constant on the original grid and the solution is therefore projected back onto this grid, before we apply the front tracking scheme in the y-direction for a time step Δt, using the (projected) solution computed in the x-direction as initial data. Each time after we apply the front tracking scheme, we project the solution back onto the original grid, and thereby obtaining a sequence of functions indexed by the number of iterations n, (the time step Δt) and the mesh size $(\Delta x, \Delta y)$.

Next, we explain in detail how the sequence of approximate solutions is constructed, and finally we prove that our numerical method generates a sequence of approximate solutions which converges to the entropy weak solution $S(t)u^0$. In particular, we obtain an explicit estimate showing how well this sequence of functions approximates $S(t)u^0$, i.e., the convergence has a rate.

3.2 Construction of approximate solutions

Let Δx and Δy be given small numbers and define a grid $\{(i\Delta x, j\Delta y)\}_{i,j\in\mathbf{Z}}$ in the x-y plane. Let π be the projection operator, from $BV(\mathbf{R}^2)$ to functions that are constant on each square

$$z_{i,j} = \{(x,y) : i\Delta x < x < (i+1)\Delta x, j\Delta y < y < (j+1)\Delta y\}$$

for $i, j \in \mathbf{Z}$, defined by

$$\pi u(x,y) = \frac{1}{\Delta x \Delta y} \int_{z_{i,j}} u(\tilde{x},\tilde{y})\, d\tilde{x}\, d\tilde{y}, \quad \forall (x,y) \in z_{i,j}.$$

$\pi u(x,y)$[1] is simply the grid average of u over the grid-cell $z_{i,j}$, and it is well-known that the error caused by the projection operator is of order $\max(\Delta x, \Delta y)$, i.e., $\|u - \pi u\|_1 = \mathcal{O}(\max(\Delta x, \Delta y))$. Since we will use the front tracking method in each direction we define f_δ and g_δ to be piecewise linear and continuous approximations to f and g, respectively: Let u_0, δ be some given real numbers with $\delta > 0$, and let $u_i = u_0 + i\delta$, $f_i = f(u_i)$ for $i = 1, \ldots, N$. We define f_δ by

$$u \in [u_i, u_{i+1}] \Rightarrow f_\delta(u) = \frac{f_{i+1} - f_i}{u_{i+1} - u_i}(u - u_i) + f_i, \quad i = 0, \ldots, N-1,$$

[1] $\pi u(x,y)$ is short for the composition of π and u; $\pi u(x,y) = (\pi \circ u)(x,y)$.

and

$$u \le u_0 \Rightarrow f_\delta(u) = f_0, \quad u \ge u_N \Rightarrow f_\delta(u) = f_N.$$

It is not necessary that each interval where f_δ is linear is equally spaced. This is just done for simplicity. The approximation should be good both in the T.V.- and the L^1-norm, i.e.,

$$\begin{aligned} \lim_{\delta\to 0} \quad & T.V._u(f_\delta - f) \quad = 0 \\ \lim_{\delta\to 0} \quad & \|f_\delta - f\|_1 \quad = 0. \end{aligned}$$

With the above definition of f_δ, these requirements are certainly fulfilled. g_δ is similarly defined. Let $S_\delta^{f,x}(t)$ and $S_\delta^{g,y}(t)$ denote the solution operators of

$$v_t + f_\delta(v)_x = 0 \tag{3.9}$$

$$w_t + g_\delta(w)_y = 0, \tag{3.10}$$

respectively, i.e., $v(\cdot,t) = S_\delta^{f,x}(t)v^0$ and $w(\cdot,t) = S_\delta^{g,y}(t)w^0$ are the entropy weak solutions of (3.9) and (3.10) with initial data v^0 and w^0, respectively. If v^0 is piecewise constant taking a finite number of values, then $S_\delta^{f,x}(t)v^0$ is identical with the solution generated by the front tracking scheme since f_δ is piecewise linear. Likewise with $S_\delta^{g,y}w^0$ when w^0 is piecewise constant. The dimensional splitting solution $\{u_\eta^n\}_{n=1}^N$, where $n\Delta t = t_n$, $N\Delta t = T < T_0$, for some fixed T_0 and $\eta = (\Delta x, \Delta y, \Delta t)$ is defined by

$$u_\eta^n(x,y) = [\pi S_\delta^{g,y}(\Delta t)\pi S_\delta^{f,x}(\Delta t)]^n u_\eta^0(x,y), \tag{3.11}$$

where $u_\eta^0(x,y) = \pi u^0(x,y)$. If, for each fixed x, $u(x,\cdot)$ is a piecewise constant function (in y) on the intervals $(j\Delta y, (j+1)\Delta y)$, $j \in \mathbf{Z}$, then we define $u_j(x)$ by

$$u_j(x) = u(x, (j+\frac{1}{2})\Delta y).$$

Similarly,

$$u_i(y) = u((i+\frac{1}{2})\Delta x, y),$$

for $u(\cdot,y)$ piecewise constant in (x) on the intervals $(i\Delta x, (i+1)\Delta x)$, $i \in \mathbf{Z}$ for each fixed y. Assume that we have finished the n'th iteration, then the next step in the method dimensional splitting consists in first applying the solution operator $S_\delta^{f,x}$ to u_j^n for each j, then project the solution back onto the grid, obtaining a piecewise constant function $u_\eta^{n+\frac{1}{2}}$ representing the solution at time $(n+\frac{1}{2})\Delta t$, and subsequently applying the solution operator $S_\delta^{g,y}$ to $u_i^{n+\frac{1}{2}}$ for each i. Finally, the result of this is projected onto the grid obtaining again a piecewise constant function u_η^{n+1} representing the solution at time $(n+1)\Delta t$, and the process is repeated. With notation introduced above: $u_\eta^{n+\frac{1}{2}}(x,y) = \pi S_\delta^{f,x}(\Delta t)u_\eta^n(x,y)$. In "computer code" the method looks like:

Algorithm 3.1 (Dimensional splitting) *In the algorithm below,* N_x *is chosen so large that* u_η^n *is constant outside the square* $[-N_x\Delta x, -N_x\Delta y] \times [N_x\Delta x, N_x\Delta y]$ *in the time interval* $[0, T]$.

1. $t = 0$;
2. $n = 0$;
3. $u_\eta^0(x, y) = \pi u^0(x, y)$;
4. *While* $t < T$
   - 4.1. *For* $j = -N_x, \ldots, N_x$
     - 4.1.1. $u_j^{n+\frac{1}{2}}(x) = S_\delta^{f,x}(\Delta t)u_j^n(x)$;
   - 4.2. $u_\eta^{n+\frac{1}{2}}(x, y) = \pi u_\eta^{n+\frac{1}{2}}(x, y)$;
   - 4.3. *For* $i = -N_x, \ldots, N_x$
     - 4.3.1. $u_i^{n+1}(y) = S_\delta^{g,y}(\Delta t)u_i^{n+\frac{1}{2}}(y)$;
   - 4.4. $u_\eta^{n+1}(x, y) = \pi u_\eta^{n+1}(x, y)$;
   - 4.5. $t = t + \Delta t$;
   - 4.6. $n = n + 1$;

The algorithm above will be explained in more detail in Chapter 4. In that chapter we implement the method on a computer and run a series of numerical tests. Our next main objective is to prove that the algorithm above may be used to produce a sequence of approximate solutions which converges to $S(t)u^0$ as the discretization parameters go to zero. This will be carried out in the next section, where also explicit error estimates will be derived.

3.3 Convergence (with error estimates)

The usual strategy used to prove convergence of a subsequence of approximate solutions is via a standard compactness argument. Assume that the initial data is bounded both in the L^∞–norm and in total variation. Let $\{u_\Delta(x, y, t)\}$ be a family of approximate solutions (constructed by some method), and let us assume that we can prove the following estimates

1. $||u_\Delta(\cdot, \cdot, t)||_\infty \leq C_1$, $\forall t \geq 0$
2. $T.V._{(x,y)}(u_\Delta(\cdot, \cdot, t)) \leq C_2$, $\forall t \geq 0$

3. $\|u_\Delta(\cdot,\cdot,t_2) - u_\Delta(\cdot,\cdot,t_1)\|_1 \le C_3|t_2 - t_1|$, $\forall t_1, t_2 \ge 0$,

where C_1, C_2 are constants independent of Δ and t, and C_3 is a constant independent of Δ, t_1, t_2. Then there is, using Helly's Theorem as in e.g. [2], a subsequence $\{u_{\Delta_j}\} \subset \{u_\Delta\}$ which converges to a function u in $L^1_{loc}(\mathbf{R}^2 \times [0,T])^2$. Actually, this approach may be used to study two types of problems: 1) the problem of showing that a numerical method produces a convergent subsequence 2) the problem of showing existence of a solution to a particular problem. To show that the limit–function u is the desired solution, we must prove that u_{Δ_j}, in the limit as $j \to \infty (\Leftrightarrow \Delta_j \to 0)$, satisfies the requirements (for instance Kružkov's definition of the entropy weak solution) to be such a solution. It may be added that this is often the hard part. Oleinik [16] used this technique to prove existence of a weak solution of a scalar conservation law in one dimension. The family of approximate solutions was constructed using Lax-Friedrich's scheme. Also Glimm [7] used this strategy in his famous existence proof for a system of strictly hyperbolic conservation laws in one dimension with small data. He used the Glimm scheme.

Convergence of our method was first established by Holden and Risebro [6] using the method described above. There they showed that the limit function u satisfied Kružkov's definition of the entropy weak solution (3.2). We will give an alternative proof of convergence which is somewhat different from the usual approach considered above. Our approach will yield an explicit convergence rate, and is based on the approximation theory developed by Kuznetsov ([9],[10])[3]. Let $\{u_\Delta\}$ be a sequence of functions, indexed over Δ, which approximates the function u. The sequence $\{u_\Delta\}$ is said to have rate of convergence $r(\Delta)$ (r is some function on $\mathbf{R}$) to u if

$$\|u_\Delta - u\| \le r(\Delta), \quad \Delta \le \Delta_0 \tag{3.12}$$

and if (3.12) can not be improved upon. Here $\|\cdot\|$ is some suitable norm. For us this means that we would like to determine a constant $\gamma > 0$ such that

$$\|S(T)u^0(\cdot,\cdot) - u^N_\eta(\cdot,\cdot)\|_1 \le C(\Delta t)^\gamma \tag{3.13}$$

for some constant C independent of Δt. Letting $\Delta t \to 0$ we have that u^N_η converges to $S(T)u^0$. With a minor abuse of the notation we will speak of a rate of convergence even though we do not know wether (3.12) may be improved or not. All of our errror estimates

[2] Helly's theorem: Let $\{u_\alpha\}$ be an infinite family of functions satisfying the estimates above (1-3). Then a subsequence $\{u_{\alpha_j}\}$ converges in L^1_{loc} to a function u.
This theorem can be applied to the family $\{u_\Delta : 0 < \Delta \le 1\}$ to obtain compactness, justifying the term compactness argument.

[3] The convergence result of Holden and Risebro yields more than it states, for the existence of the solution $S(t)u^0$ of (3.3) is established while proving convergence. Our technique presuppose the existence of $S(t)u^0$. But, of course, it has been known for a long time that $S(t)u^0$ exists, cf. for instance Kružkov [8] (1970).

will be of the form (3.12), but we have no information concerning the optimality of these estimates.

Besides proving a rate of convergence for our numerical method, we will also obtain a convergence rate for the method of dimensional splitting with exact 1D solution operators (Theorem 3.1 will be improved). This is, of course, only of theoretical interest. A consequence though, is that this result provides us with a upper bound on the convergence rate for any numerical method based on the splitting (3.8).

Let us start by presenting an argument which implies that our numerical method can not in general be first order accurate. Much of our work was motivated by a desire to prove a first order rate of convergence. In one dimension there are no known methods, besides front tracking, which are first order accurate. The hope was that this feature could be adopted to the method of dimensional splitting when the 1D solution operators are replaced by the front tracking scheme. Before we look at the argument let us briefly describe Godunov's method, for one-dimensional conservation laws, which is the key to the argument.

Let u_j^n be a approximation of $u(x,t)$ for $x \in I_j = [x_{j-1}, x_j)$ at time $t_n = n\Delta t$. Let us assume that we know $u_j^n \ \forall j \in \mathbf{Z}$, and that our aim is to calculate $\{u_j^{n+1}\}_{j\in\mathbf{Z}}$. The sequence $\{u_j^n\}_{j\in\mathbf{Z}}$ defines a piecewise constant function, and the conservation law can be solved with $\{u_j^n\}_{j\in\mathbf{Z}}$ as initial data. Let v be the exact solution of the following problem

$$\begin{aligned} v_t + f(v)_x &= 0 \quad t_n < t < t_{n+1} \\ v(x,t_n) &= \sum_{j\in\mathbf{Z}} \chi_{I_j}(x) u_j^n \end{aligned} \tag{3.14}$$

where χ_{I_j} is the characteristic function of I_j. We will solve this initial value problem for $t \in [t_n, t_{n+1})$. If Δt is sufficiently small, no waves (fronts) will interact with each other and we have a sequence of Riemann problems which we are able to solve exactly, see Figure 3.1. Let us assume that Δt is so small that we can determine $v(x, t_{n+1})$ exactly. To repeat this procedure from time level t_{n+1} to time level t_{n+2}, etc., we must approximate the function $v(\cdot, t_{n+1})$ with a piecewise constant function. By representing $v(\cdot, t_{n+1})$ as a piecewise constant function we will again have a sequence of Riemann problems, this time at time level t_{n+1}, on the form (3.14) and the process can be repeated. The approximation of $v(\cdot, t_{n+1})$ with a piecewise constant function can be done in many different ways, in Godunov's method this is done by letting u_j^{n+1} be the average of $v(\cdot, t_{n+1})$ over the grid-cell I_j, i.e.,

$$u_j^{n+1} = \frac{1}{\Delta x} \int_{I_j} v(x, t_{n+1})\, dx.$$

Just note that this method can be written in conservative form by integrating (3.14) over the rectangle $I_j \times [t_n, t_{n+1})$, which in practice implies that this algorithm can be

considerably simplified, cf. [13]. Lucier [15] has proven the following error estimate for Godunov's method

$$\|u(\cdot,T)-u^N(\cdot,T)\|_1 \le (\Delta x + \frac{2}{\sqrt{3}}(\frac{\Delta x}{\Delta t})^{1/2}(\Delta x T)^{1/2})T.V._x(u^0)$$

where $N\Delta t = T$ and u^N is the solution obtained using the Godunov scheme. The error in Godunov's method is therefore in general of order $(\Delta x)^{1/2}$. Observe that this estimate is rather sharp: one can show that Godunov's method applied to the problem with $f(u) = u$, $u^0(x) = \chi_{(-\infty,0]}(x)$ and $\Delta t/\Delta x = 1/2$ has asymptotic error rate of $1.1283(T\Delta x)^{1/2}$, compared to the above estimate of $1.6330(T\Delta x)^{1/2}$, cf. [15].

Now let us return to the argument which reads as follows. If we use Godunov's method in the splitting (3.8) instead of front tracking, we can not expect to obtain a convergence rate which is better than $1/2$. The reason for this is that the splitting (3.8) with Godunov's scheme reduces to an one-dimensional Godunov method if one of the flux functions vanishes. By looking at the description of Godunov's method we observe that it makes no difference if we use the front tracking scheme or the Godunov scheme in the splitting procedure if Δt is chosen so small that there are no wave interactions. This can be obtained by letting the discretization parameters obey the strict CFL condition

$$\frac{\Delta t}{\Delta x}\max|f'(u)| \le \frac{1}{2}.$$

This implies that we can not, without restrictions on the discretization parameters anyway, expect that our numerical method has a convergence rate of 1. This conclusion will be confirmed in the forthcoming sections.

3.3.1 Convergence of the method of dimensional splitting with exact 1D solution operators

The problem with Theorem 3.1, from our point of view anyway, is that it only expresses the convergence of the dimensional splitting solution $\{u^n_{\Delta t}\}_{n=1}^N$ to the entropy weak solution, it says nothing about the rate of convergence. This is due to the fact that the theorem is proved via a compactness argument. This argument only implies that a sequence of approximate solutions, constructed using the splitting (3.8), lies in some compact set and therefore must contain a convergent subsequence. The limit–function of this subsequence is then proved to satisfy Kružkov's definition of the entropy weak solution (3.2). In this section our aim is to give an explicit estimate of the difference between $u^N_{\Delta t}$ and $S(T)u^0$. The estimate will be proved via an entirely different technique than used in the proof of convergence by Crandall and Majda. A rigorous proof will show that the rate of convergence for the method of dimensional splitting with exact 1D solution operators is no less than $1/2$, i.e.,

Theorem 3.2 *Assume that* $u^0 \in L^1(\mathbf{R}^2) \cap L^\infty(\mathbf{R}^2) \cap BV(\mathbf{R}^2)$ *and* $f,g \in C^1(I)$, *where* $I = [-\|u^0\|_\infty, \|u^0\|_\infty]$. *If* $N\Delta t = T < T_0$ *for some fixed* T_0, *then as* $\Delta t \to 0$ *and* $N \to \infty$,

$$\|S(T)u^0(\cdot,\cdot) - u^N_{\Delta t}(\cdot,\cdot)\|_1 \le K(\Delta t)^{1/2},$$

where K *is a finite constant independent of* Δt, *but dependent on* T_0, $T.V._{(x,y)}(u^0)$, $\|f'\|_\infty$ *and* $\|g'\|_\infty$.

One consequence of this theorem is that no numerical method based on dimensional splitting will in the general case have an error estimate which is better than $\mathcal{O}((\Delta t)^{1/2})$.

In the analysis of the speed of convergence we have to consider the approximate solution $u^n_{\Delta t}$ as function of t, and not only defined at the discrete t-values $t = n\Delta t$. We let $u_{\Delta t} = u_{\Delta t}(x,y,t)$, given by

$$u_\eta(x,y,t) = \begin{cases} S^{f,x}(2(t - n\Delta t))u^n_{\Delta t}(x,y) & \text{for all } t \in [n\Delta t, (n+\frac{1}{2})\Delta t) \\ S^{g,y}(2(t-(n+\frac{1}{2})\Delta t))u^{n+\frac{1}{2}}_{\Delta t}(x,y) & \text{for all } t \in [(n+\frac{1}{2})\Delta t, (n+1)\Delta t) \end{cases}$$

where we recall that

$$\begin{aligned} u^n_{\Delta t}(x,y) &= [S^{g,y}(\Delta t)S^{f,x}(\Delta t)]^n u^0(x,y) \\ u^{n+\frac{1}{2}}_{\Delta t}(x,y) &= S^{f,x}(\Delta t)[S^{g,y}(\Delta t)S^{f,x}(\Delta t)]^n u^0(x,y), \end{aligned}$$

to be our approximate solution. It can easily be seen that $u_{\Delta t}$ coincide with $u^n_{\Delta t}$ when $t = t_n$, $n = 0, 1 \ldots, N$. Before proving the theorem, we will first show that

1. $u_{\Delta t}$ obey a maximum-principle,
2. $u_{\Delta t}$ has the TVD property
3. $u_{\Delta t}$ has L^1–norm which is Lipschitz continuous in the time variable.

