Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft Institute of Applied Mathematics

Literature study

A literature study submitted to the
Delft Institute of Applied Mathematics
in partial fulfillment of the requirements

for the degree

MASTER OF SCIENCE
in
APPLIED MATHEMATICS

by

Jeroen Wille

Delft, the Netherlands
March 2009

MSc THESIS APPLIED MATHEMATICS

Literature study

Jeroen Wille

Delft University of Technology

Daily supervisors	**Responsible professor**
Dr.ir. F. J. Vermolen	Prof. dr. ir. C. Vuik
Dr.ir. J. K. Ryan	

March 2009 Delft, the Netherlands

Contents

1 Introduction **1**

2 Scalar Advection Equation **2**
2.1 One dimensional . . . 2
2.1.1 Perturbed . . . 3
2.1.2 streamline upwind Petrov Galerkin . . . 4
2.1.3 Perturbed SUPG . . . 5
2.1.4 Variable velocity . . . 5
2.2 Two dimensional . . . 6
2.2.1 Perturbed . . . 7
2.2.2 SUPG . . . 8
2.2.3 Perturbed SUPG . . . 8

3 Discontinuity approach **10**
3.1 Discontinuous Galerkin . . . 10
3.2 Application . . . 11
3.3 Shock Detection . . . 12
3.4 Application to a 1D scalar conservation equation . . . 13
3.5 Limiter in 1D . . . 14
3.5.1 Overview . . . 15

4 Results **17**
4.1 1D scalar advection . . . 17
4.2 2D scalar advection . . . 21
4.3 DG approach . . . 23

5 Conclusions and further research **24**

1 Introduction

The goal of this research is to make a model for the water flowing beneath the building of the EWI faculty which is used as heating or cooling depending on the season. This report discusses the relevant literature used as a basis for this model and also dicusses prelimanary results obtained by applying this material.

The groundwater flow can be modelled by a multi-phase flow. Currently, we meglect further investigation of this model and instead focus on the numerical methods that will be used. Our model consits of two systems. The first system has a discontinuity that arises and the second system is for the velocity of the flow. The outline of this literature study is as follows.

In section 2 we discuss the scalar advection equation in one and two dimensions where we also will apply the SUPG method and add a disturbance to the equation. We can then use these results for the system of the velocity of the flow. In section 3 we will discuss the discontinuous Galerkin method and a slope detection method combined with a limiter. These can be used to treat the first system of our model where we will encounter a discontinuity. In section 4 we show and discuss some of the results that were obtained by implementing the metioned methods. Finally in section 5 we will draw some conclusions and state some directions for the remainder of the Master thesis.

2 Scalar Advection Equation

In this section we will consider the scalar advection equation in one and two dimensions. For these equations we will apply the finite element method, for more information we refer to [2]. Furthermore we will apply streamline upwind Petrov Galerkin to deal with discontinuities and we will examine at the equations when some small disturbance is added.

2.1 One dimensional scalar advection equation

Consider the following equation

$$u_t + u_x = 0, \tag{1}$$

on the unit rod. This is a simplification of $u_t + \bar{u} \cdot u_x = 0$ with $\bar{u} = 1$. We take a Dirichlet boundary condition at $x = 0$ so $u(0) = a$ with $a \in \mathbb{R}$ being some constant and we take an initial condition $u(x, 0) = g(x)$. We want to solve this sytem using the Finite Element Method. To do so we multiply equation (1) by some test function v, subject to $v(0) = 0$, and we integrate the equation over the domain $[0, 1]$ to get

$$\frac{\partial}{\partial t}\int_0^1 uvdx = -\int_0^1 \frac{\partial u}{\partial x}vdx. \tag{2}$$

We will approximate the solution u by $u = \sum_{j=1}^{n} u_j\varphi_j + a\varphi_0$ where φ_j is a basis function. Next we substitute $u = \sum_{j=1}^{n+n_b} u_j\varphi_j + a\varphi_0$ and $v = \varphi_i$ for $i = 1, \ldots, n$ to obtain

$$\sum_{j=1}^{n} \frac{\partial u_j}{\partial t}\int_0^1 \varphi_i\varphi_j dx = -\sum_{j=1}^{n} u_j \int_0^1 \frac{\partial \varphi_j}{\partial x}\varphi_i dx - \frac{\partial}{\partial t}\int_0^1 a\varphi_0\varphi_i dx - a\int_0^1 \frac{\partial \varphi_0}{\partial x}\varphi_i dx, \quad i = 1, \ldots, n. \tag{3}$$