The reason for this is that crucial parts of the analysis are based on these three important facts about the approximate solution, which also is true for the exact solution of (3.3).

Lemma 3.1 *There holds*

$$\|u_{\Delta t}(\cdot,\cdot,t)\|_\infty \le \|u^0\|_\infty.$$

Proof This holds since $S^{f,x}$ and $S^{g,y}$ do not introduce new maxima or minima. ∎

Lemma 3.2 *We have*

$$T.V._{(x,y)}(u_{\Delta t}(\cdot,\cdot,t)) \le T.V._{(x,y)}(u^0).$$

Proof First recall that the total variation of a function $h(x)$ of one variable may be defined as

$$T.V._x(h) = \limsup_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_{\mathbf{R}} |h(x) - h(x-\varepsilon)| \, dx. \tag{3.15}$$

Assume first that $t \in [n\Delta t, (n+\frac{1}{2})\Delta t]$. Since the one-dimensional solution has the TVD property we have that

$$T.V._x(u_{\Delta t}(\cdot, y, t)) \leq T.V._x(u_{\Delta t}(\cdot, y, n\Delta t)) \tag{3.16}$$

Furthermore, observe that $u_{\Delta t}(x, y-\varepsilon, t)$ is the solution of

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

with initial data

$$u(x, 0) = u_{\Delta t}(\cdot, y-\varepsilon, n\Delta t)$$

and that $u_{\Delta t}(x, y, t)$ is a solution of

$$v_t + f(v)_x = 0$$

$$v(x, 0) = u_{\Delta t}(\cdot, y, n\Delta t).$$

The stability estimate

$$\int_{\mathbf{R}} |u - v| \, dx \leq \int_{\mathbf{R}} |u(x, 0) - v(x, 0)| \, dx$$

now implies that

$$\int_{\mathbf{R}} |u_{\Delta t}(x, y, t) - u_{\Delta}(x, y-\varepsilon, t)| \, dx \leq \int_{\mathbf{R}} |u_{\Delta t}(x, y, n\Delta t) - u_{\Delta t}(x, y-\varepsilon, n\Delta t)| \, dx.$$

Since this inequality holds for all $\varepsilon > 0$, we have after multiplying both sides by $1/\varepsilon$, integrating in the y–direction, changing the order of integration and taking lim sup on both sides of the inequality (and using Lebesgue's dominated convergence theorem), that

$$\int_{\mathbf{R}} T.V._y(u_{\Delta t}(x, \cdot, t)) \, dx \leq \int_{\mathbf{R}} T.V._y(u_{\Delta t}(x, \cdot, n\Delta t)) \, dx, \tag{3.17}$$

Integrating (3.16) with respect to y and then adding it to (3.17) yields

$$\int_{\mathbf{R}} T.V._x(u_{\Delta t}(\cdot, y, t)) \, dy + \int_{\mathbf{R}} T.V._y(u_{\Delta t}(x, \cdot, t)) \, dx \leq$$
$$\int_{\mathbf{R}} T.V._x(u_{\Delta t}(\cdot, y, n\Delta t)) \, dy + \int_{\mathbf{R}} T.V._y(u_{\Delta t}(x, \cdot, n\Delta t)) \, dx. \tag{3.18}$$

Performing the same analysis for $t \in [(n+\frac{1}{2})\Delta t, (n+1)\Delta t]$ and combining this with (3.18), implies that (3.18) holds for all $t \in [n\Delta t, (n+1)\Delta t]$. The lemma now follows by induction. ∎

From now on let u^0 be a bounded measurable function on $\mathbf{R}^2$, let $|u^0(x,y)| \le B$, and let the flux functions f and g be continuously differentiable in the interval $|u| \le B$, while

$$A = \max\{\max_{|u|\le B}|f'(u)|, \max_{|u|\le B}|g'(u)|\}.$$

Lemma 3.3 *$u_{\Delta t}$ has L^1–norm which is Lipschitz continuous in the time variable, i.e.,*

$$||u_{\Delta t}(\cdot,\cdot,t_2) - u_{\Delta t}(\cdot,\cdot,t_1)||_1 \le A(t_2 - t_1).$$

Proof Recall that if $v = v(x,t)$ is the solution of a one-dimensional conservation law with flux function f, initial data v^0 and $t_2 \ge t_1$, then

$$||v(\cdot,t_2) - v(\cdot,t_1)||_1 \le C(t_2 - t_1)T.V._x(u^0),$$

where $C = \max_{|v|\le||v^0||_\infty}|f'(v)|$. Assume first that $t_1, t_2 \in [n\Delta t, (n+\frac{1}{2})\Delta t)$ or $t_1, t_2 \in [(n+\frac{1}{2})\Delta t, (n+1)\Delta t)$, $t_2 \ge t_1$, then

$$\begin{aligned} ||u_{\Delta t}(\cdot,\cdot,t_2) - u_{\Delta t}(\cdot,\cdot,t_1)||_1 &\le A(t_2-t_1)T.V._{(x,y)}(u_{\Delta t}(\cdot,\cdot,n\Delta t)) \\ &\le A(t_2-t_1)T.V._{(x,y)}(u^0), \end{aligned} \tag{3.19}$$

where we also have used the preceding lemma which says that $u_{\Delta t}$ has the TVD property. Furthermore, if $t_1 \in [n\Delta t, (n+\frac{1}{2})\Delta t]$ and $t_2 \in [(n+\frac{1}{2})\Delta t, (n+1)\Delta t]$, then using (3.19) we obtain

$$\begin{aligned} ||u_{\Delta t}(\cdot,\cdot,t_2) - u_{\Delta t}(\cdot,\cdot,t_1)||_1 &\le ||u_{\Delta t}(\cdot,\cdot,t_2) - u_{\Delta t}(\cdot,\cdot,(n+\frac{1}{2})\Delta t)||_1 \\ &+ ||u_{\Delta t}(\cdot,\cdot,(n+\frac{1}{2})\Delta t) - u_{\Delta t}(\cdot,\cdot,t_1)||_1 \\ &\le A((t_2 - (n+\frac{1}{2})\Delta t) + ((n+\frac{1}{2})\Delta t - t_1))T.V._{(x,y)}(u^0) \\ &= A(t_2-t_1)T.V._{(x,y)}(u^0). \end{aligned} \tag{3.20}$$

More generally assume that $t_1, t_2 \in [0,T]$, $t_2 > t_1$. Then there exists an integer n such that $t_1 \in [n\Delta t, (n+1)\Delta t]$, and an integer m such that $t_2 \in [m\Delta t, (m+1)\Delta t]$ (we may assume $n \ne m$). Using the triangle inequality and (3.20) we achieve the desired result:

$$\begin{aligned} ||u_{\Delta t}(\cdot,\cdot,t_2) - u_{\Delta t}(\cdot,\cdot,t_1)||_1 \le ||u_{\Delta t}(\cdot,\cdot,t_2) - u_{\Delta t}(\cdot,\cdot,t_m)||_1 + ||u_{\Delta t}(\cdot,\cdot,t_m) - u_{\Delta t}(\cdot,\cdot,t_{m-1})||_1 \\ + \cdots + ||u_{\Delta t}(\cdot,\cdot,t_{n+2}) - u_{\Delta t}(\cdot,\cdot,t_{n+1})||_1 + ||u_{\Delta t}(\cdot,\cdot,t_{n+1}) - u_{\Delta t}(\cdot,\cdot,t_1)||_1 \\ \le A(t_2-t_1)T.V._{(x,y)}(u^0). \end{aligned}$$

∎

KUZNETSOV'S APPROXIMATION THEORY: The proof of Theorem 3.2 is based on some rather general theory developed by Kuznetsov ([9],[10]). This theory has already been used to estimate the accuracy of a certain class of approximate methods including, in particular, the method of vanishing viscosity, the method of smoothing, and several widely used finite difference methods, cf. [10]. For instance, the error estimate for Godunov's method was obtained using this technique, cf. [15]. We will first present the general theory and then apply it to our approximate solution $u_{\Delta t}$. So let us begin by introducing some notation. We will consider bounded weak solutions of (3.3) in the half-space $\mathbf{R}^2 \times \mathbf{R}^+$. We select from the set of equivalent solutions those that are continuous in t in the sense of $L^1(\mathbf{R}^2)$ at each point of the halfline $t \geq 0$. Let

$$\mathcal{K}(A) = \{u : u \in L^1(\mathbf{R}^2), |f'(u)|, |g'(u)| \leq A\}.$$

With the L^1–norm on $\mathbf{R}^2$ we introduce the moduli of continuity:

$$\begin{aligned}
\lambda_1(\alpha, u) &= \int_{\mathbf{R}^2} |u(x+\alpha, y) - u(x,y)|\, dx\, dy, \\
\lambda_2(\alpha, u) &= \int_{\mathbf{R}^2} |u(x, y+\alpha) - u(x,y)|\, dy\, dx, \\
\lambda(\delta, u) &= \sup_{|\alpha| \leq \delta} \lambda_1(\alpha, u) + \sup_{|\alpha| \leq \delta} \lambda_2(\alpha, u), \\
\nu_t(\delta, u) &= \sup_{|\alpha| \leq \delta, \alpha \geq -t} \|u(\cdot,\cdot, t+\alpha) - u(\cdot,\cdot,t)\|_1
\end{aligned}$$

If $u : t \to u(t)$ is a solution of (3.3) with $u^0 \in \mathcal{K}(A)$, then, as is well known (cf. [10]), $u(t) \in \mathcal{K}(A)$ and

$$\lambda_i(\delta, u) \leq \lambda_i(\delta, u^0), \quad \nu_t(\delta, u) \leq \tilde{\lambda}(\delta, u^0),$$

where

$$\tilde{\lambda}(\delta, u) = \inf_{\varepsilon} \left\{ 2\lambda(\varepsilon, u) + \frac{\delta A \lambda(\varepsilon, u)}{\varepsilon} \right\}.$$

We denote by $\mathcal{U}$ the space whose elements are functions $u : t \to u(t)$, with domain $[0, \infty)$ and range $L^1(\mathbf{R}^2)$, for which the limit values $u(t\pm)$, exist and which for the sake of definiteness are assumed to be continuous on the right (observe that $u_{\Delta t} \in \mathcal{U}$). For the functions in $\mathcal{U}$, we will use in conjunction with the moduli of continuity ν_t, the moduli

$$\nu_t^{\pm}(\delta, u) = \sup_{0 < \alpha < \delta} \|u(\cdot,\cdot, t \pm \alpha) - u(\cdot,\cdot, t\pm)\|_1.$$

We introduce a non–negative function $\omega \in C_0^\infty$, satisfying $\omega(\sigma) = \omega(-\sigma)$, $\omega(\sigma) \equiv 0$ for $|\sigma| \geq 1$ and $\int_{\mathbf{R}} \omega(\sigma)\, d\sigma = 1$. For $\varepsilon > 0$ let

$$\omega_\varepsilon(\sigma) = \frac{1}{\varepsilon} \omega\left(\frac{1}{\varepsilon}\right)$$

then

$$\begin{aligned}
&\text{a)} \quad \omega_\varepsilon \in C_0^\infty, \quad \omega(\sigma) \geq 0 \quad \forall \sigma \in \mathbf{R} \\
&\text{b)} \quad \int_{\mathbf{R}} \omega_\varepsilon(\sigma) d\sigma = 1 \\
&\text{c)} \quad \omega_\varepsilon(\sigma) \equiv 0 \quad \text{for} \quad |\sigma| \geq \varepsilon \\
&\text{d)} \quad \omega_\varepsilon(\sigma) \leq M_\omega/\varepsilon, \quad |\omega_\varepsilon'(\sigma)| \leq M_\omega/\varepsilon^2,
\end{aligned}$$

where M_ω is a finite constant independent ε. For $\varepsilon > 0$, $\varepsilon_0 > 0$ we define the non–negative test function

$$\Omega_{\varepsilon,\varepsilon_0} = \Omega_{\varepsilon,\varepsilon_0}(x - x', y - y', t - t') = \omega_\varepsilon(x - x')\omega_\varepsilon(y - y')\omega_{\varepsilon_0}(t - t').$$

Furthermore, we introduce with $u, u_\tau \in \mathcal{U}$, $\phi \in C_0^1(\mathbf{R}^2 \times \mathbf{R}^+)$, $\phi \geq 0$, and $a \in \mathbf{R}$ the notation:

$$\begin{aligned}
\mathcal{D}_{T,f,g}(\phi, u, a) &= \int_{\mathbf{R}^2} \int_0^T [\phi_t |u - a| + \phi_x F_1(u, a) + \phi_y F_2(u, a)] \, dt \, dx \, dy \\
&+ \int_{\mathbf{R}^2} \phi(x, y, 0)|u^0(x, y) - a| \, dx \, dy \\
&- \int_{\mathbf{R}^2} \phi(x, y, T)|u(x, y, T - 0) - a| \, dx \, dy, \quad \text{and} \\
\mathcal{D}^{\varepsilon,\varepsilon_0}_{T,f,g}(\Omega_{\varepsilon,\varepsilon_0}, u_\tau, u) &= \int_{\mathbf{R}^2} \int_0^T \mathcal{D}_{T,f,g}(\Omega_{\varepsilon,\varepsilon_0}(x - x', y - y', t - t'), u_\tau, u(x', y', t') \, dt' \, dx' \, dy',
\end{aligned}$$

where we have for notational convenience introduced the auxiliary "flux–functions" $F_1(u, a) = \operatorname{sign}(u - a)(f(u) - f(a))$ and $F_2(u, a) = \operatorname{sign}(u - a)(g(u) - g(a))$. $\mathcal{D}_{T,f,g}$ is sometimes called the "Kružkov form" of (3.3), and if u is the exact solution of (3.3), then $\mathcal{D}_{T,f,g}(\phi, u, a) \geq 0$. Since there will be no ambiguity, the f, g–subscript on $\mathcal{D}_{T,f,g}$ and $\mathcal{D}^{\varepsilon,\varepsilon_0}_{T,f,g}$ will be dropped. The next theorem is due to Kuznetsov [10], and will be our main tool in proving Theorem 3.2.

Theorem 3.3 (Kuznetsov) *For $0 < \varepsilon_0 < T$, $\varepsilon > 0$, $u_\tau \in \mathcal{U}$, and if u is a solution of (3.3), then*

$$\begin{aligned}
\|u(\cdot, \cdot, T) - u_\tau(\cdot, \cdot, T-)\|_1 &\leq \|u^0 - u_\tau^0\|_1 + 2\lambda(\varepsilon, u^0) + \tilde{\lambda}(\varepsilon_0, u^0) \\
&+ \frac{\nu_T^-(\varepsilon_0, u_\tau) + \nu_0^+(\varepsilon_0, u_\tau)}{2} - \mathcal{D}_T^{\varepsilon,\varepsilon_0}(\Omega_{\varepsilon,\varepsilon_0}, u_\tau, u). \quad (3.21)
\end{aligned}$$

For a detailed proof see [10]. In the case when u_τ is a solution of (3.3), $\mathcal{D}_T^{\varepsilon,\varepsilon_0} \geq 0$, so that on neglecting $\mathcal{D}_T^{\varepsilon,\varepsilon_0}$ in (3.21) and letting ε and ε_0 tend to zero, we obtain the well known stability estimate

$$\|u(\cdot, \cdot, T) - u_\tau(\cdot, \cdot, T)\|_1 \leq \|u_\tau^0 - u^0\|_1.$$

In the general case, by using the inequality (3.21) we can estimate how close the function u_τ is to the exact solution of (3.3), in terms of the discrepancy $\mathcal{D}_T^{\varepsilon,\varepsilon_0}$. If $\{u_\tau\}$ is a sequence of functions for which $\mathcal{D}_T^{\varepsilon,\varepsilon_0}(u_\tau) \to 0$ as $\tau \to 0$ for fixed ε and ε_0, then by making use of the arbitrariness in the choice of ε and ε_0, we could in general estimate the rate of convergence to zero of $\|u - u_\tau\|$. We will now apply Kuznetsov's theorem to derive the error estimate stated in Theorem 3.2. The difficulty in proving the theorem is to obtain a proper estimate on $\mathcal{D}_T^{\varepsilon,\varepsilon_0}$. Nevertheless, our next lemma provides us with such a bound on $\mathcal{D}_T^{\varepsilon,\varepsilon_0}$ in terms of ε, ε_0 and Δt.

Lemma 3.4 *Assume that* $u^0 \in L^1(\mathbf{R}^2) \cap L^\infty(\mathbf{R}^2) \cap BV(\mathbf{R}^2)$ *and* $f, g \in C^1(I)$, *where* $I = [-\|u^0\|_\infty, \|u^0\|_\infty]$. *If* $N\Delta t = T < T_0$ *for some fixed* T_0, *then*

$$|\mathcal{D}_T^{\varepsilon,\varepsilon_0}(\Omega_{\varepsilon,\varepsilon_0}, u_{\Delta t}, u)| \le \Delta t \left\{ \frac{K_1(\Delta t + 2\varepsilon_0)}{\varepsilon_0^2} + \frac{K_2}{\varepsilon} \right\},$$

where K_1 *and* K_2 *are finite constants independent of* Δt, *but dependent on* T_0,$T.V._{(x,y)}(u^0)$, $\|f'\|_\infty$ *and* $\|g'\|_\infty$.

Proof of Theorem 3.2 The difference between $u_{\Delta t}$ and u, in L^1–norm, is given by (3.21). So let us establish bounds on the remaining terms in this inequality. We have the following estimates of the moduli of continuity:

$$\begin{aligned}
\lambda(\varepsilon, u^0) &= \sup_{|\alpha|<\varepsilon} \alpha \int_{\mathbf{R}^2} \frac{u^0(x+\alpha, y) - u^0(x,y)}{\alpha}\, dx\, dy \\
&+ \sup_{|\alpha|<\varepsilon} \alpha \int_{\mathbf{R}^2} \frac{u^0(x, y+\alpha) - u^0(x,y)}{\alpha}\, dx\, dy \\
&\le \sup_{|\alpha|<\varepsilon} \alpha \int_{\mathbf{R}} T.V._x(u^0(\cdot, y))\, dy + \sup_{|\alpha|<\varepsilon} \alpha \int_{\mathbf{R}} T.V._y(u^0(x,\cdot))\, dx \\
&= \varepsilon T.V._{(x,y)}(u^0) \\
\tilde{\lambda}(\varepsilon_0, u^0) &= \inf_\delta \left\{ 2\lambda(\delta, u^0) + \frac{\varepsilon_0 A \lambda(\delta, u^0)}{\delta} \right\} \\
&\le \inf_\delta \left\{ 2\delta T.V._{(x,y)}(u^0) + \frac{\varepsilon_0 A \delta T.V._{(x,y)}(u^0)}{\delta} \right\} = \varepsilon_0 T.V._{(x,y)}(u^0) A \\
\nu_t^{\pm}(\varepsilon_0, u_{\Delta t}) &= \sup_{0<\alpha<\varepsilon_0} \|u_{\Delta t}(\cdot,\cdot, t \pm \alpha) - u_{\Delta t}(\cdot,\cdot, t\pm)\|_1 \\
&\le \sup_{0<\alpha<\varepsilon_0} \alpha A T.V._{(x,y)}(u^0) = \varepsilon_0 A T.V._{(x,y)}(u^0)
\end{aligned}$$

Note that $u_{\Delta t}(\cdot,\cdot,0) = u(\cdot,\cdot,0)$. We have in the latter estimate used the fact that $u_{\Delta t}$ has L^1–norm which is Lipschitz continuous in the time variable with Lipschitz constant C, and that $u_{\Delta t}$ is in $BV(\mathbf{R}^2)$, with BV–norm bounded independently of t (the TVD property). These estimates together with Lemma 3.4 now yield

$$\|S(T)u^0(\cdot,\cdot) - u_{\Delta t}(\cdot,\cdot,T)\|_1 \le C_1\varepsilon_0 + C_2\varepsilon + \frac{K_1 \Delta t(\Delta t + 2\varepsilon_0)}{\varepsilon_0^2} + \frac{K_2 \Delta t}{\varepsilon},$$

for some constants C_i, K_i, $i = 1, 2$. Setting $\varepsilon_0 = \varepsilon$ and letting $\varepsilon = (\Delta t)^{1/2}$ ($0 < \varepsilon_0 < T \Rightarrow T > (\Delta t)^{1/2}$) now imply, for some constant K, that

$$\|S(T)u^0(\cdot,\cdot) - u_{\Delta t}(\cdot,\cdot,T)\|_1 \le C_1(\Delta t)^{1/2} + C_2(\Delta t)^{1/2} + K_1((\Delta t) + 2(\Delta t)^{1/2}) + K_2(\Delta t)^{1/2} \le K(\Delta t)^{1/2},$$

as $\Delta t \to 0$. Theorem 3.2 now follows since $u_{\Delta t}(\cdot,\cdot,t)$ coincide with $u^n_{\Delta t}(\cdot,\cdot)$ when $t = t_n$, $n = 0, \dots, N$. ∎

Proof of Lemma 3.4 We first establish a bound on $\mathcal{D}_T(\phi, u_{\Delta t}, a)$ for arbitrary test functions ϕ and constants a. We begin by recording a fact needed in the proof. Assume $v^0 \in L^1(\mathbf{R}^2) \cap L^\infty(\mathbf{R}^2) \cap BV(\mathbf{R}^2)$, $w^0 \in L^1(\mathbf{R}^2) \cap L^\infty(\mathbf{R}^2) \cap BV(\mathbf{R}^2)$, and let $v(x,y,t) \equiv S^{f,x}(t)v^0(\cdot,y)$, $w(x,y,t) \equiv S^{g,y}(t)w^0(x,\cdot)$, denote the corresponding entropy weak solutions of the one–dimensional conservation laws. Under the above assumptions we have the following (cf. [3]): If $\phi \in C^1_0(\mathbf{R}^2 \times \mathbf{R}^+)$, $\phi \ge 0$ and $t_2 \ge t_1 \ge 0$, then

$$\begin{aligned}\int_{\mathbf{R}^2}\int_{t_1}^{t_2} |v-a| + F_1(v,a)\phi_x \, dt\, dx\, dy &\ge \int_{\mathbf{R}^2} |v(x,y,t_2) - a|\phi(x,y,t_2)\, dx\, dy \\ &\quad - \int_{\mathbf{R}^2} |v(x,y,t_1) - a|\phi(x,y,t_1)\, dx\, dy \quad (3.22)\\ \int_{\mathbf{R}^2}\int_{t_1}^{t_2} |w-a| + F_2(w,a)\phi_y \, dt\, dx\, dy &\ge \int_{\mathbf{R}^2} |w(x,y,t_2) - a|\phi(x,y,t_2)\, dx\, dy \\ &\quad - \int_{\mathbf{R}^2} |w(x,y,t_1) - a|\phi(x,y,t_1)\, dx\, dy \quad (3.23)\end{aligned}$$

Let us now consider a new test function $\tilde\phi$ defined by $\tilde\phi(x,y,t) = \phi(x,y,t/2)$, furthermore, define $v_n(t) \equiv S^{f,x}(t)[S^{g,y}(\Delta t)S^{f,x}(\Delta t)]^n u^0$ and
$w_n(t) \equiv S^{g,y}(t)S^{f,x}(\Delta t)[S^{g,y}(\Delta t)S^{f,x}(\Delta t)]^n u^0$ and apply (3.22) and (3.23) to $v_n(t)$ and $w_n(t)$ respectively, to obtain (we put $\tau = 2(t - n\Delta t)$, $d\tau = 2dt \Rightarrow 1/2\phi_t(\tau/2 + n\Delta t) = \phi_\tau(1/2(\tau + 2n\Delta t)) = \tilde\phi_\tau(\tau + 2n\Delta t)$)

$$\int_{\mathbf{R}^2}\int_{n\Delta t}^{(n+\frac12)\Delta t} \frac12\phi_t|u_{\Delta t} - a| + \phi_x F_1(u_{\Delta t}, a)\, dt\, dx\, dy \equiv$$

$$\frac12\int_{\mathbf{R}^2}\int_0^{\Delta t} \tilde\phi_\tau(x,y,\tau + 2n\Delta t)|v_n(\tau) - a| + \tilde\phi_x(x,y,\tau + 2n\Delta t)F_1(v_n(\tau), a)\, d\tau\, dx\, dy$$

$$\ge \frac12\int_{\mathbf{R}^2} \phi(x,y,(n+\frac12)\Delta t)|u_{\Delta t}(x,y,(n+\frac12)\Delta t) - a|\, dx\, dy$$

$$-\frac12\int_{\mathbf{R}^2} \phi(x,y,n\Delta t)|u_{\Delta t}(x,y,n\Delta t) - a|\, dx\, dy \quad (3.24)$$

and similarly

$$\int_{\mathbf{R}^2}\int_{(n+\frac12)\Delta t}^{(n+1)\Delta t} \frac12\phi_t|u_{\Delta t} - a| + \phi_y F_2(u_{\Delta t}, a)\, dt\, dx\, dy$$

$$\geq \frac{1}{2}\int_{\mathbf{R}^2} \phi(x,y,(n+1)\Delta t)|u_{\Delta t}(x,y,(n+1)\Delta t) - a|\,dx\,dy$$
$$-\frac{1}{2}\int_{\mathbf{R}^2} \phi(x,y,(n+\frac{1}{2})\Delta t)|u_{\Delta t}(x,y,(n+\frac{1}{2})\Delta t) - a|\,dx\,dy. \tag{3.25}$$

We add the inequalities (3.24) and (3.25) and sum over n. We observe that the right hand side of the resulting inequality is a telescoping sum, letting χ_n denote the characteristic function of $\{(x,y,t) : n\Delta t \leq t \leq (n+\frac{1}{2})\Delta t\}$, then the resulting inequality has the form

$$\int_{\mathbf{R}^2}\int_0^T \frac{1}{2}\phi_t|u_{\Delta t} - a|\,dt\,dx\,dy +$$
$$\sum_{n=0}^{N-1}\int_{\mathbf{R}^2}\int_{n\Delta t}^{(n+1)\Delta t} (\chi_n F_1(u_{\Delta t},a)\phi_x + (1-\chi_n)F_2(u_{\Delta t},a)\phi_y)\,dt\,dx\,dy \geq$$
$$\frac{1}{2}\int_{\mathbf{R}^2}\phi(x,y,T)|u_{\Delta t}(x,y,T) - a|\,dx\,dy - \frac{1}{2}\int_{\mathbf{R}^2}\phi(x,y,0)|u_{\Delta t}(x,y,0) - a|\,dx\,dy. \tag{3.26}$$

This inequality was shown by Crandall and Majda in their fundamental article on dimensional splitting [3]. There they also showed that passing to the limit in this inequality, leads to (3.4). Multiplying both sides of (3.26) by 2 and adding equal terms on both sides, yields

$$\int_{\mathbf{R}^2}\int_0^T \phi_t|u_{\Delta t} - a| + \phi_x F_1(u_{\Delta t},a) + \phi_y F_2(u_{\Delta t},a)\,dt\,dx\,dy$$
$$\geq -\left\{\sum_{n=0}^{N-1}\int_{\mathbf{R}^2}\int_{n\Delta t}^{(n+1)\Delta t} 2\chi_n F_1(u_{\Delta t},a)\phi_x\,dt\,dx\,dy - \int_{\mathbf{R}^2}\int_0^T F_1(u_{\Delta t},a)\phi_x\,dt\,dx\,dy\right\}$$
$$-\left\{\sum_{n=0}^{N-1}\int_{\mathbf{R}^2}\int_{n\Delta t}^{(n+1)\Delta t} 2(1-\chi_n)F_2(u_{\Delta t},a)\phi_y\,dt\,dx\,dy - \int_{\mathbf{R}^2}\int_0^T F_2(u_{\Delta t},a)\phi_y\,dt\,dx\,dy\right\}$$
$$+\int_{\mathbf{R}^2}\phi(x,y,T)|u_{\Delta t}(x,y,T) - a|\,dx\,dy - \int_{\mathbf{R}^2}\phi(x,y,0)|u_{\Delta t}(x,y,0) - a|\,dx\,dy.$$

This implies that

$$-\mathcal{D}_T(\phi,u_{\Delta t},a) \leq \mathcal{I}_1 + \mathcal{I}_2,$$

where $\mathcal{I}_i = \mathcal{I}_i(\phi,u_{\Delta t},a)$, $i = 1,2$ is defined by

$$\mathcal{I}_1 = \sum_{n=0}^{N-1}\int_{\mathbf{R}^2}\int_{n\Delta t}^{(n+1)\Delta t} 2\chi_n F_1(u_{\Delta t},a)\phi_x\,dt\,dx\,dy - \int_{\mathbf{R}^2}\int_0^T F_1(u_{\Delta t},a)\phi_x\,dt\,dx\,dy$$
$$\mathcal{I}_2 = \sum_{n=0}^{N-1}\int_{\mathbf{R}^2}\int_{n\Delta t}^{(n+1)\Delta t} 2(1-\chi_n)F_2(u_{\Delta t},a)\phi_y\,dt\,dx\,dy - \int_{\mathbf{R}^2}\int_0^T F_2(u_{\Delta t},a)\phi_y\,dt\,dx\,dy.$$

In the remaining analysis we substitute in $u(x',y',t')$ for a and $\Omega_{\varepsilon,\varepsilon_0}$ for ϕ. $\mathcal{I}_1$ and $\mathcal{I}_2$ then take the form

$$\mathcal{I}_1 = \sum_{n=0}^{N-1}\int_{\mathbf{R}^2}\omega_\varepsilon(x-x')_x\omega_\varepsilon(y-y')\Big(\int_{n\Delta t}^{(n+\frac{1}{2})\Delta t} 2F_1(u_{\Delta t},u(x',y',t'))\omega_{\varepsilon_0}(t-t')\,dt$$

$$- \int_{n\Delta t}^{(n+1)\Delta t} F_1(u_{\Delta t}, u(x', y', t'))\omega_{\varepsilon_0}(t - t')\, dt)dx\, dy$$

$$\mathcal{I}_2 = \sum_{n=0}^{N-1} \int_{\mathbf{R}^2} \omega_\varepsilon(x - x')\omega_\varepsilon(y - y')_y (\int_{(n+\frac{1}{2})\Delta t}^{(n+1)\Delta t} 2F_1(u_{\Delta t}, u(x', y', t'))\omega_{\varepsilon_0}(t - t')\, dt$$

$$- \int_{n\Delta t}^{(n+1)\Delta t} F_1(u_{\Delta t}, u(x', y', t'))\omega_{\varepsilon_0}(t - t')\, dt)dx\, dy.$$

Furthermore, write

$$\begin{aligned} F_1(u_{\Delta t}(t), u(x', y', t')) &= F_1(u_{\Delta t}(n\Delta t), u(x', y', t')) \\ &+ (F_1(u_{\Delta t}(t), u(x', y', t')) - F_1(u_{\Delta t}(n\Delta t), u(x', y', t'))). \end{aligned} \tag{3.27}$$

This puts us in a position to employ the following lemma to $\mathcal{I}_1$ (and $\mathcal{I}_2$).

Lemma 3.5 *If the function* $f(u)$ *satisfies an Lipschitz condition on the interval* $[-M, M]$ *with constant* L*, then the function* $F(u, v) = \operatorname{sign}(u - v)[f(u) - f(v)]$ *also satisfies the Lipschitz condition in* u *and* v *with the constant* L*.*

Proof To prove this, it suffices to take into account that $F_u(u, v) = \operatorname{sign}(u - v)f'(u)$ for fixed $v \in [-M, M]$ and almost all $u \in [-M, M]$, and that $F_v(u, v) = \operatorname{sign}(v - u)f'(v)$ for fixed $u \in [-M, M]$ and almost all $v \in [-M, M]$. ■

Define $\mathcal{I}_1^{\varepsilon,\varepsilon_0} = \mathcal{I}_1^{\varepsilon,\varepsilon_0}(\Omega_{\varepsilon,\varepsilon_0}, u_{\Delta t}, u)$ by

$$\mathcal{I}_1^{\varepsilon,\varepsilon_0} = \int_{\mathbf{R}^2} \int_0^T \mathcal{I}_1(\Omega_{\varepsilon,\varepsilon_0}(x - x', y - y', t - t'), u_{\Delta t}(x, y, t), u(x', y', t'))\, dt'\, dx'\, dy',$$

and $\mathcal{I}_2^{\varepsilon,\varepsilon_0}$ likewise, only with $\mathcal{I}_2$ replaced by $\mathcal{I}_1$. Observe that

$$-\mathcal{D}_T^{\varepsilon,\varepsilon_0}(\Omega_{\varepsilon,\varepsilon_0}, u_{\Delta t}, u) \le \mathcal{I}_1^{\varepsilon,\varepsilon_0} + \mathcal{I}_2^{\varepsilon,\varepsilon_0}.$$

The substitution of (3.27) into $\mathcal{I}_1^{\varepsilon,\varepsilon_0}$ yields the equality:

$$\begin{aligned} &|\mathcal{I}_1^{\varepsilon,\varepsilon_0}| = \\ &\quad |\sum_{n=0}^{N-1} \int_{\mathbf{R}^2} \int_{\mathbf{R}^2} \omega_\varepsilon(x - x')_x \omega_\varepsilon(y - y') \times \\ &\quad (\int_0^T \int_{n\Delta t}^{(n+\frac{1}{2})\Delta t} 2F_1(u_{\Delta t}(n\Delta t), u(x', y', t'))\omega_{\varepsilon_0}(t - t')\, dt\, dt' \\ &\quad - \int_0^T \int_{n\Delta t}^{(n+1)\Delta t} F_1(u_{\Delta t}(n\Delta t), u(x', y', t'))\omega_{\varepsilon_0}(t - t')\, dt\, dt')\, dx\, dy\, dx'\, dy' \\ &\quad + \sum_{n=0}^{N-1} \int_{\mathbf{R}^2} \int_{\mathbf{R}^2} \omega_\varepsilon(x - x')_x \omega_\varepsilon(y - y') \times \end{aligned}$$

$$
\begin{aligned}
&\Big(\int_0^T \int_{n\Delta t}^{(n+\frac{1}{2})\Delta t} 2(F_1(u_{\Delta t}(t), u(x', y', t')) - F_1(u_{\Delta t}(n\Delta t), u(x', y', t'))\omega_{\varepsilon_0}(t - t')\,dt\,dt' \\
&- \int_0^T \int_{n\Delta t}^{(n+1)\Delta t} (F_1(u_{\Delta t}(t), u(x', y', t')) - F_1(u_{\Delta t}(n\Delta t), u(x', y', t')))\omega_{\varepsilon_0}(t - t')\,dt\,dt') \times \\
&dx\,dy\,dx'\,dy'\Big| \\
&\equiv |\mathcal{I}_{1,1}^{\varepsilon,\varepsilon_0} + \mathcal{I}_{1,2}^{\varepsilon,\varepsilon_0}|.
\end{aligned}
$$