We are going to use linear functions on a equidistant grid. We therefore have over an element of size h two non zero test functions, which gives us a 2×2 element matrix. For the element mass matrix we simply obtain

$$M_{ij} = \int_0^1 \varphi_i\varphi_j dx = \int_{x_i}^{x_{i+1}} \varphi_i\varphi_j dx \overset{NC}{=} \frac{h}{2}(\varphi_i\varphi_j(x_i) + \varphi_i\varphi_j(x_{i+1})) = \frac{h}{2}\delta_{ij}. \tag{4}$$

For the element stiffness matrix we have

$$S = \begin{bmatrix} 0.5 & -0.5 \\ 0.5 & -0.5 \end{bmatrix}. \tag{5}$$

To solve this we are going to use implicit Euler. This gives us

$$\begin{aligned} M\frac{\left(u^{k+1} - u^k\right)}{\Delta t} &= Su^{k+1} + f, \\ &\Rightarrow \\ (M - \Delta t S)u^{k+1} &= Mu^k + \Delta t \cdot f. \end{aligned} \tag{6}$$

This system can be solved directly by inverting $(M - \Delta t S)$ for small n or iteratively for large n.

2.1.1 Perturbed

If we add a small disturbance to this problem, equation (1) becomes

$$u_t + u_x = \varepsilon u_{xx}. \tag{7}$$

Another boundary condition is necessary and as we will see

$$\frac{\partial u}{\partial x}(1) = b, \tag{8}$$

for some $b \in \mathbb{R}$, will appear to be a natural boundary condition for this problem. We rewrite the equation to separate time and space derivatives, multiply both sides with a test function v and integrate over the domain to obtain

$$\frac{\partial}{\partial t}\int_0^1 uvdx = \varepsilon \int_0^1 u_{xx}vdx - \int_0^1 \frac{\partial u}{\partial x}vdx. \tag{9}$$

This time we will apply integration by parts first. This will give us

$$\frac{\partial}{\partial t}\int_0^1 uvdx = \varepsilon u_x v\Big|_0^1 - \varepsilon \int_0^1 u_x v_x dx - \int_0^1 \frac{\partial u}{\partial x}vdx. \tag{10}$$

Since we have an essential boundary condition at $x = 0$ the test function will be zero at $x = 0$. Now we see the need for the natural boundary condition in (8) which will give us

$$\frac{\partial}{\partial t}\int_0^1 uvdx = \varepsilon bv(1) - \varepsilon \int_0^1 u_x v_x dx - \int_0^1 \frac{\partial u}{\partial x}vdx. \tag{11}$$

Again we substitute $u = \sum_{j=1}^n u_j\varphi_j + a\varphi_0$ and $v = \varphi_i$ for $i = 1, \ldots, n$. This gives us

$$\begin{aligned}\sum_{j=1}^n \frac{\partial u_j}{\partial t}\int_0^1 \varphi_i\varphi_j dx = \varepsilon b\varphi_i(1) - \varepsilon\sum_{j=1}^n u_j\int_0^1 \frac{\partial\varphi_i}{\partial x}\frac{\partial\varphi_j}{\partial x}dx - \sum_{j=1}^n u_j\int_0^1 \frac{\partial\varphi_j}{\partial x}\varphi_i dx \\ - \frac{\partial}{\partial t}\int_0^1 a\varphi_0\varphi_i dx - \varepsilon a\int_0^1 \frac{\partial\varphi_i}{\partial x}\frac{\partial\varphi_0}{\partial x}dx - a\int_0^1 \frac{\partial\varphi_0}{\partial x}\varphi_i dx, \quad i = 1, \ldots, n.\end{aligned} \tag{12}$$

Again, we take linear triangles and a equidistant grid and this gives us the same mass matrix as in (4). The stiffness matrix now gets an extra term

$$S_{ij} = \begin{bmatrix} 0.5 & -0.5 \\ 0.5 & -0.5 \end{bmatrix} + \varepsilon \begin{bmatrix} \frac{-1}{h} & \frac{1}{h} \\ \frac{1}{h} & \frac{-1}{h} \end{bmatrix}. \tag{13}$$

Lastly, note that the right hand side vector f gets an extra term in the last position. This gives

$$f = f + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \varepsilon b \end{pmatrix}. \tag{14}$$

2.1.2 streamline upwind Petrov Galerkin

If we have a continuous initial condition we will obtain a smooth solution. However, if we have a discontiuous initial condition we see that we obtain wiggles in our solution. When working with concentrations this is not desirable since, for example, negative concentrations do not exist, and our solution should satisfy certain maximum principles. To treat this we implemented streamline upwind Petrov Galerkin, abbreviated by SUPG. Using SUPG we will not obtain wiggles anymore, however we then have smearing.