To estimate $\mathcal{I}_{1,1}^{\varepsilon,\varepsilon_0}$ we use the fact that $\omega_{\varepsilon_0}(t-t')$ locally has a Taylor expansion. For each $t \in [n\Delta t, (n+1)\Delta t]$, $t' \in [0, T]$ let $\omega_{\varepsilon_0}(t - t') = \omega_{\varepsilon_0}(n\Delta t - t') + \omega'_{\varepsilon_0}(z_n - t')(t - n\Delta t)$, where z_n is some number between $n\Delta t$ and t.[4] This implies that

$$
\begin{aligned}
&\int_0^T \int_{n\Delta t}^{(n+\frac{1}{2})\Delta t} 2F_1(u_{\Delta t}(n\Delta t), u(x', y', t'))\omega_{\varepsilon_0}(t - t')\,dt\,dt' \\
&\quad - \int_0^T \int_{n\Delta t}^{(n+1)\Delta t} F_1(u_{\Delta t}(n\Delta t), u(x', y', t'))\omega_{\varepsilon_0}(t - t')\,dt\,dt' \\
&\quad = \int_0^T 2F_1(u_{\Delta t}(n\Delta t), u(x', y', t')) \int_{n\Delta t}^{(n+\frac{1}{2})\Delta t} \omega_{\varepsilon_0}(n\Delta t - t') + \omega'_{\varepsilon_0}(z_n - t')(t - n\Delta t)\,dt\,dt' \\
&\quad - \int_0^T F_1(u_{\Delta t}(n\Delta t), u(x', y', t')) \int_{n\Delta t}^{(n+1)\Delta t} \omega_{\varepsilon_0}(n\Delta t - t') + \omega'_{\varepsilon_0}(z_n - t')(t - n\Delta t)\,dt\,dt' \\
&\quad = \Delta t \int_0^T F_1(u_{\Delta t}(n\Delta t), u(x', y', t'))\omega_{\varepsilon_0}(n\Delta t - t')\,dt' \\
&\quad + \int_0^T \int_{n\Delta t}^{(n+\frac{1}{2})\Delta t} 2F_1(u_{\Delta t}(n\Delta t), u(x', y', t'))\omega'_{\varepsilon_0}(z_n - t')(t - n\Delta t)\,dt\,dt' \\
&\quad - \Delta t \int_0^T F_1(u_{\Delta t}(n\Delta t), u(x', y', t'))\omega_{\varepsilon_0}(n\Delta t - t')\,dt' \\
&\quad - \int_0^T \int_{n\Delta t}^{(n+1)\Delta t} F_1(u_{\Delta t}(n\Delta t), u(x', y', t'))\omega'_{\varepsilon_0}(z_n - t')(t - n\Delta t)\,dt\,dt' \\
&\quad = \int_0^T \int_{n\Delta t}^{(n+\frac{1}{2})\Delta t} 2F_1(u_{\Delta t}(n\Delta t), u(x', y', t'))\omega'_{\varepsilon_0}(z_n - t')(t - n\Delta t)\,dt\,dt' \\
&\quad - \int_0^T \int_{n\Delta t}^{(n+1)\Delta t} F_1(u_{\Delta t}(n\Delta t), u(x', y', t'))\omega'_{\varepsilon_0}(z_n - t')(t - n\Delta t)\,dt\,dt'. \qquad (3.28)
\end{aligned}
$$

Using (3.28), we have

$$
\begin{aligned}
|\mathcal{I}_{1,1}^{\varepsilon,\varepsilon_0}| = |2\sum_{n=0}^{N-1} \int_{\mathbf{R}^2} \int_{\mathbf{R}^2} \int_0^T \int_{n\Delta t}^{(n+\frac{1}{2})\Delta t} &\omega_\varepsilon(x - x')_x \omega_\varepsilon(y - y')F_1(u_{\Delta t}(x, y, n\Delta t), u(x', y', t')) \\
&\omega'_{\varepsilon_0}(z_n - t')(t - n\Delta t)\,dt\,dt'\,dx\,dy\,dx'\,dy'
\end{aligned}
$$

[4] Note that $\omega'_{\varepsilon_0}(t - t') = \frac{d}{dt}(\omega_{\varepsilon_0}(t - t'))$.

$$-\sum_{n=0}^{N-1}\int_{\mathbf{R}^2}\int_{\mathbf{R}^2}\int_0^T\int_{n\Delta t}^{(n+1)\Delta t}\omega_\varepsilon(x-x')_x\omega_\varepsilon(y-y')F_1(u_{\Delta t}(x,y,n\Delta t),u(x',y',t'))\times$$
$$\omega'_{\varepsilon_0}(z_n-t')(t-n\Delta t)\,dt\,dt'\,dx\,dy\,dx'\,dy'|$$
$$\equiv|\mathcal{I}^{\varepsilon,\varepsilon_0}_{1,1,1}-\mathcal{I}^{\varepsilon,\varepsilon_0}_{1,1,2}|.$$

Furthermore, since $\int_{\mathbf{R}}\omega_\varepsilon(x-x')_x\,dx=0$, we have

$$2\sum_{n=0}^{N-1}\int_{\mathbf{R}^2}\int_{\mathbf{R}^2}\int_0^T\int_{n\Delta t}^{(n+\frac12)\Delta t}\omega_\varepsilon(x-x')_x\omega_\varepsilon(y-y')F_1(u_{\Delta t}(x',y,n\Delta t),u(x',y',t'))$$
$$\omega'_{\varepsilon_0}(z_n-t')(t-n\Delta t)\,dt\,dt'\,dx\,dy\,dx'\,dy'=0.$$

Using this fact and the fact that F_1 satisfies a Lipschitz condition with constant A, it follows that:

$$|\mathcal{I}^{\varepsilon,\varepsilon_0}_{1,1,1}|=$$
$$2|\sum_{n=0}^{N-1}\int_{\mathbf{R}^2}\int_{\mathbf{R}^2}\int_0^T\int_{n\Delta t}^{(n+\frac12)\Delta t}\omega_\varepsilon(x-x')_x\omega_\varepsilon(y-y')\omega'_{\varepsilon_0}(z_n-t')(t-n\Delta t))\times$$
$$(F_1(u_{\Delta t}(x,y,n\Delta t),u(x',y',t'))-F_1(u_{\Delta t}(x',y,n\Delta t),u(x',y',t'))\,dt\,dt'\,dx\,dy\,dx'\,dy'|$$
$$\le 2A\Delta t\sum_{n=0}^{N-1}\int_{\mathbf{R}^2}\int_{\mathbf{R}^2}\int_0^T\int_{n\Delta t}^{(n+\frac12)\Delta t}|\omega_\varepsilon(x-x')_x|\omega_\varepsilon(y-y')|u_{\Delta t}(x,y,n\Delta t)-u_{\Delta t}(x',y,n\Delta t)|\times$$
$$|\omega'_{\varepsilon_0}(z_n-t')|\,dt\,dt'\,dx\,dy\,dx'\,dy'.$$

Observe that

$$\begin{aligned}
&\int_{\mathbf{R}^2}\int_{\mathbf{R}^2}|\omega_\varepsilon(x-x')_x||u_{\Delta t}(x,y,n\Delta t)-u_{\Delta t}(x',y,n\Delta t)|\omega_\varepsilon(y-y')\,dx\,dy\,dx'\,dy'\\
&\quad=\int_{\mathbf{R}^2}\int_{\mathbf{R}}|\omega'_\varepsilon(z)||u_{\Delta t}(x,y,n\Delta t)-u_{\Delta t}(x-z,y,n\Delta t)|\,dx\,dy\,dz\\
&\quad=\int_{\mathbf{R}^2}|\omega'_\varepsilon(z)||z|\int_{\mathbf{R}}\frac{|u_{\Delta t}(x,y,n\Delta t)-u_{\Delta t}(x-z,y,n\Delta t)|}{|z|}\,dx\,dy\,dz\\
&\quad\le\int_{\mathbf{R}^2}|\omega'_\varepsilon(z)||z|T.V._x(u_{\Delta t}(\cdot,y,n\Delta t))\,dz\,dy=\int_{\mathbf{R}}T.V._x(u_{\Delta t}(\cdot,y,n\Delta t))\,dy\\
&\quad\le\int_{\mathbf{R}}T.V._x(u_{\Delta t}(\cdot,y,n\Delta t))\,dy+\int_{\mathbf{R}}T.V._y(u_{\Delta t}(x,\cdot,n\Delta t))\,dx\\
&\quad=T.V._{(x,y)}(u_{\Delta t}(\cdot,\cdot,n\Delta t))\le T.V._{(x,y)}(u^0),
\end{aligned}\tag{3.29}$$

where we have used Fubini's theorem, $\int_{\mathbf{R}}\omega_\varepsilon(y-y')\,dy'=1$, the definition of the total variation (3.15), integration by parts to show that $\int_{\mathbf{R}}|\omega'_\varepsilon(z)||z|\,dz=1$, and the TVD property of $u_{\Delta t}$. With this inequality we obtain that

$$|\mathcal{I}^{\varepsilon,\varepsilon_0}_{1,1,1}|\ \le\ 2T.V._{(x,y)}(u^0)A\Delta t\sum_{n=0}^{N-1}\int_{n\Delta t}^{(n+\frac12)\Delta t}\int_0^T|\omega'_{\varepsilon_0}(z_n-t')|\,dt'\,dt.$$

We can not simply perform the integration of this expression (with respect to t or t') since z_n is depending on t and t'; $z_n = z_n(t,t')$. So fix $t \in [n\Delta t, (n+\frac{1}{2})\Delta t]$, using the properties of ω_{ε_0} we have that $\omega'_{\varepsilon_0}(z_n(t,t') - t') \equiv 0$ for $|z_n(t,t') - t'| \geq \varepsilon_0$. Let us therefore determine the largest set of t' such that $|z_n(t,t') - t'| < \varepsilon_0$. We have, since $z_n(t,t') \in (n\Delta t, (n+1)\Delta t)$ for all $t \in [n\Delta t, (n+1)\Delta t]$ and $t' \in [0,T]$, that

$$\begin{array}{rcl} |z_n(t,t') - t'| & < & \varepsilon_0 \\ & \Downarrow & \\ z_n(t,t') - \varepsilon_0 < & t' & < z_n(t,t') + \varepsilon_0 \\ & \Downarrow & \\ n\Delta t - \varepsilon_0 < & t' & < (n+1)\Delta t + \varepsilon_0. \end{array}$$

This implies that ($\omega'_{\varepsilon_0}(z_n(t,t') - t') \equiv 0$ when t' is not in $[n\Delta t - \varepsilon_0, (n+1)\Delta t + \varepsilon_0]$ [5])

$$\begin{aligned} |\mathcal{I}^{\varepsilon,\varepsilon_0}_{1,1,1}| &\leq 2T.V._{(x,y)}(u^0)A\Delta t \sum_{n=0}^{N-1} \int_{n\Delta t}^{(n+\frac{1}{2})\Delta t} \int_{n\Delta t - \varepsilon_0}^{(n+1)\Delta t + \varepsilon_0} |\omega'_{\varepsilon_0}(z_n(t,t') - t')|\, dt'\, dt \\ &\leq 2T.V._{(x,y)}(u^0)A\Delta t \sum_{n=0}^{N-1} \int_{n\Delta t}^{(n+\frac{1}{2})\Delta t} \int_{n\Delta t - \varepsilon_0}^{(n+1)\Delta t + \varepsilon_0} \frac{M_\omega}{\varepsilon_0^2}\, dt'\, dt \\ &\leq T.V._{(x,y)}(u^0)AT\Delta t \frac{M_\omega(\Delta t + 2\varepsilon_0)}{\varepsilon_0^2}. \end{aligned}$$

$|\mathcal{I}^{\varepsilon,\varepsilon_0}_{1,1,2}|$ is bounded by the same expression which implies that

$$|\mathcal{I}^{\varepsilon,\varepsilon_0}_{1,1}| \leq |\mathcal{I}^{\varepsilon,\varepsilon_0}_{1,1,1}| + |\mathcal{I}^{\varepsilon,\varepsilon_0}_{1,1,2}| \leq 2AT.V._{(x,y)}(u^0)T\Delta t \frac{M_\omega(\Delta t + 2\varepsilon_0)}{\varepsilon_0^2}.$$

To finally bound $|\mathcal{I}^{\varepsilon,\varepsilon_0}_1|$ we must first derive a proper upper bound on $|\mathcal{I}^{\varepsilon,\varepsilon_0}_{1,2}|$. Recall that $|\mathcal{I}^{\varepsilon,\varepsilon_0}_{1,2}|$ takes the form

$$\begin{aligned} |\mathcal{I}^{\varepsilon,\varepsilon_0}_{1,2}| = |\sum_{n=0}^{N-1} \int_{\mathbf{R}^2} \int_{\mathbf{R}^2} \int_0^T \int_{n\Delta t}^{(n+\frac{1}{2})\Delta t} & 2(F_1(u_{\Delta t}(t), u(x',y',t')) - F_1(u_{\Delta t}(n\Delta t), u(x',y',t')) \times \\ & \omega_\varepsilon(x-x')_x \omega_\varepsilon(y-y')\omega_{\varepsilon_0}(t-t')\, dt\, dt'\, dx\, dy\, dx'\, dy' \\ - \sum_{n=0}^{N-1} \int_{\mathbf{R}^2} \int_{\mathbf{R}^2} \int_0^T \int_{n\Delta t}^{(n+1)\Delta t} & (F_1(u_{\Delta t}(t), u(x',y',t')) - F_1(u_{\Delta t}(n\Delta t), u(x',y',t')) \times \\ & \omega_\varepsilon(x-x')_x \omega_\varepsilon(y-y')\omega_{\varepsilon_0}(t-t')\, dt\, dt'\, dx\, dy\, dx'\, dy'| \\ & \equiv |\mathcal{I}^{\varepsilon,\varepsilon_0}_{1,2,1} - \mathcal{I}^{\varepsilon,\varepsilon_0}_{1,2,2}|. \end{aligned}$$

[5] Actually, $\omega'_{\varepsilon_0}(z_n(t,t') - t') \equiv 0$ when t' is not in $[n\Delta t - \varepsilon_0, (n+\frac{1}{2})\Delta t + \varepsilon_0]$. But by considering $[n\Delta t - \varepsilon_0, (n+1)\Delta t + \varepsilon_0]$ we have automatically covered the estimation of $\mathcal{I}^{\varepsilon,\varepsilon_0}_{1,1,2}$.

If we once more make use of Fubini's theorem and exploit the fact that $u_{\Delta t}$ has L^1–norm which is Lipschitz continuous in the time variable, then we have the following

$$
\begin{aligned}
&|\mathcal{I}_{1,2,1}^{\varepsilon,\varepsilon_0}| = \\
&\quad |2\sum_{n=0}^{N-1}\int_{\mathbf{R}^2}\int_{\mathbf{R}^2}\int_0^T\int_{n\Delta t}^{(n+\frac{1}{2})\Delta t}(F_1(u_{\Delta t}(t),u(x',y',t'))-F_1(u_{\Delta t}(n\Delta t),u(x',y',t')))\times \\
&\quad \omega_\varepsilon(x-x')_x\omega_\varepsilon(y-y')\omega_{\varepsilon_0}(t-t')\,dt\,dt'\,dx\,dy\,dx'\,dy'| \\
&\quad \le 2A\sum_{n=0}^{N-1}\int_{n\Delta t}^{(n+\frac{1}{2})\Delta t}\int_{\mathbf{R}^2}\int_{\mathbf{R}^2}\int_0^T|u_{\Delta t}(x,y,t)-u_{\Delta t}(x,y,n\Delta t)||\omega_\varepsilon(x-x')_x|\omega_\varepsilon(y-y')\times \\
&\quad \omega_{\varepsilon_0}(t-t')\,dt'\,dx'\,dy'\,dx\,dy\,dt \\
&\quad \le 4A\frac{M_\omega}{\varepsilon}\sum_{n=0}^{N-1}\int_{n\Delta t}^{(n+\frac{1}{2})\Delta t}\int_{\mathbf{R}^2}|u_{\Delta t}(x,y,t)-u_{\Delta t}(x,y,n\Delta t)|\,dx\,dy\,dt \\
&\quad \le 4A\frac{M_\omega}{\varepsilon}\sum_{n=0}^{N-1}\int_{n\Delta t}^{(n+\frac{1}{2})\Delta t}A|t-n\Delta t|T.V._{(x,y)}(u^0)\,dt \le T.V._{(x,y)}(u^0)A^2T\Delta t\frac{M_\omega}{\varepsilon},
\end{aligned}
$$

where we have used that $\int_{\mathbf{R}}\omega_\varepsilon(y-y')\,dy'=1$, $\int_0^T\omega_{\varepsilon_0}(t-t')\,dt'\le 1$ and $\int_{\mathbf{R}}|\omega_\varepsilon(x-x')_x|\,dx'\le 2M_\omega/\varepsilon$. Similarly

$$
\begin{aligned}
&|\mathcal{I}_{1,2,2}^{\varepsilon,\varepsilon_0}| = \\
&\quad |\sum_{n=0}^{N-1}\int_{\mathbf{R}^2}\int_{\mathbf{R}^2}\int_0^T\int_{n\Delta t}^{(n+1)\Delta t}(F_1(u_{\Delta t}(t),u(x',y',t'))-F_1(u_{\Delta t}(n\Delta t),u(x',y',t'))\times \\
&\quad \omega_\varepsilon(x-x')_x\omega_\varepsilon(y-y')\omega_{\varepsilon_0}(t-t')\,dt\,dt'\,dx\,dy\,dx'\,dy'| \\
&\quad \le 2T.V._{(x,y)}(u^0)A^2T\Delta t\frac{M_\omega}{\varepsilon}.
\end{aligned}
$$

From this we conclude that the following estimate holds:

$$
|\mathcal{I}_{1,2}^{\varepsilon,\varepsilon_0}| \le |\mathcal{I}_{1,2,1}^{\varepsilon,\varepsilon_0}| + |\mathcal{I}_{1,2,2}^{\varepsilon,\varepsilon_0}| \le 3T.V._{(x,y)}(u^0)A^2T\Delta t\frac{M_\omega}{\varepsilon}.
$$

And therefore

$$
\begin{aligned}
|\mathcal{I}_1^{\varepsilon,\varepsilon_0}| &\le |\mathcal{I}_{1,1}^{\varepsilon,\varepsilon_0}| + |\mathcal{I}_{1,2}^{\varepsilon,\varepsilon_0}| \le 2T.V._{(x,y)}(u^0)AT\Delta t\frac{M_\omega(\Delta t+2\varepsilon_0)}{\varepsilon_0^2} + 3T.V._{(x,y)}(u^0)A^2T\Delta t\frac{M_\omega}{\varepsilon} \\
&= \Delta t\left\{\frac{K_1'(\Delta t+2\varepsilon_0)}{\varepsilon_0^2}+\frac{K_2'}{\varepsilon}\right\}.
\end{aligned}
$$

An identical analysis for $\mathcal{I}_2^{\varepsilon,\varepsilon_0}$ results in the same bound and the desired estimate follows:

$$
|\mathcal{D}_T^{\varepsilon,\varepsilon_0}(\Omega_{\varepsilon,\varepsilon_0},u_{\Delta t},u)| \le |\mathcal{I}_1^{\varepsilon,\varepsilon_0}| + |\mathcal{I}_2^{\varepsilon,\varepsilon_0}| \le \Delta t\left\{\frac{K_1(\Delta t+2\varepsilon_0)}{\varepsilon_0^2}+\frac{K_2}{\varepsilon}\right\},
$$

for some constants K_1 and K_2. This concludes the proof of Lemma 3.4. ∎

3.3.2 Convergence of the method of dimensional splitting for front tracking

We will again use the theory developed by Kuznetsov , but this time applied to the approximate solution generated by the method of dimensional splitting with the exact 1D solution operators replaced by the front tracking scheme; $\{u_\eta^n\}_{n=1}^N$. One hypothesis in Kuznetsov's theorem (as presented above) demands that f and g are locally C^1, but a refined argument shows that the flux functions only need to be Lipschitz continuous. We will therefore first derive an error estimate for our numerical method when it is applied to problem (3.3) with the flux functions f and g piecewise linear. More precisely we prove the following lemma:

Lemma 3.6 *Assume that (i)* $u^0 \in L^1(\mathbf{R}^2) \cap L^\infty(\mathbf{R}^2) \cap BV(\mathbf{R}^2)$ *and (ii)* f *and* g *are piecewise linear and continuous functions on* $I = [-\|u^0\|_\infty, \|u^0\|_\infty]$. *If* $N\Delta t = T < T_0$ *for some fixed* T_0 *and* $\Delta x = \Delta y = \mathcal{O}(\Delta t)$, *then as* $\eta = (\Delta x, \Delta y, \Delta t) \to 0$ *and* $N \to \infty$,

$$\|S(T)u^0(\cdot,\cdot) - u_\eta^N(\cdot,\cdot)\|_1 \leq K((\Delta t)^{1/2} + (\Delta x)^{1/2}),$$

where K *is a finite constant independent of* η, *but dependent on* T_0, $T.V._{(x,y)}(u^0)$, $\|f'\|_\infty$ *and* $\|g'\|_\infty$.

Proof We first have to decide how to define our approximate solution between $n\Delta t$ and $(n+1)\Delta t$, $n = 0, \ldots, N-1$. Essentially the same definition as before will be used. Let $u_\eta = u_\eta(x,y,t)$ be given by

$$u_\eta(x,y,t) = \begin{cases} S^{f,x}(2(t-n\Delta t))u_\eta^n(x,y) & \text{for all } t \in [n\Delta t, (n+\frac{1}{2})\Delta t) \\ S^{g,y}(2(t-(n+\frac{1}{2})\Delta t))u_\eta^{n+\frac{1}{2}}(x,y) & \text{for all } t \in [(n+\frac{1}{2})\Delta t, (n+1)\Delta t), \end{cases}$$

where

$$\begin{aligned} u_\eta^n(x,y) &= [\pi S^{g,y}(\Delta t)\pi S^{f,x}(\Delta t)]^n u_\eta^0(x,y) \\ u_\eta^{n+\frac{1}{2}}(x,y) &= \pi S^{f,x}(\Delta t)[\pi S^{g,y}(\Delta t)\pi S^{f,x}(\Delta t)]^n u_\eta^0(x,y), \end{aligned}$$

where $u_\eta^0(x,y) = \pi u^0(x,y)$. Observe that both left and right limits exist (in the sense of L^1) and that u_η is right continuous, i.e. $u_\eta \in \mathcal{U}$. As for $u_{\Delta t}$, u_η has the following three properties:

1. $\|u_\eta(\cdot,\cdot,t)\|_\infty \leq \|u_\eta^0\|_\infty$
2. $T.V._{(x,y)}(u_\eta(\cdot,\cdot,t)) \leq T.V._{(x,y)}(u_\eta^0)$
3. $\|u_\eta(\cdot,\cdot,t_2) - u_\eta(\cdot,\cdot,t_1)\|_1 \leq C(t_2 - t_1)$, [6]

[6] In Holden and Risebro [6] it is shown that $\|u_\eta^m - u_\eta^n\|_1 \leq C(\Delta t + \Delta x)(m-n)$. This means that the Lipschitz condition (3) only holds when $\Delta x = \mathcal{O}((\Delta t)^\gamma)$ for any $\gamma \geq 1$.

for a proof of this cf. [6]. Proceeding exactly in the same way as before, having in mind that $u_\eta(n\Delta t-)$, $u_\eta((n+\frac{1}{2})\Delta t-)$ (left limits) are not equal to $u_\eta(n\Delta t+)$, $u_\eta((n+\frac{1}{2})\Delta t+)$ (right limits) respectively, we obtain that

$$-\mathcal{D}_T(\phi, u_\eta, a) \le \mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3 + \mathcal{I}_4$$

where $\mathcal{I}_i = \mathcal{I}_i(\phi, u_\eta, a)$, $i = 1, 2, 3, 4$ is defined by

$$\begin{aligned}
\mathcal{I}_1 &= \sum_{n=0}^{N-1} \int_{\mathbf{R}^2} \int_{n\Delta t}^{(n+1)\Delta t} 2\chi_n F_1(u_\eta, a)\phi_x \, dt\, dx\, dy - \int_{\mathbf{R}^2} \int_0^T F_1(u_\eta, a)\phi_x \, dt\, dx\, dy \\
\mathcal{I}_2 &= \sum_{n=0}^{N-1} \int_{\mathbf{R}^2} \int_{n\Delta t}^{(n+1)\Delta t} 2(1-\chi_n) F_2(u_\eta, a)\phi_y \, dt\, dx\, dy - \int_{\mathbf{R}^2} \int_0^T F_2(u_\eta, a)\phi_y \, dt\, dx\, dy \\
\mathcal{I}_3 &= -\sum_{n=1}^{N-1} \int_{\mathbf{R}^2} \phi(x, y, n\Delta t)(|u_\eta(n\Delta t-) - a| - |u_\eta(n\Delta t+) - a|) \, dx\, dy \\
\mathcal{I}_4 &= -\sum_{n=0}^{N-1} \int_{\mathbf{R}^2} \phi(x, y, (n+\frac{1}{2})\Delta t)(|u_\eta((n+\frac{1}{2})\Delta t-) - a| - |u_\eta((n+\frac{1}{2})\Delta t+) - a|) \, dx\, dy.
\end{aligned}$$

We observe that $\mathcal{I}_1$ and $\mathcal{I}_2$ are identical with the two terms estimated in Lemma 3.4. The latter two terms $\mathcal{I}_3$ and $\mathcal{I}_4$ represent the error introduced by the projection operator, and if left limits are equal to right limits, i.e. the solution is not projected back onto the grid at each time step (we are in a situation covered by Theorem 3.2), then these two terms cancel out (as, of course, they should do). In the remaining analysis we substitute in $u(x', y', t')$ for a and $\Omega_{\varepsilon,\varepsilon_0} = \omega_\varepsilon(x - x')\omega_\varepsilon(y - y')\omega_{\varepsilon_0}(t - t')$ for ϕ. If we define $\mathcal{I}_i^{\varepsilon,\varepsilon_0} = \mathcal{I}_i^{\varepsilon,\varepsilon_0}(\Omega_{\varepsilon,\varepsilon_0}, u_\eta, u)$, $i = 1, 2, 3, 4$ by

$$\mathcal{I}_i^{\varepsilon,\varepsilon_0} = \int_{\mathbf{R}^2} \int_0^T \mathcal{I}_i(\Omega_{\varepsilon,\varepsilon_0}(x - x', y - y', t - t'), u_\eta(x, y, t), u(x', y', t'))\, dt'\, dx'\, dy',$$

then,

$$-\mathcal{D}_T^{\varepsilon,\varepsilon_0}(\Omega_{\varepsilon,\varepsilon_0}, u_\eta, u) \le \mathcal{I}_1^{\varepsilon,\varepsilon_0} + \mathcal{I}_2^{\varepsilon,\varepsilon_0} + \mathcal{I}_3^{\varepsilon,\varepsilon_0} + \mathcal{I}_4^{\varepsilon,\varepsilon_0}.$$

The first two terms have already been proven to have proper upper bounds. So let us estimate $\mathcal{I}_3^{\varepsilon,\varepsilon_0}$ and $\mathcal{I}_4^{\varepsilon,\varepsilon_0}$.[7] We start with $\mathcal{I}_3^{\varepsilon,\varepsilon_0}$ ($\mathcal{I}_4^{\varepsilon,\varepsilon_0}$ will follow in the same way). $\mathcal{I}_3^{\varepsilon,\varepsilon_0}$ takes the form

$$\begin{aligned}
\mathcal{I}_3^{\varepsilon,\varepsilon_0} = \sum_{n=1}^{N-1} \int_0^T \int_{\mathbf{R}^4} &(|u_\eta(x, y, n\Delta t+) - u(x', y', t')| - |u_\eta(x, y, n\Delta t-) - u(x', y', t')|) \times \\
&\Omega_{\varepsilon,\varepsilon_0}(x - x', y - y', n\Delta t - t')\, dx\, dy\, dx'\, dy'\, dt'.
\end{aligned}$$

[7] The method used to estimate $\mathcal{I}_3^{\varepsilon,\varepsilon_0}$ and $\mathcal{I}_4^{\varepsilon,\varepsilon_0}$ is inspired by Lucier's paper [15]. In this paper he uses Kuznetsov's theorem to prove error estimates for the methods of Glimm, Goudunov and LeVeque.

Now recall that

$$u_\eta(x,y,n\Delta t+) = \frac{1}{\Delta x \Delta y}\int_{I_i^x}\int_{I_j^y} u_\eta(X,Y,n\Delta t-)\,dY\,dX$$

for all $(x,y) \in (i\Delta x,(i+1)\Delta x)\times(j\Delta y,(j+1)\Delta y) \equiv I_i^x \times I_j^y$. Substitution of this into $\mathcal{I}_3^{\varepsilon,\varepsilon_0}$, yields ($\int_{I_i^x\times I_j^y}$ is short for $\int_{I_i^x}\int_{I_j^y}$)

$$\begin{aligned}
\mathcal{I}_3^{\varepsilon,\varepsilon_0} &= \sum_{n=1}^{N-1}\int_0^T\int_{\mathbf{R}^2}\sum_{i,j}\int_{I_i^x\times I_j^y}(|\frac{1}{\Delta x\Delta y}\int_{I_i^x\times I_j^y} u_\eta(X,Y,n\Delta t-)\,dY\,dX - u(x',y',t')|-\\
&\quad |u_\eta(x,y,n\Delta t-)-u(x',y',t')|)\Omega_{\varepsilon,\varepsilon_0}(x-x',y-y',n\Delta t-t')\,dy\,dx\,dx'\,dy'\,dt'\\
&= \sum_{n=1}^{N-1}\int_0^T\int_{\mathbf{R}^2}\sum_{i,j}\int_{I_i^x\times I_j^y}(|\frac{1}{\Delta x\Delta y}\int_{I_i^x\times I_j^y}(u_\eta(X,Y,n\Delta t-)-u(x',y',t'))\,dY\,dX| -\\
&\quad \frac{1}{\Delta x\Delta y}\int_{I_i^x\times I_j^y}|u_\eta(x,y,n\Delta t-)-u(x',y',t')|\,dY\,dX)\Omega_{\varepsilon,\varepsilon_0}(x-x',y-y',n\Delta t-t')\times\\
&\quad dy\,dx\,dx'\,dy'\,dt'\\
&\le \sum_{n=1}^{N-1}\int_0^T\int_{\mathbf{R}^2}\Omega_{\varepsilon,\varepsilon_0}(x-x',y-y',n\Delta t-t')\sum_{i,j}\int_{I_i^x\times I_j^y}\int_{I_i^x\times I_j^y}\frac{1}{\Delta x\Delta y}\times\\
&\quad (|u_\eta(X,Y,n\Delta t-)-u(x',y',t')|-|u_\eta(x,y,n\Delta t-)-u(x',y',t')|)\,dY\,dX\,dy\,dx\,dx'\,dy'\,dt'\\
&= \frac{1}{2}\sum_{n=1}^{N-1}\int_0^T\int_{\mathbf{R}^2}\Omega_{\varepsilon,\varepsilon_0}(x-x',y-y',n\Delta t-t')\sum_{i,j}\int_{I_i^x\times I_j^y}\int_{I_i^x\times I_j^y}\frac{1}{\Delta x\Delta y}\times\\
&\quad (|u_\eta(X,Y,n\Delta t-)-u(x',y',t')|-|u_\eta(x,y,n\Delta t-)-u(x',y',t')|)\,dY\,dX\,dy\,dx\,dx'\,dy'\,dt'\\
&\quad +\frac{1}{2}\sum_{n=1}^{N-1}\int_0^T\int_{\mathbf{R}^2}\Omega_{\varepsilon,\varepsilon_0}(X-x',Y-y',n\Delta t-t')\sum_{i,j}\int_{I_i^x\times I_j^y}\int_{I_i^x\times I_j^y}\frac{1}{\Delta x\Delta y}\times\\
&\quad (|u_\eta(x,y,n\Delta t-)-u(x',y',t')|-|u_\eta(X,Y,n\Delta t-)-u(x',y',t')|)\,dy\,dx\,dY\,dX\,dx'\,dy'\,dt'\\
&= \frac{1}{2}\sum_{n=1}^{N-1}\int_0^T\int_{\mathbf{R}^2}(\Omega_{\varepsilon,\varepsilon_0}(x-x',y-y',n\Delta t-t')-\Omega_{\varepsilon,\varepsilon_0}(X-x',Y-y',n\Delta t-t'))\times\\
&\quad \sum_{i,j}\int_{I_i^x\times I_j^y}\int_{I_i^x\times I_j^y}\frac{1}{\Delta x\Delta y}(|u_\eta(X,Y,n\Delta t-)-u(x',y',t')|-|u_\eta(x,y,n\Delta t-)-u(x',y',t')|)\times\\
&\quad dY\,dX\,dy\,dx\,dx'\,dy'\,dt'
\end{aligned}$$

Taking absolute values and using the inverse triangle inequality we achieve

$$|\mathcal{I}_3^{\varepsilon,\varepsilon_0}| \le \frac{1}{2}\sum_{n=1}^{N-1}\int_0^T\int_{\mathbf{R}^2}\sum_{i,j}\int_{I_i^x\times I_j^y}\int_{I_i^x\times I_j^y}\frac{1}{\Delta x\Delta y}|u_\eta(X,Y,n\Delta t-)-u_\eta(x,y,n\Delta t-)|\times$$
$$|\Omega_{\varepsilon,\varepsilon_0}(x-x',y-y',n\Delta t-t')-\Omega_{\varepsilon,\varepsilon_0}(X-x',Y-y',n\Delta t-t')|\,dY\,dX\,dy\,dx\,dx'\,dy'\,dt'$$

If we integrate over x', y', t' and assume $(x, y) \in I_i^x \times I_j^y$, then we find that

$$\begin{aligned}
&\int_0^T \int_{\mathbf{R}^2} |(\Omega_{\varepsilon,\varepsilon_0}(x-x', y-y', n\Delta t - t') - \Omega_{\varepsilon,\varepsilon_0}(X-x', Y-y', n\Delta t - t'))| \, dx' \, dy' \, dt' = \\
&\quad \int_0^T |\omega_{\varepsilon_0}(n\Delta t - t')| \int_{\mathbf{R}^2} |\omega_\varepsilon(x-x')\omega_\varepsilon(y-y') - \omega_\varepsilon(X-x')\omega_\varepsilon(Y-y')| \, dx' \, dy' \, dt' \\
&\quad \le \int_{\mathbf{R}^2} |\omega_\varepsilon(x-x') - \omega_\varepsilon(X-x')||\omega_\varepsilon(y-y')| \, dy' \, dx' \\
&\quad + \int_{\mathbf{R}^2} |\omega_\varepsilon(y-y') - \omega_\varepsilon(Y-y')||\omega_\varepsilon(X-x')| \, dx' \, dy' \\
&\quad \le \int_{\mathbf{R}} |\omega_\varepsilon(x-x') - \omega_\varepsilon(X-x')| \, dx' + \int_{\mathbf{R}} |\omega_\varepsilon(y-y') - \omega_\varepsilon(Y-y')| \, dy \\
&\quad \le (\Delta x + \Delta y)\frac{\|\omega'\|_1}{\varepsilon}.
\end{aligned}$$

Trivially

$$\begin{aligned}
|u_\eta(X, Y, n\Delta t-) - u_\eta(x, y, n\Delta t-)| &= |u_\eta(x, Y, n\Delta t-) - u_\eta(x, y, n\Delta t-)| \\
&\le T.V._{(y \in I_j^y)}(u_\eta(x, \cdot, n\Delta t-)).
\end{aligned}$$

Therefore it follows that

$$\begin{aligned}
|\mathcal{I}_3^{\varepsilon,\varepsilon_0}| &\le \frac{1}{2}\sum_{n=1}^{N-1}\sum_{i,j}\int_{I_i^x \times I_j^y}\int_{I_j^y} \frac{1}{\Delta x \Delta y}(\Delta x + \Delta y)\frac{\|\omega'\|_1}{\varepsilon} \times \\
&\quad \Big(\int_{I_i^x} T.V._{(y \in I_j^y)}(u_\eta(x, \cdot, n\Delta t-)) \, dx\Big) \, dy \, dX \, dY \\
&\le \frac{1}{2}\sum_{n=1}^{N-1}\sum_{i,j}(\Delta x + \Delta y)\Delta y \frac{\|\omega'\|_1}{\varepsilon} T.V._{(x \in I_i^x, y \in I_j^y)}(u_\eta(\cdot, \cdot, n\Delta t-)) \\
&\le \frac{1}{2}\sum_{n=1}^{N-1}(\Delta x + \Delta y)\Delta y \frac{\|\omega'\|_1}{\varepsilon} T.V._{(x,y)}(u_\eta(\cdot, \cdot, n\Delta t-)) \\
&\le \frac{1}{2}(\Delta x + \Delta y)\Delta y \frac{T}{\Delta t}\frac{\|\omega'\|_1}{\varepsilon} T.V._{(x,y)}(u_\eta^0),
\end{aligned}$$

where we in the last passage have used the TVD property of u_η. An identical analysis shows that $|\mathcal{I}_4^{\varepsilon,\varepsilon_0}|$ is bounded by almost the same expression,

$$|\mathcal{I}_4^{\varepsilon,\varepsilon_0}| \le \frac{1}{2}(\Delta x + \Delta y)\Delta x \frac{T}{\Delta t}\frac{\|\omega'\|_1}{\varepsilon} T.V._{(x,y)}(u_\eta^0),$$

which leads to

$$|\mathcal{I}_3^{\varepsilon,\varepsilon_0}| + |\mathcal{I}_4^{\varepsilon,\varepsilon_0}| \le \frac{K_3(\Delta x + \Delta y)^2}{\varepsilon \Delta t}$$

for some constant K_3. We have already shown (Lemma 3.4 [8]) that

$$|\mathcal{I}_1^{\varepsilon,\varepsilon_0}| + |\mathcal{I}_2^{\varepsilon,\varepsilon_0}| \leq \Delta t \left\{ \frac{K_1(\Delta t + 2\varepsilon_0)}{\varepsilon_0^2} + \frac{K_2}{\varepsilon} \right\},$$

and $|\mathcal{D}_T^{\varepsilon,\varepsilon_0}(\Omega_{\varepsilon,\varepsilon_0}, u_\eta, u)|$ is bounded by $\sum_{i=1}^4 |\mathcal{I}_i^{\varepsilon,\varepsilon_0}|$. Using the approximation theorem of Kuznetsov with the same bounds on the moduli of continuity as in the proof of Theorem 3.2, we achieve (after assuming $\Delta x = \Delta y$) that

$$\|S(T)u^0(\cdot,\cdot) - u_\eta(\cdot,\cdot,T-)\|_1 \leq \|u^0 - u_\eta^0\|_1 + K_0\varepsilon_0 + \tilde{K}_0\varepsilon + \frac{K_1\Delta t(\Delta t + 2\varepsilon_0)}{\varepsilon_0^2} + \frac{K_2\Delta t}{\varepsilon} + \frac{4K_3(\Delta x)^2}{\varepsilon\Delta t}.$$

If we let $\varepsilon_0 = (\Delta t)^{1/2}$ and $\varepsilon = (\Delta x)^{1/2}$ (this is just one possible choice, the point is however, that both ε and ε_0 have to be of order $1/2$ either in Δx or Δt to obtain an optimal estimate), then for some constants $\tilde{K}$ and $\tilde{C}$,

$$\begin{aligned} \|S(T)u^0(\cdot,\cdot) - u_\eta(\cdot,\cdot,T-)\|_1 &\leq C_0\Delta x + K_0(\Delta t)^{1/2} + \tilde{K}_0(\Delta x)^{1/2} + K_1\Delta t + 2K_1(\Delta t)^{1/2} \\ &+ K_2(\Delta t)^{1/2}(\frac{\Delta t}{\Delta x})^{1/2} + 4K_3(\Delta x)^{1/2}(\frac{\Delta x}{\Delta t}) \\ &\leq \tilde{K}(\Delta t)^{1/2}\left\{(1 + (\frac{\Delta t}{\Delta x})^{1/2}\right\} + \tilde{C}(\Delta x)^{1/2}\left\{1 + (\frac{\Delta x}{\Delta t})\right\} \end{aligned}$$

as $\Delta x, \Delta t \to 0$. If we assume $\Delta x = \mathcal{O}(\Delta t)$, then for some constant K,

$$\|S(T)u^0(\cdot,\cdot) - u_\eta(\cdot,\cdot,T-)\|_1 \leq K((\Delta t)^{1/2} + (\Delta x)^{1/2}).$$

This concludes the proof of Lemma 3.6, since $u_\eta(\cdot,\cdot,t)$ coincides with $u_\eta^n(\cdot,\cdot)$ when $t = t_n$.