The idea of SUPG is similar to the finite element method. However, instead of multiplying the equation with a test function v we now multiply the equation with the same test function v plus an extra function p which is continuous over the element. After integrating over the domain our equation now becomes

$$\int_0^1 u_t v dx + \int_0^1 u_t p dx = -\int_0^1 u_x v dx - \int_0^1 u_x p dx. \tag{15}$$

Since p is only continuous over an element, two of the integrals may not be subjected to partial integration. Therefore, we split the integral into smaller integrals over the elements and hence the contributions over the element boundaries are neglected. Equation (15) then becomes

$$\int_0^1 u_t v dx + \sum_{k=1}^{n_{el}} \int_{el_k} u_t p dx = -\int_0^1 u_x v dx - \sum_{k=1}^{n_{el}} \int_{el_k} u_x p dx. \tag{16}$$

A very common choice for p is $\frac{h}{2}\frac{\partial \varphi_i}{\partial x}$ or, more generally, when s is the velocity, $p = \frac{h\xi}{2}\frac{s}{\|s\|}\frac{\partial \varphi_i}{\partial x}$, so that the function follows the direction of the flow and ξ denotes some fraction we choose. Substituting this into the equation, and taking $u = \sum_{j=1}^{n} u_j\varphi_j + a\varphi_0$ and $v = \varphi_i$, $i = 1, \ldots, n$, as before, equation (16) yields

$$\begin{aligned}
&\frac{\partial}{\partial t}\sum_{j=1}^{n} u_j \int_0^1 \varphi_i\varphi_j dx + \frac{\partial}{\partial t}\frac{h\xi}{2}\sum_{j=1}^{n}\sum_{k=1}^{n_{el}} u_j \int_{el_k} \varphi_j \frac{\partial \varphi_i}{\partial x} dx \quad = \\
&-\sum_{j=1}^{n} u_j \int_0^1 \frac{\partial \varphi_j}{\partial x}\varphi_i dx - \frac{h\xi}{2}\sum_{j=1}^{n}\sum_{k=1}^{n_{el}} u_j \int_{el_k} \frac{\partial \varphi_i}{\partial x}\frac{\partial \varphi_j}{\partial x} dx \\
&\quad -\frac{\partial}{\partial t}\int_0^1 a\varphi_0\varphi_i dx - \frac{\partial}{\partial t}\frac{h\xi}{2}\sum_{k=1}^{n_{el}} \int_{el_k} a\varphi_0 \frac{\partial \varphi_i}{\partial x} dx \\
&\quad -a\int_0^1 \frac{\partial \varphi_0}{\partial x}\varphi_i dx - \frac{h\xi}{2}\sum_{k=1}^{n_{el}} a\int_{el_k} \frac{\partial \varphi_i}{\partial x}\frac{\partial \varphi_0}{\partial x} dx, \qquad i = 1, \ldots, n.
\end{aligned} \tag{17}$$

This means that the mass matrix and the stiffness matrix have an extra contribution from the integrals over the elements. Taking $\xi = 1$ we obtain for the mass matrix

$$M_{ij} = \begin{bmatrix} \frac{h}{2} & 0 \\ 0 & \frac{h}{2} \end{bmatrix} + \begin{bmatrix} \frac{-h}{4} & \frac{-h}{4} \\ \frac{h}{4} & \frac{h}{4} \end{bmatrix} = \frac{h}{4}\begin{bmatrix} 0 & -1 \\ 1 & 4 \end{bmatrix}. \tag{18}$$

For the stiffness matrix we have

$$S = \begin{bmatrix} 0.5 & -0.5 \\ 0.5 & -0.5 \end{bmatrix} + \begin{bmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix} \tag{19}$$

2.1.3 Perturbed SUPG

We will apply SUPG as in the previous section to the perturbed equation $u_t + u_x = \varepsilon u_{xx}$. We obtain two extra terms

$$\varepsilon \int_0^1 u_{xx} v dx + \varepsilon \int_0^1 u_{xx} p dx \tag{20}$$

and apply the same approach. Since the linear elements will have a zero for the second derivative, the second integral will vanish. The first integral is identical to the integral part in equation (9) and so we can apply the same approach with the same result. In total we need to add the term

$$\frac{\varepsilon}{h} \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \tag{21}$$

to the stiffness matrix and the right hand side vector will be identical to the right hand side in (14).