■

As before let $S(t)u^0$ denote the exact solution of (3.3). Furthermore, let $S_\delta(t)u^0$ denote the exact solution of (3.3), but with the flux functions f and g replaced with their piecewise linear and continuous approximations f_δ and g_δ respectively. The previous lemma says that $\|S_\delta(T)u^0(\cdot,\cdot) - u_\eta^N(\cdot,\cdot)\|_1$ is of order $(\Delta t)^{1/2}$. To extend this lemma to the case where the flux functions not necessarily is piecewise linear, we have to bound $\|S(T)u^0(\cdot,\cdot) - S_\delta(T)u^0(\cdot,\cdot)\|_1$. In one space dimension, Lucier [15] has proven that

$$\|u(t) - v(t)\|_1 \leq \|u^0 - v^0\|_1 + t\|f - g\|_{Lip} \min\left\{T.V._x(u^0), T.V._x(v^0)\right\},$$

[8] The proof of Lemma 3.4 depends heavily upon two fundamental properties, namely that $u_{\Delta t}(\cdot,\cdot,t)$ is in $BV(\mathbf{R}^2)$ with BV–norm bounded independently of t and that $u_{\Delta t}$ is Lipschitz continuous (in the sense of L^1). The reason that the proof carries over to the situation where $u_{\Delta t}$ is replaced by u_η is that u_η also has these two properties.

where u and v are exact solutions of the scalar conservation law with flux functions f and g, intial data u^0 and v^0, respectively. We will extend this result to the n–dimensional case. The proof will be a straightforward generalisation of the argument given by Lucier [15].

Proposition 3.1 *Let f_i, g_i, $i = 1, \dots, n$, be Lipschitz continuous functions, $u^0, v^0 \in BV(\mathbf{R}^n)$ and u and v are solutions of*

$$\begin{aligned} u_t + \sum_{i=1}^{n} f_i(u)_{x_i} &= 0 \quad x \in \mathbf{R}^n \\ u(x,0) &= u^0(x) \quad x \in \mathbf{R}^n \end{aligned} \tag{3.30}$$

and

$$\begin{aligned} v_t + \sum_{i=1}^{n} g_i(v)_{x_i} &= 0 \quad x \in \mathbf{R}^n \\ v(x,0) &= v^0(x) \quad x \in \mathbf{R}^n, \end{aligned} \tag{3.31}$$

then for any $T > 0$,

$$\|u(T) - v(T)\|_1 \le \|u^0 - v^0\|_1 + T \max_{1 \le i \le n} \|f_i - g_i\|_{\mathrm{Lip}} \min\left\{ T.V._x(u^0), T.V._x(v^0) \right\}.$$

Proof With $f = (f_1, \dots, f_n)$ and $g = (g_1, \dots, g_n)$ let $\mathcal{D}_{T,f}$ and $\mathcal{D}_{T,g}$ be the n–dimensional Kružkov form of (3.30) and (3.31) respectively, i.e.,

$$\begin{aligned} \mathcal{D}_{T,f}(\phi, u, a) = \int_{\mathbf{R}^n} \int_0^T [\phi_t |u - a| + \operatorname{sign}(u - a) \sum_{i=0}^{n} (f_i(u) - f_i(a)) \phi_{x_i}] \, dt \, d^n x \\ + \int_{\mathbf{R}^n} |u^0 - a| \, d^n x - \int_{\mathbf{R}^n} |u(T) - a| \, d^n x. \end{aligned}$$

$\mathcal{D}_{T,g}$ is given by the same expression, but with f_i replaced by g_i. With $\Omega_{\varepsilon,\varepsilon_0} = \Omega_{\varepsilon,\varepsilon_0}(x - x', t - t') = \omega_\varepsilon(x_1 - x_1') \cdots \omega_\varepsilon(x_n - x_n') \omega_{\varepsilon_0}(t - t')$, we define $\mathcal{D}^{\varepsilon,\varepsilon_0}_{T,f}$ by

$$\mathcal{D}^{\varepsilon,\varepsilon_0}_{T,f}(\Omega_{\varepsilon,\varepsilon_0}, v, u) = \int_{\mathbf{R}^n} \int_{\mathbf{R}^n} \int_0^T \int_0^T \mathcal{D}_{T,f}(\Omega_{\varepsilon,\varepsilon_0}(x - x', t - t'), v(x,t), u(x',t')) \, dt \, dt' \, d^n x \, d^n x',$$

and similarly for $\mathcal{D}^{\varepsilon,\varepsilon_0}_{T,g}$. Kuznetsov's theorem, which in case when u and v are exact solutions and therefore possess a linear moduli of continuity, states that

$$\|u(\cdot, t) - v(\cdot, t)\|_1 \le \|u^0 - v^0\|_1 + C_1 \varepsilon + C_2 \varepsilon_0 - \mathcal{D}^{\varepsilon,\varepsilon_0}_{T,f}(\Omega_{\varepsilon,\varepsilon_0}, v, u). \tag{3.32}$$

Since u and v are entropy weak solutions, consequently $\mathcal{D}_{T,f}(\phi, u, a) \ge 0$ and $\mathcal{D}_{T,g}(\phi, v, a) \ge 0$ for all constants a and for all $\phi \in C_0^1(\mathbf{R}^2 \times \mathbf{R}^+)$, $\phi \ge 0$, we have that

$$\begin{aligned} \mathcal{D}_{T,f}(\phi, v, a) &\ge \mathcal{D}_{T,f}(\phi, v, a) - \mathcal{D}_{T,g}(\phi, v, a) \\ &= \int_{\mathbf{R}^n} \int_0^T \operatorname{sign}(v - a) \sum_{i=0}^{n} (f_i(v) - f_i(a) - (g_i(v) - g_i(a))) \phi_{x_i}) \, dt \, d^n x. \end{aligned}$$

Let

$$\Delta_{f_i,g_i}(v,a) = \mathrm{sign}(v-a)(h_i(v)-h_i(a)).$$

where $h_i = f_i - g_i$. Then after setting $a = u(x',t')$, $\phi = \Omega_{\varepsilon,\varepsilon_0}(x-x',t-t')$ and integrating over x' and t', we obtain

$$\begin{aligned}
&|\mathcal{D}^{\varepsilon,\varepsilon_0}_{T,f}(\Omega_{\varepsilon,\varepsilon_0},v,u(x',t')| \\
&\quad\le |\int_{\mathbf{R}^n}\int_{\mathbf{R}^n}\int_0^T\int_0^T \sum_{i=0}^n \Delta_{f_i,g_i}((v(x,t),u(x',t'))\Omega_{\varepsilon,\varepsilon_0}(x-x',t-t')_{x_i}\,dt\,dt'\,d^nx\,d^nx'| \\
&\quad= |\int_{\mathbf{R}^n}\int_{\mathbf{R}^n}\int_0^T\int_0^T \sum_{i=0}^n \frac{\partial}{\partial x_i}[\Delta_{f_i,g_i}(v(x,t),u(x',t'))]\Omega_{\varepsilon,\varepsilon_0}(x-x',t-t')\,dt\,dt'\,d^nx\,d^nx'|
\end{aligned}$$

where we have used integration by parts once. Our next step is to bound

$$|\frac{\partial}{\partial x_i}[\Delta_{f_i,g_i}(v(x,t),u(x',t'))]|$$

as a measure. Denote by e_i the i'th unit vector in $\mathbf{R}^n$, and let $h \in \mathbf{R}$ be a given positive number. Because $v(\cdot,t) \in BV(\mathbf{R}^n)$ for each t, we have that $v(x_1,\dots,x_{i-1},\cdot,x_{i+1},\dots,x_n,t) \in BV(\mathbf{R})$ for each t and almost all $x_j \in \mathbf{R}$, $j = 1,\dots,n$, $j \ne i$ (we consider $v(x,t) = v(x_1,\dots,x_i,\dots,x_n,t)$ as a function of x_i, $t, x_1,\dots,x_{i-1},x_{i+1},\dots,x_n$ are to be considered as fixed parameters.) Therefore, for each t and almost all $x_j \in \mathbf{R}$, $j = 1,\dots,n$, $j \ne i$, we can find an increasing function of x_i, $v(x_1,\dots,x_{i-1},\cdot,x_{i+1},\dots,x_n,t)^+$, and a decreasing function of x_i, $v(x_1,\dots,x_{i-1},\cdot,x_{i+1},\dots,x_n,t)^-$, so that $v(x_1,\dots,x_{i-1},\cdot,x_{i+1},\dots,x_n,t) = v(x_1,\dots,x_{i-1},\cdot,x_{i+1},\dots,x_n,t)^+ + v(x_1,\dots,x_{i-1},\cdot,x_{i+1},\dots,x_n,t)^-$. Define
$v(x_1,\dots,x_{i-1},\cdot,x_{i+1},\dots,x_n,t)^t = v(x_1,\dots,x_{i-1},\cdot,x_{i+1},\dots,x_n,t)^+$
$- v(x_1,\dots,x_{i-1},\cdot,x_{i+1},\dots,x_n,t)^-$. If we use the fact that $\mathrm{sign}(a-b)(h(a)-h(b)) = h(a\vee b)-h(a\wedge b)$, where $a\vee b = \max(a,b)$ and $a\wedge b = \min(a,b)$, then a case analysis shows that (this argument is identical with the one given by Lucier [15])

$$\begin{aligned}
|\Delta_{f_i,g_i}(v(x+he_i,t),u(x',t')) - \Delta_{f_i,g_i}(v(x,t),u(x',t'))| \;&= \\
|h_i(v(x+he_i,t)\vee u(x',t')) - h_i(v(x+he_i,t)\wedge u(x',t')) & \\
-(h_i(v(x,t)\vee u(x',t')) - h_i(v(x,t)\wedge u(x',t')))| & \\
\le |h_i(v(x+he_i,t)\vee u(x',t')) - h_i(v(x,t)\vee u(x',t'))| & \\
+|h_i(v(x+he_i,t)\wedge u(x',t')) - h_i(v(x,t)\wedge u(x',t'))| & \\
\le \|h_i\|_{\mathrm{Lip}}|v(x+he_i,t)^t - v(x,t)^t|. &
\end{aligned}$$

$|\frac{\partial}{\partial x_i}[\Delta_{f_i,g_i}(v(x,t),u(x',t'))]|$ is therefore bounded by $\|h_i\|_{\mathrm{Lip}}|(\partial v/\partial x_i)(x,t)|$ as a measure for all t and almost all $x_j \in \mathbf{R}$, $j = 1,\dots,n$, $j \ne i$. This implies that $\mathcal{D}^{\varepsilon,\varepsilon_0}_{T,f}(\Omega_{\varepsilon,\varepsilon_0},v,u)$ may be bounded by the integral of the convolution of the measures $\|h_i\|_{\mathrm{Lip}}|\nu_{x_i}|$ and $\Omega_{\varepsilon,\varepsilon_0}$,

i.e. ($M = \max_{1\le i\le n} \|h_i\|_{\mathrm{Lip}}$),

$$\begin{aligned}
\mathcal{D}^{\varepsilon,\varepsilon_0}_{T,f}(\Omega_{\varepsilon,\varepsilon_0}, v, u) &\le M\int_0^T\int_0^T \sum_{i=0}^{n} \Big|\int_{\mathbf{R}^n}\int_{\mathbf{R}^n} \frac{\partial v}{\partial x_i}(x,t)\Omega_{\varepsilon,\varepsilon_0}(x-x', t-t')\,d^n x\,d^n x'\Big|\,dt\,dt' \\
&= M\int_0^T\int_0^T \sum_{i=0}^{n} \Big|\int_{\mathbf{R}^n}\int_{\mathbf{R}^n} v(x,t)\Omega_{\varepsilon,\varepsilon_0}(x-x', t-t')_{x_i}\,d^n x\,d^n x'\Big|\,dt\,dt' \\
&\le M\int_0^T\int_0^T \sum_{i=0}^{n} \int_{\mathbf{R}^{n-1}} \mathrm{T.V.}_{x_i}(v(x_1,\dots,x_{i-1},\cdot,x_{i+1},\dots,x_n,t))\omega_{\varepsilon_0}(t-t')\,d^{n-1}x\,dt\,dt' \\
&\le M\int_0^T \sum_{i=0}^{n} \int_{\mathbf{R}^{n-1}} \mathrm{T.V.}_{x_i}(v(x_1,\dots,x_{i-1},\cdot,x_{i+1},\dots,x_n,t))\,d^{n-1}x\,dt \\
&\le M\mathrm{T.V.}_x(v^0)T.
\end{aligned}$$

The argument needed to show that

$$\Big|\int_{\mathbf{R}^n}\int_{\mathbf{R}^n} v(x,t)\Omega_{\varepsilon,\varepsilon_0}(x-x', t-t')_{x_i}\,d^n x\,d^n x'\Big| \le$$
$$\int_{\mathbf{R}^{n-1}} \mathrm{T.V.}_{x_i}(v(x_1,\dots,x_{i-1},\cdot,x_{i+1},\dots,x_n,t))\omega_{\varepsilon_0}(t-t')\,d^{n-1}x$$

is similar to the one given in the proof of Lemma 3.4 (cf. the calculations leading to eq. 3.29). Letting $\varepsilon, \varepsilon_0$ tend to zero in (3.32), and using symmetry, we achieve the desired result. ∎

We are now in the position of proving the main result which says that our numerical method (for arbitrary flux functions and initial data) produces approximate solutions which converges to the entropy weak solution with a rate of convergence which is no less than 1/2. More precisely:

Theorem 3.4 (Convergence) *If $u^0 \in L^1(\mathbf{R}^2)\cap L^\infty(\mathbf{R}^2)\cap BV(\mathbf{R}^2)$, f and g are piecewise C^2 on $I = [-\|u^0\|_\infty, \|u^0\|_\infty]$, $N\Delta t = T < T_0$ for some fixed T_0, $\Delta x = \Delta y = \mathcal{O}(\Delta t)$, and δ is the parameter measuring the polygonal approximation of the flux functions, then as $\eta = (\Delta x, \Delta y, \Delta t) \to 0$ and $N \to \infty$,*

$$\|S(T)u^0(\cdot,\cdot) - u^N_\eta(\cdot,\cdot)\|_1 \quad \le C((\Delta t)^{1/2} + (\Delta x)^{1/2} + \delta), \tag{3.33}$$

where C is a finite constant independent of η, but dependent on T_0, $\mathrm{T.V.}_{(x,y)}(u^0)$, $\|f'\|_\infty$ and $\|g'\|_\infty$.

Proof If f_δ is piecewise linear interpolant of f (f Lipschitz continuous and piecewise C^2) with breakpoints at $i\delta$, then

$$\|f - f_\delta\|_{\mathrm{Lip}} \le \frac{\delta}{2}\|f''\|_\infty.$$

Therefore

$$\|S(T)u^0(\cdot,\cdot) - S_\delta(T)u_\eta^0(\cdot,\cdot)\|_1 \leq T\delta \min\left\{T.V._{(x,y)}(u^0), T.V._{(x,y)}(v^0)\right\} \leq C_1\delta$$

for some constant C_1, and by using Lemma 3.6 the desired result follows:

$$\begin{aligned}\|S(T)u^0(\cdot,\cdot) - u_\eta^N(\cdot,\cdot)\|_1 &\leq \|S(T)u^0(\cdot,\cdot) - S_\delta(T)u^0(\cdot,\cdot)\|_1 + \|S_\delta(T)u^0(\cdot,\cdot) - u_\eta^N(\cdot,\cdot)\|_1 \\ &\leq C_1\delta + K((\Delta t)^{1/2} + (\Delta x)^{1/2}) \\ &\leq C((\Delta t)^{1/2} + (\Delta x)^{1/2} + \delta)\end{aligned}$$

for some constant C, as $\eta \to 0$. ■

A few remarks are in order. First, the error estimate (3.33) is not a suprising result seen in the light of the argument given at the beginning of this chapter, where we compared our method with Goudunov's scheme. Assuming that the discretization parameters obey the strict CFL condition, then the method of dimensional splitting together with with front tracking is the same as using the Godunov scheme and dimensional splitting, and our method can therefore not have a convergence rate which is better than $1/2$. Secondly, it is obvious that there are two sources of error in the dimensional splitting - front tracking method. The intrinsic error involved in using the dimensional splitting technique and the spatial discretization error involved in using the front tracking scheme and the projection operator. In the general case these two sources of error interact in a complex fashion. It is therefore interesting to see how they are represented in (3.33). The $(\Delta t)^{1/2}$-term represents the first source of error, the $((\Delta x)^{1/2} + \delta)$-term the second. δ is the error due to the fact that we replace the flux functions with some approximate functions, and $(\Delta x)^{1/2}$ represents the (global) error caused by the projection operator.

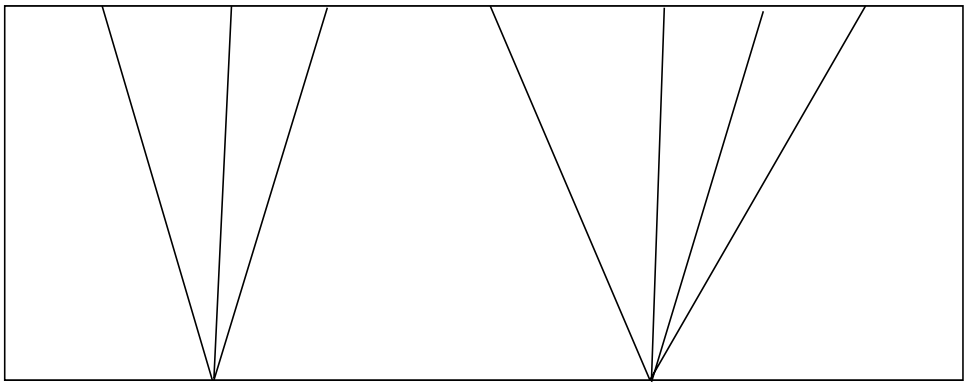

Figure 3.1: When the initial data is piecewise constant, we can locally in time determine the solution of the initial value problem by solving Riemann problems.

Chapter 4

Numerical experiments

Kunsten å gjøre hva man skal, består i å
ville hva man skal og gjøre hva man vil.

– Piet Hein –

In this chapter we present some examples where our numerical method from the previous chapter has been used. The purpose of the numerical experiments is to explore some of the properties of the method, such as convergence rates and computional cost. In the previous chapter a rigorous argument showed that the rate of convergence is no less than $1/2$. This result has to be considered as a "worst case estimate". Meaning that the method may in many cases have a convergence rate which is noticeably better, but never worse. One of our main objectives is therefore to try to extract, from numerical experiments, some information conserning the optimality of this estimate of the convergence rate. Another property we would like to investigate is the CPU-time needed to obtain a certain degree of accuracy. The rate of convergence should always be compared with computional cost.

We will measure the error (denoted by $\mathcal{E}$), at a fixed time T, in a relative L^1-norm, i.e.,

$$\mathcal{E} = \frac{||u_\eta^n - u^*||_1}{||u^*||_1}$$

where $n\Delta t = T$, u_η^n is the solution generated by our numerical method and u^* is some reference solution. To estimate the rate of convergence, we assume that the error on a grid with gridsize Δx (we always use $\delta = C\Delta x$ for some fixed constant C, and $\Delta y = \Delta x$) is of the form

$$\mathcal{E} = \alpha_1(\Delta x)^{\beta_1},$$

where α_1 is a constant independent of Δx, β_1 the rate of convergence. In each of our expamples we will compute approximate solutions for $\Delta x = 2^{-k}$, $k = 2, \ldots, 8$. Furthermore, two relations between Δt and Δx will be considered:

Case 1. $\frac{\Delta t}{\Delta t}\max(\max_u |f'(u)|, \max_u |g'(u)|) \leq \frac{1}{2}$ (strict CFL condition) This corresponds to the relation used in the convergence theorem from the previous chapter. In this case no shock collisions are allowed take place.

Case 2. $\Delta t = 0.95(\Delta x)^{1/2}$. Here we take advantage of the fact that no CFL condition is needed to make the method stable. The effect of using this relation is that the time step will be larger than the space step, and in this way we allow shock collisions to occur.

The parameters α_1 and β_1 are determined using a standard regression analysis. [1] In the examples where the exact solution is not explicitly known, we will use our method to calculate the reference solution using a very fine grid: $\delta = \Delta x = \Delta y = 2^{-9}$, $\Delta t = 0.95(\Delta x)^{1/2}$. [2] The reason for using $\Delta t = 0.95(\Delta x)^{1/2}$ to calculate the reference solution is that numerical experiments indicate that the method is considerably improved, in every respect, using this relation between the space step and the time step (the essential fact here is that the time step should be larger than the space step). The effect of using a very fine grid solution as the reference solution is that convergence rate will be to large when the coarser grid approaches the grid used to calculate the reference solution. Similarly, if the grid become to coarse we may not pick up asymptotic effects. We will therefore use a regression analysis on the observation for $k = 4, 6, 7$, to compute α_1 and the rate of convergence β_1 (in the case where the exact solution is known we use a regression analysis on the observations for $k = 4, 5, 6, 7, 8$). Similarly to the error $\mathcal{E}$, we assume that the CPU-time (measured in seconds) used by the method on a particular problem (denoted by $\mathcal{T}$) is of the form:

$$\mathcal{T} = \alpha_2 \mathcal{E}^{\beta_2}$$

for some constants α_2 and β_2. The paramerters α_2 and β_2 are determined in the same way as α_1 and β_1.

4.1 Implementation

The method of dimensional splitting using the front tracking scheme as approximate 1D solution operators (Algorithm 3.1) was implemented in the programming language C on a SILICON GRAPHICS (INDY) work-station, MIPS R4000/100 MHZ/32 MB RAM.

[1] Define (x_Δ, y_Δ) and $\tilde{\alpha}$ by $x_\Delta = \ln(\Delta)$, $y_\Delta = \ln(\mathcal{E})$ and $\tilde{\alpha}_1 = \ln(\alpha_1)$. α_1 and β_1 is determined (using the method of least squares) such that the straight line $y = \tilde{\alpha}_1 + \beta_1 x$ fit the data set $\{(x_\Delta, y_\Delta)\}$ in a best possible way. For a brief introduction to regression analysis cf. Larsen and Marx [12]

[2] Preferably we would like to use $\delta = \Delta x = \Delta y = 2^{-10}$, $\Delta t = 0.95(\Delta x)^{1/2}$ (or even smaller grid parameters) to calculate the reference solution, but due to hardware limitations we had to settle with 2^{-9}.

The program has two parts; an initialization and an iteration part.

Initialization: The initialization part consists in constructing piecewise constant initial data and piecewise linear flux functions. The construction of a piecewise constant approximation to $u^0(x,y)$ can be done in several ways. In the present implementation we have simply divided the rectangle $[a,b]\times[c,d]$ into N_xN_y equally sized rectangles, and let the constant value on each such rectangle be the initial function evaluated at the center of the rectangle. The piecewise linear flux functions, f_δ and g_δ, are constructed as explained in Chapter 3. To represent f_δ and g_δ on a computer we have introduced a so called point object consisting of a u-value and a f-value. A piecewise linear flux function is then stored as a list of point objects, where each object represents a breakepoint. Using this (data) structure it is easy to generalize to the case where the flux functions is linear on intervals of different sizes.

Iteration: The iteration part consists in using the front tracking scheme, first, in the x-direction and then, after projecting the solution back onto the grid, in the y-direction. Furthermore, the solution obtained using front tracking in the y-direction is projected back onto the grid. This process is then repeated a finite number of times. A few remarks on the implementation of the front tracking scheme (Algorithm 2.1) may be instructive. Since the initial data is piecewise constant, the front tracking method is to solve a series of Riemann problems and "to track" the shock collisions. To solve a Riemann problem with left state u_L and right state u_R ($u_L < u_R$) we have to find the convex envelope of the piecewise linear flux function f_δ (or g_δ) restricted to the interval $[u_L,u_R]$. If both u_L and u_R coincide with some u-value in the breakpoint list, then traverse the list (starting with the $[(u_L,f_\delta(u_L))-(u_R,f_\delta(u_R))]$–front) to determine the convex envelope. If the flux function is linear on the intervals $[u_i,u_{i+1}]$, $i=0,\dots,N-1$ and $u_L\in[u_k,u_{k+1}]$, $u_R\in[u_l,u_{l+1}]$ for some $k,l\in\{0,\dots,N-1\}$ (and neither u_L nor u_R coincide with a u-value in the breakpoint list), then let $(u_L,f_\delta(u_L))$ take $(u_k,f_\delta(u_k))$'s place in the list and let $(u_R,f_\delta(u_R))$ take the place of $(u_{l+1},f_\delta(u_{l+1}))$. We then traverse the list (starting with the k'th object) to determine the convex envelope as before. After the envelope is found restore the old points in the breakpoint list.

The solutions of the Riemann problems defined by the initial data are organized in the X-list. To organize the corresponding T-list (according to increasing values of collision times) we have simply used a standard insertion sort, but of course to optimize the program with respect to consumed CPU-time one should know the size of the X-list and then switch back and forth between using an insertion sort and a faster sorting algorithm (Quick sort ?) depending on the number of elements in the X-list. To keep the T-list small we only store the fronts which have collision time less than $t=t_{max}$ ($t_{max}=\Delta t$ in the dim. splitt. alg.). In the iteration part of Algorithm 2.1 we often have to locate a specific front object in the T-list. As for the sorting routine, we have used a plain sequential search algorithm, but again a more sophisticated search technique could have

been used. Since the T-list is sorted according to increasing values of collision times, one possibility is to use a binary search.

Next, we will consider three examples where our method has been used. The first one is a linear problem. This example is used to convince ourselves and the reader that the method is first order accurate on smooth solutions The second example is also linear, but with discontinuous initial data. It is known that Godunov's method behave badly (slow convergence rate, etc.) on this problem in one dimension, cf. [15]. This example is therefore used to obtain a worst possible convergence rate for our method. The last examples is a true non-linear problem where shock collisions take place (in Case 2), and perhaps the most interesting example with respect to practical use of the method.

4.2 Example 1

We will first consider the scalar avection equation

$$u_t + u_x + u_y = 0,$$

with initial data

$$u(x,y,0) = u^0(x,y) = \begin{cases} \sin(\pi x)\sin(\pi y) & 0 < x < 1,\ 0 < y < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

The exact solution is simply a translation of the initial data, i.e.,

$$S(t)u^0(x,y) = u^0(x-t, y-t).$$

The computional domain is taken to be $[-0.5, 2] \times [-0.5, 2]$ and we compute approximate solutions to $S(T)u^0$ for $T = 0.5$. As mentioned before we will study the behaviour of the method using two different relations between the space step and the time step. We first let Δx and Δt obey the strict CFL condition; $\Delta t = 0.5\Delta x$ (Case 1), i.e., our method is essentially the same as Godunov's method. Since the true solution in this example is smooth, we expect that the convergence rate is fairly close to 1. The numerical results are presented in Table **??**. There we list the gridsize Δx, the error $\mathcal{E}$, the CPU-time $\mathcal{T}$, the total number of solved Riemann problems and the total number shock collisions that occurred during the simulation. The experiments suggest a convergence rate which is 0.92, and a CPU-time which increases, as the error tends to zero, like $\mathcal{O}(\mathcal{E}^{-3.3})$.

Next, we let $\Delta t = 0.95(\Delta x)^{1/2}$ (Case 2), i.e., we use a time step that is larger than the spacestep. The results are presented in Table **??**. We observe that the convergence rate is slightly improved; $\beta_1 = 1.04$, but more importantly the CPU-time has decreased dramatically. The CPU-time now increases, as the error tends to zero, like $\mathcal{O}(\mathcal{E}^{-2.3})$. To

obtain an error of e.g., 0.02 we need a CPU-time of 4200 seconds in Case 1, while in Case 2 we need 4 seconds. The results from Case 1 and 2 are also summarized in the table below.

Case	$\mathcal{E} = \alpha_1(\Delta x)^{\beta_1}$ (Error)	$\mathcal{T} = \alpha_2\mathcal{E}^{\beta_2}$ (CPU-time)
1	$\alpha_1 = 3.13,\ \beta_1 = 0.92$	$\alpha_2 = 0.01,\ \beta_2 = -3.3$
2	$\alpha_1 = 1.63,\ \beta_1 = 1.04$	$\alpha_2 = 0.0007,\ \beta_2 = -2.29$

The CPU-time used by the method may also be expressed in terms of the gridsize Δx. In Case 1 we get $\mathcal{T} = \mathcal{O}((\Delta x)^{-3.04})$ and in Case 2, $\mathcal{T} = \mathcal{O}((\Delta x)^{-2.38})$. In the concluding remarks we will argue that the CPU-time should be of order $(\Delta x)^{-\gamma}$, where γ in Case 1 is close to 3 and in Case 2 close to 2.5.

4.3 Example 2

Again we consider the scalar advection equation

$$u_t + u_x + u_y = 0,$$

but this time with initial data

$$u(x,y,0) = u^0(x,y) = \begin{cases} 1 & x \le 0 \\ 0 & x > 0. \end{cases}$$

The exact solution is given by

$$S(t)u^0(x,y) = \begin{cases} 1 & x \le t \\ 0 & x > t. \end{cases}$$

The computional domain is taken to be $[-0.5, 1.5] \times [-0.5, 1.5]$ and $T = 0.5$. This is essentially a one-dimensional Riemann problem. It is known that the (1D) Godunov scheme has a convergence rate of $1/2$ on this problem. The results of the simulations are presented in Table **??** and **??** and summarized in the table below.

Case	$\mathcal{E} = \alpha_1(\Delta x)^{\beta_1}$ (Error)	$\mathcal{T} = \alpha_2\mathcal{E}^{\beta_2}$ (CPU-time)
1	$\alpha_1 = 0.4,\ \beta_1 = 0.50$	$\alpha_2 = 0.00000037,\ \beta_2 = -5.9$
2	$\alpha_1 = 0.4,\ \beta_1 = 0.83$	$\alpha_2 = 0.000002,\ \beta_2 = -2.7$

As observed in Example 1, the overall performance of the method is dramatically improved in Case 2 compared to Case 1. Especially, the CPU-time needed to obtain certain degree of accuracy has decreased by several factors in the exponent. To obtain an error of e.g., 0.04 we need a CPU-time of 136 seconds in Case 1, while in Case 2 wee need

0.04 seconds. This time the increase in the convergence rate is also noticeable. Furthermore, one should also observe that we here have an example where our method has an convergence rate of $1/2$. This means that the error estimate proved in Chapter 3 is in some sense optimal. We can as in Example 1 also express the CPU-time as a function of the gridsize Δx instead of the error $\mathcal{E}$. In Case 1 we get $\mathcal{T} = \mathcal{O}((\Delta x)^{-2.95})$ and in Case 2, $\mathcal{T} = \mathcal{O}((\Delta x)^{-2.24})$. This is in fairly good agreement with the predictions made in Example 1.

A nice feature with the front tracking method is that it does not smear out discontinuous parts of the solution (shocks), which is a major problem for many finite difference methods. Since the projection operator π is simply the grid cell average, our method will also have this problem This is clearly illustrated in Figure **??**. Of course this "smearing effect" can be, more or less, avoided by using a more sophisticated projection operator which "tries to discover" where the solution is smooth and where it is discontinuous. A question that immediately comes to mind is wether the projection operator (grid cell average) is the only reason for our dimensional splitting method having a convergence rate of only $1/2$. The answer is probably that it is a combination of the projection operator and the splitting technique not being good enough. The reason for not being completely sure of this, is that we do not know if the proven error estimate (Theorem 3.2) for the semi-discrete method is optimal.

4.4 Example 3

We now present an example where we generate approximate solutions to the equation

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

with initial data

$$u(x, y, 0) = u^0 = \begin{cases} 0 & x^2 + y^2 < 0.23 \\ 1 & \text{elsewhere,} \end{cases}$$

and

$$\begin{aligned} g(u) &= \frac{u^2}{u^2 + (1-u)^2} \\ f(u) &= g(u)(1 - (1-u)^2). \end{aligned}$$

The flux functions f and g both have an S-shaped form with $f(0) = g(0) = 0$ and $f(1) = g(1) = 1$. This problem is motivated from two-phase flow in porous medium with gravitation in the y-direction. This time the exact solution is not known, so we use our numerical method to compute the reference solution using a very fine grid; $\delta =$

$\Delta y = \Delta x = 2^{-9}$, $\Delta t = 0.95(\Delta x)^{1/2}$. The computional domain is the square $[-1.5, 1.5] \times [-1.5, 1.5]$ and we generate approximate solutions at time $T = 0.5$. Since most practical problems are indeed non-linear, it is interesting to see how our method performs on a non-linear example. It is worth noting that the solutions of the 1D equations consist of both shock and rarefaction waves. The results are presented in Table **??** and **??**, and summarized in the table below.

Case	$\mathcal{E} = \alpha_1(\Delta x)^{\beta_1}$ (Error)	$\mathcal{T} = \alpha_2 \mathcal{E}^{\beta_2}$ (CPU-time)
1	$\alpha_1 = 1.27$, $\beta_1 = 0.83$	$\alpha_2 = 0.0008$, $\beta_2 = -3.46$
2	$\alpha_1 = 1.75$, $\beta_1 = 0.87$	$\alpha_2 = 0.0005$, $\beta_2 = -2.78$

Since the problem is non-linear shock collisions take place in Case 2. This is illustrated in Table **??**, where we see that the number of shock collisions increases monotonically with decreasing Δx. The convergence rates in Case 1 and Case 2 are roughly the same, but again we see an impressive decrease in consumed CPU-time in Case 2 compared with Case 1. To achieve an accuracy of 0.02 we need a CPU-time of 353 seconds in Case 1, while in Case 2 we only need 14 seconds. If we again express the CPU-time as a function of Δx we obtain in Case 1, $\mathcal{T} = \mathcal{O}((\Delta x)^{-2.87})$ and in Case 2, $\mathcal{T} = \mathcal{O}((\Delta x)^{-2.42})$. Again close to the predictions made in Example 1.

4.5 Concluding remarks

The numerical examples indicates that the convergence rate always lies in the interval $[1/2, 1]$. The estimate of the convergence rate stated in Chapter 3 (Theorem 3.4) can therefore be considered as a reasonable result for the general problem. In particular we found an example where the convergence rate turned out to be $1/2$. The proven estimate of the convergence rate is therefore [3] the best estimate we can hope for having in mind that this estimate should be independent of the flux functions, the initial data and the relationship between the discretization paramteters. Furthermore, it seems that the convergence rate is improved when we allow the time step to be larger than the space step, but not so much that we achieve, in the general case, a convergence rate of 1. In Case 2 the convergence rate is, in each of the three examples, noticeable better than $1/2$.