2.1.4 Variable velocity

In this section we will look at the following equation

$$u_t + v(t,x) u_x = 0. \tag{22}$$

In this equation we have a time dependent speed v. Although it can be space dependent as well, we will see that for a 1-dimensional problem the speed will be constant in space.
We need a way to derive the speed at a certain time t. For this we have Darcy's Law

$$v = -\lambda(u) \frac{\partial p}{\partial x}, \tag{23}$$

where $\lambda(u)$ is some function depending on the solution u and p is the pressure. Furthermore we know that the gradient of the velocity should be equal to zero which in one-dimension gives us

$$\frac{\partial v}{\partial x} = -\frac{\partial}{\partial x}\left(\lambda(u) \frac{\partial p}{\partial x}\right) = 0. \tag{24}$$

Solving this system at every time step, we can derive the velocity for the next time step. Since we have $\frac{\partial v}{\partial x} = 0$ we can only conclude that v should be constant in space. This makes it somewhat easier to calculate the velocity after the system has been solved for the pressure. We have

$$\begin{aligned} v &= -\lambda(u) \frac{\partial p}{\partial x} \\ -\frac{v}{\lambda(u)} &= \frac{\partial p}{\partial x} \\ -\int_0^1 \frac{v}{\lambda(u)} dx &= \int_0^1 \frac{\partial p}{\partial x} dx \\ -v \int_0^1 \frac{dx}{\lambda(u)} &= p(1) - p(0) \\ v &= \frac{p(0) - p(1)}{\int_0^1 \frac{dx}{\lambda(u)}} \end{aligned} \tag{25}$$

$$\tag{26}$$

In general one uses the following function for $\lambda(u)$ for which the integral can be approximated using a Newton-Cotes rule.

$$\lambda(u) = \frac{u^2}{(1-u)^2 + u^2} \tag{27}$$

Using a Dirichlet boundary condition on $x = 0$, in order to obtain a positive speed, and a negative Neumann boundary condition on $x = 1$, for instance $\frac{\partial p}{\partial x} = a$, $a < 0$, we can solve system (24) using finite elements in order to determine $p(1)$. Multiplying with a test function w and integrating over the domain we get

$$\begin{aligned} \int_0^1 -\frac{\partial}{\partial x}\left(\lambda(u)\frac{\partial p}{\partial x}\right) w dx &= 0, \\ \int_0^1 \lambda(u)\frac{\partial p}{\partial x}\frac{\partial w}{\partial x} dx &= \lambda(u)\frac{\partial p}{\partial x} w\Big|_0^1, \\ \int_0^1 \lambda(u)\frac{\partial p}{\partial x}\frac{\partial w}{\partial x} dx &= a\lambda(u(1))w(1). \end{aligned} \tag{28}$$

Substituting $p = \sum_{j=1}^n p_j \varphi_j$ and $w = \varphi_i$, for $i = 1, \ldots, n$ we get

$$\sum_{j=1}^n p_j \int_0^1 \lambda(u)\frac{\partial \varphi_j}{\partial x}\frac{\partial \varphi_i}{\partial x} dx = a\lambda(u(1))\varphi_i(1), \quad i = 1, \ldots, n \tag{29}$$

This leads to the following element matrix using linear triangles and an equidistant grid:

$$S_{ij} = \begin{bmatrix} \frac{\lambda(u(x_i))}{h} & \frac{-(\lambda(u(x_i))+\lambda(u(x_j)))}{2h} \\ \frac{-(\lambda(u(x_i))+\lambda(u(x_j)))}{2h} & \frac{\lambda(u(x_j))}{h} \end{bmatrix}. \tag{30}$$

For the right hand side we get

$$f = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ a\lambda(u(1)) \end{pmatrix}. \tag{31}$$

The rest of the syste remains the same as in the previous sections.

2.2 Two dimensional scalar advection equation

We look at the two dimensional scalar advection equation

$$u_t + v_1 \cdot u_x + v_2 \cdot u_y = 0. \tag{32}$$

Or, in more general terms

$$u_t + \mathbf{v} \cdot \nabla u = 0, \tag{33}$$

with the boundary condition $u = f$ on Γ for some function f.
We multiply this equation with a test function v, subject to $v|_\Gamma = 0$, and integrate over the domain to obtain

$$\int_\Omega u_t v d\Omega = -\int_\Omega v\mathbf{v}\nabla u d\Omega. \tag{34}$$