Next, the examples indicate that the CPU-time as a function of the gridsize Δx is of order $\mathcal{O}((\Delta x)^{-\gamma})$ where γ is between 2.5 and 3 in Case 1, and between 2 and 2.5 in Case 2. That this is a fairly reasonable results can be argued for by inspecting Algorithm 3.1. If we assume that the CPU-time used by the front tracking method on a particular 1D problem is roughly $\mathcal{O}((\Delta x)^{-1})$ (lines 4.1.1 and 4.3.1 in Algorithm 3.1), then the CPU-time used

[3] Confer also the argument given on Page 47 where similarities between Godunov's scheme and the dimensional splitting – front tracking method are discussed.

to advance the solution one time step Δt is $\mathcal{O}((\Delta x)^{-2})$ (lines 4.1 and 4.3 in Algorithm 3.1). Therefore the total amount of CPU-time used to calculate an approximate solution to the 2D problem is $\mathcal{O}((\Delta x)^{-2}(\Delta t)^{-1})$ (line 4 in Algorithm 3.1). In Case 1 this means that the CPU-time should be of order $(\Delta x)^{-3.0}$ and in Case 2 of order $(\Delta x)^{-2.5}$.

Using a time step that is larger than the space step makes our method very fast. It uses very little CPU-time to obtain a certain degree of accuracy compared with Godunov's method (the case where the space step and the time step satisfy the strict CFL condition). This together with the fact that the convergence rate in Case 2 seems to be noticeable better than 1/2 indicate that we here have a method which works exellently on most problems. To further improve the method one could use a more sophisticated projection operator to avoid excessive smearing of discontinuities. This has in fact been done with great sucess in a black oil resovoir simulator where dimensional splitting together with front tracking has been used to solve the saturation equation, cf. [1] for details. It is also worth mentioning that it is trivial to extend our method to higher dimensions. In the table below we have summarized some of the properties of our two methods; front tracking (1D) and dimensional splitting (together with front tracking) (2D):

Property	Front tracking (1D)	Dimensional splitting (2D)
Smearing of discontinuities	no	yes (can be improved)
CFL condition	no	no
Convergence rate	1	1/2 (in most cases better)

Finally, what does the future hold for this line of field ? There are two things that comes to mind. First of all, the numerical method should be extended to systems of conservation laws. To begin with from a numerical point of view, anyway. For systems of conservation laws in several space dimensions very little (nothing ?) is known what mathematical theory concerns. For instance it does not exist something similar to the Kružkov/Kuznetsov–theory, which is based on the vanishing viscosity method, for systems of conservation laws (not even for systems in one space dimension). We recall that the proofs of the error estimates stated in this paper were entirely built on this theory. So it seems difficult, at the time of this writing, to derive rigorously estimates like those obtained for scalar conservation laws for systems of conservation laws. But, of course, this can be done numerically. Likewise can also other properties, either connected to the numerical method or to the system of conservation laws, be studied.

Secondly, an extensive comparison-study should be carried out where we compare, in every respect, our numerical method with finite difference methods. This should be done for both scalar problems and for systems.

Δx	$\mathcal{E}$ (Error)	$\mathcal{T}$ (CPU-time)	# Riemann problems	# shock collisions
2^{-2}	0.73770	0.02	322	0
2^{-3}	0.42822	0.14	2484	0
2^{-4}	0.24246	0.98	19496	0
2^{-5}	0.13241	7.59	154448	0
2^{-6}	0.07107	63.12	1229472	0
2^{-7}	0.03710	509.22	9811264	0
2^{-8}	0.01919	4207.97	78391936	0

Table 4.1: Example 1, ($\Delta t = 0.5\Delta x$, $\delta = 4\Delta x$). Based on a standard regression analysis we compute: $\mathcal{E} = 3.13(\Delta x)^{0.92}$ with an estimated error of 0.001. and $\mathcal{T} = 0.01\mathcal{E}^{-3.3}$ with an estimated error of 0.01.

Δx	$\mathcal{E}$ (Error)	$\mathcal{T}$ (CPU-time)	# Riemann problems	# shock collisions
2^{-2}	0.37902	0.01	111	0
2^{-3}	0.20442	0.03	343	0
2^{-4}	0.09385	0.18	1894	0
2^{-5}	0.04336	0.66	6838	0
2^{-6}	0.02209	4.41	44575	0
2^{-7}	0.01047	20.65	206753	0
2^{-8}	0.00525	121.27	1223949	0

Table 4.2: Example 1, ($\Delta t = 0.95(\Delta x)^{1/2}$, $\delta = 4\Delta x$). Based on a standard regression analysis we compute: $\mathcal{E} = 1.63(\Delta x)^{1.04}$ with an estimated error of 0.002 and $\mathcal{T} = 0.0007\mathcal{E}^{-2.29}$ with an estimated error of 0.14.

Δx	$\mathcal{E}$ (Error)	$\mathcal{T}$ (CPU-time)	# Riemann problems	# shock collisions
2^{-2}	0.18747	0.02	80	0
2^{-3}	0.13760	0.05	576	0
2^{-4}	0.09818	0.3	4352	0
2^{-5}	0.07106	2.24	34320	0
2^{-6}	0.04958	17.53	268580	0
2^{-7}	0.03499	136.26	1961666	0
2^{-8}	0.02462	1662.41	13954402	0

Table 4.3: Example 2, ($\Delta t = 0.5\Delta x$, $\delta = 4\Delta x$). Based on a standard regression analysis we compute: $\mathcal{E} = 0.4(\Delta x)^{0.50}$ with an estimated error of 0.0003 and $\mathcal{T} = 0.00000037\mathcal{E}^{-5.9}$ with an estimated error of 0.006.

Δx	$\mathcal{E}$ (Error)	$\mathcal{T}$ (CPU-time)	# Riemann problems	# shock collisions
2^{-2}	0.05369	0.00	24	0
2^{-3}	0.05926	0.02	48	0
2^{-4}	0.03704	0.04	192	0
2^{-5}	0.02402	0.11	390	0
2^{-6}	0.01409	0.63	1950	0
2^{-7}	0.00747	2.96	5439	0
2^{-8}	0.00380	17.67	23310	0

Table 4.4: Example 2, ($\Delta t = 0.95(\Delta x)^{1/2}$, $\delta = 4\Delta x$). Based on a standard regression analysis we compute: $\mathcal{E} = 0.4(\Delta x)^{0.83}$ with an estimated error of 0.03 and $\mathcal{T} = 0.000002\mathcal{E}^{-2.7}$ with an estimated error of 0.05.

Δx	$\mathcal{E}$ (Error)	$\mathcal{T}$ (CPU-time)	# Riemann problems	# shock collisions
2^{-2}	0.41443	0.03	260	0
2^{-3}	0.31566	0.14	2030	0
2^{-4}	0.12661	0.93	15167	0
2^{-5}	0.07369	6.66	108002	0
2^{-6}	0.04218	47.72	704366	0
2^{-7}	0.02266	352.95	4936281	0
2^{-8}	0.01662	2802.58	35766444	0

Table 4.5: Example 3, ($\Delta t = 0.5\Delta x$, $\delta = 4\Delta x$). Based on a standard regression analysis we compute: $\mathcal{E} = 1.27(\Delta x)^{0.83}$ with an estimated error of 0.02 and $\mathcal{T} = 0.0008\mathcal{E}^{-3.46}$ with an estimated error of 0.02.

Δx	$\mathcal{E}$ (Error)	$\mathcal{T}$ (CPU-time)	# Riemann problems	# shock collisions
2^{-2}	0.37902	0.021	78	0
2^{-3}	0.32070	0.03	171	0
2^{-4}	0.14407	0.1	995	93
2^{-5}	0.09252	0.38	3736	710
2^{-6}	0.05085	2.86	29486	3930
2^{-7}	0.02356	14.35	144292	19658
2^{-8}	0.00976	94.52	986366	93537

Table 4.6: Example 3, ($\Delta t = 0.95(\Delta x)^{1/2}$, $\delta = 4\Delta x$). Based on a standard regression analysis we compute: $\mathcal{E} = 1.75(\Delta x)^{0.87}$ with an estimated error of 0.03 and $\mathcal{T} = 0.0005\mathcal{E}^{-2.78}$ with an estimated error of 0.17.

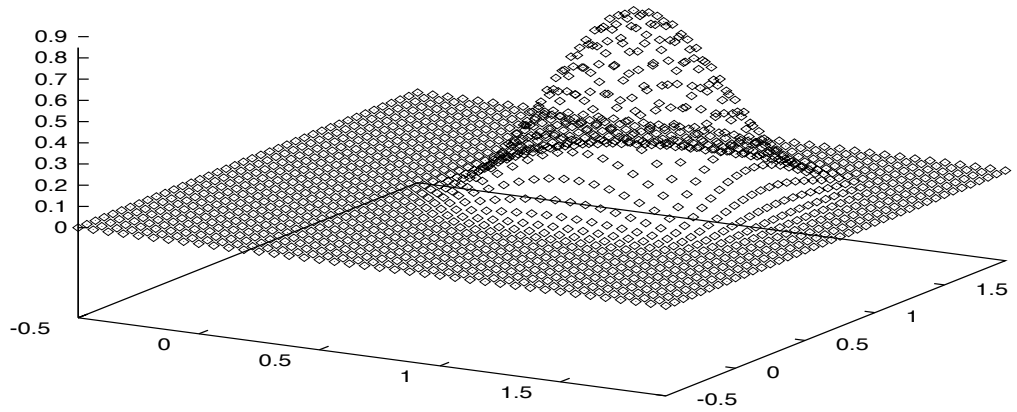

Figure 4.1: A 3D plot of an approximate solution from Example 1, $\Delta x = 2^{-4}$, $\Delta t = 0.5\Delta x$, $\delta = 4\Delta x$

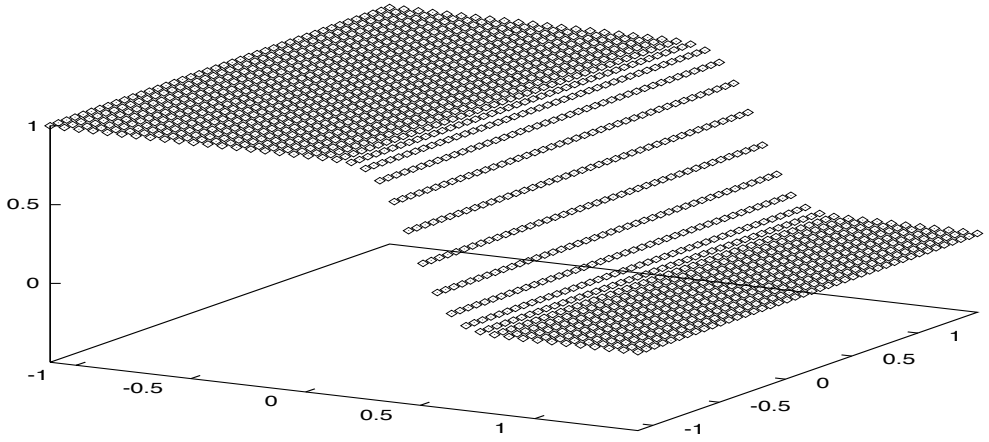

Figure 4.2: A 3D plot of an approximate solution from Example 2, $\Delta x = 2^{-4}$, $\Delta t = 0.5\Delta x$, $\delta = 4\Delta x$

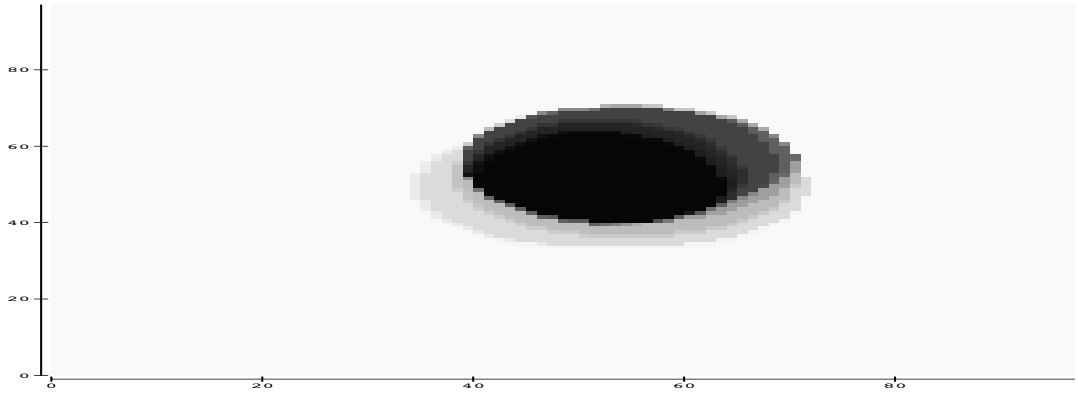

Figure 4.3: A density-plot of an approximate solution from Example 3, $\Delta x = 2^{-5}$, $\Delta t = 0.95(\Delta x)^{1/2}$, $\delta = 4\Delta x$, $T = 0.17$. The different values are shown on a grey scale such that black corresponds 0 and white corresponds to 1. The labels on the axes denote gridblocks, the lower left-hand corner has coordinates $(-1.5, -1.5)$ and the upper right-hand corner $(1.5, 1.5)$.

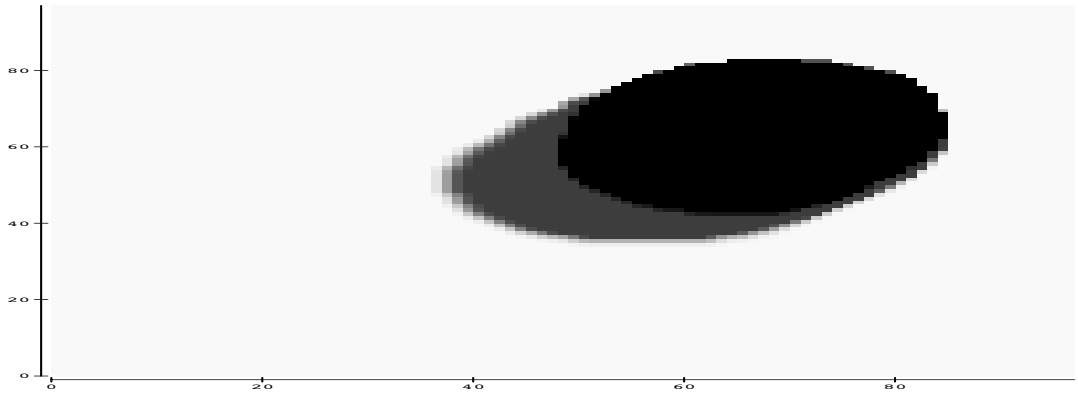

Figure 4.4: A density-plot of an approximate solution from Example 3, $\Delta x = 2^{-5}$, $\Delta t = 0.95(\Delta x)^{1/2}$, $\delta = 4\Delta x$, $T = 0.5$. The different values are shown on a grey scale such that black corresponds 0 and white corresponds to 1. The labels on the axes denote gridblocks, the lower left-hand corner has coordinates $(-1.5, -1.5)$ and the upper right-hand corner $(1.5, 1.5)$.

Bibliography

[1] F. Bradvedt, K. Bradtvedt, etc. *Frontline and Frontsim two full scale, two-phase, black oil simulators based on front tracking*, Surv. Math. Ind. (1993), pp. 185-215.

[2] E. Conway and J. Smoller *Global solutions of the Cauchy problem for quasilinear first order equations in several space variables*, Comm. Pure App. Math. 19 (1966), pp. 95-105.

[3] M. Crandall and A. Majda, *The method of fractional steps for conservation laws*, Numer. Math. 34 (1980), pp. 285-314.

[4] C. M. Dafermos, *Polygonal approximation of solutions of the initial value problem for a conservation law*, J. Math. Anal. Appl., 38, pp. 33–41.

[5] H. Holden, L. Holden and R. Høegh-Krohn, *A numerical method for first order nonlinear scalar conservation laws in on space dimension*, Comput. Math. Appl. 15 (1988), pp. 592-662.

[6] H. Holden and N. H. Risebro, *A method of fractional steps for scalar conservation laws without the CFL condition*, Mathematics of computation, Volume 60, Number 201, January 1993, pp. 221–232.

[7] J.Glimm, *Solutions for nonlinear hyperbolic systems*, Comm. Pure Appl. Math. 18 (1965), pp. 697–715.

[8] S. N. Kružkov, *First order quasilinear equations with several space variables*, Math. USSR. Sb. 10 (1970), pp. 217-243.

[9] N. N. Kuznetsov, *On stable methods of solving a quasi–linear equation of first order in a class of discontinuous functions*, Soviet Math. Dokl. Vol. 16 (1975), No. 6, pp. 1569-1573.

[10] N. N. Kuznetsov, *Accuracy of some approximative methods for computing the weak solutions of a first–order quasi–linear equation*, U.S.S.R. Comput. Math. and Math. Phys. Dokl. (1976), pp. 105-119.

[11] H. P. Langtangen oog A. Tveito *Numeriske metoder i kontinumsmekanikk,* Kompendium i ME-IN 324.

[12] R. Larsen and M. Marx *An introduction to mathematical statistics and its applications*, Prentice-Hall International editions 1986.

[13] Randall J. LeVeque, *Numerical Methods for Conservation Laws*, Birkhäueser 1990.

[14] Randall J. LeVeque, *Time split Methods for partial differential equations*, PhD thesis, Stanford 1982.

[15] B. J. Lucier, *Error bounds for the methods of Glimm, Godunov and LeVeque* SIAM J. Num. Anal, 22 (1985), pp. 1074-1081.

[16] O. Oleinik, *Discontinuous solutions of nonlinear partial differential equations*, Amer. Math. Soc, Transl. Ser. 2, 26 (1963), pp. 95–172.

[17] N. H. Risebro, *The partial differential equation* $u_t + \sum_{i=1}^{n} f_i(x)_{x_i} = 0$ - *a numerical method*, Cand. Scient. Thesis, University of Oslo, 1987.

[18] N. H. Risebro and A. Tveito, *Front tracking applied to a non-strictly hyperbolic system of conservation laws*, SIAM J. Sci. Stat. Comput., Vol. 1, No. 6, (1991), pp. 1401-1419.

[19] Walther Rudin, *Principles of Mathematical Analysis*, Third edition, McGraw-Hill.

[20] J. Smoller, *Shock Waves and Reaction–Diffusion Equations*, Springer–Verlag, New York 1983.

[21] Zhen-Huan Teng, *On the accuracy of fractional steps methods for conservation laws in two dimensions*, Siam. J. Numer. Anal., Vol. 31, No. 1, pp. 43-63, February 1994.