Next we substitute $u = \sum_{j=1}^{n+n_b} u_j \varphi_j$, where n_b denotes the number of boundary nodes, and $v = \varphi_i$ for $i = 1, \ldots, n$ to obtain

$$\frac{\partial}{\partial t} \sum_{j=1}^{n+n_b} u_j \int_\Omega \varphi_i \varphi_j d\Omega = - \sum_{j=1}^{n+n_b} u_j \int_\Omega \varphi_i \mathbf{v} \nabla \varphi_j d\Omega, \quad i = 1, \ldots, n. \tag{35}$$

We are going to use linear triangular elements so $\varphi_i = a_0^i + a_1^i x + a_2^i y$. For the element mass matrix we have

$$M_{ij} = \int_\Omega \varphi_i \varphi_j d\Omega \overset{NC}{=} \frac{|\Delta|}{6} \sum_{k=1}^{3} \varphi_i \varphi_j (\bar{x}_k) = \frac{|\Delta|}{6} \delta_{ij} \tag{36}$$

For the element stiffness matrix we have

$$S_{ij} = -\frac{|\Delta|}{6} \left(v_1 a_1^j + v_2 a_2^j \right). \tag{37}$$

To solve this, we will use implicit Euler, so we have

$$M \frac{\left(u^{k+1} - u^k\right)}{\Delta t} = S u^{k+1} + f \Rightarrow (M - \Delta t S) u^{k+1} = M u^k + \Delta t \cdot f. \tag{38}$$

2.2.1 Perturbed

We can add a small disturbance to this problem. and then equation (33) becomes

$$u_t + \mathbf{v} \nabla u = \varepsilon \mathrm{div}(\nabla u). \tag{39}$$

Again, we need another boundary condition and as we will see this boundary condition is

$$\frac{\partial u}{\partial n} = g, \tag{40}$$

for some function g on Γ_2, whereas the Dirichlet boundary condition will now only be applied to $\Gamma = \Gamma_1$, will appear to be a natural boundary condition for this problem.
Multiply by some test function v, subject to $v|_{\Gamma_1} = 0$, and integrate over the domain to obtain

$$\int_\Omega u_t v d\Omega = - \int_\Omega v \mathbf{v} \nabla u + \int_\Omega \varepsilon \mathrm{div}(\nabla u) v d\Omega. \tag{41}$$

This time we can apply Green's theorem on the latter integral to obtain

$$\int_\Omega u_t v d\Omega = - \int_\Omega v \mathbf{v} \nabla u - \varepsilon \int_\Omega \nabla u \nabla v d\Omega + \varepsilon \oint_\Gamma v \frac{\partial u}{\partial n} d\Gamma \tag{42}$$

Now we split the boundary Γ over the two separate boundaries in order to apply boundary conditions $u = f$ on Γ_1 and (40) on the latter integral to obtain

$$\int_\Omega u_t v d\Omega = - \int_\Omega v \mathbf{v} \nabla u - \varepsilon \int_\Omega \nabla u \nabla v d\Omega + \varepsilon \oint_{\Gamma_2} v g d\Gamma \tag{43}$$

We substitute $u = \sum_{j=1}^{n+n_b} u_j \varphi_j$ and $v = \varphi_i$ for $i = 1, \ldots, n$ and obtain

$$\begin{aligned} \frac{\partial}{\partial t} \sum_{j=1}^{n+n_b} u_j \int_\Omega \varphi_i \varphi_j d\Omega \quad &= \quad - \sum_{j=1}^{n+n_b} u_j \int_\Omega \varphi_i \mathbf{v} \nabla \varphi_j - \varepsilon \sum_{j=1}^{n+n_b} u_j \int_\Omega \nabla \varphi_j \nabla \varphi_i d\Omega \\ &\quad + \varepsilon \oint_{\Gamma_2} \varphi_i g d\Gamma, \quad i = 1, \ldots, n. \end{aligned} \tag{44}$$

We now have an extra term in the element stiffness matrix and we have a non-zero element vector. So we now have

$$S_{ij} = S_{ij} - \varepsilon\frac{|\Delta|}{2}(a_1^i a_1^j + a_2^i a_2^j) \quad (45)$$

$$F = \frac{\varepsilon h}{2}\begin{pmatrix} g(x_{i-1}) \\ g(x_i) \end{pmatrix} \quad (46)$$

2.2.2 SUPG

We can also apply the SUPG method to the two-dimensional problem. We take $v = v + p$ where p may be discontinuous over the elements. However, p need not be constant anymore. Similar to the one dimensional problem we have two extra integrals. One integral containing the time derivative times p, the other containing the space derivative times p. For the time derivative integral we have

$$\int_\Omega u_t p d\Omega = \frac{\partial}{\partial t}\int_\Omega up d\Omega. \quad (47)$$

For the space derivative integral we have

$$-\int_\Omega \mathbf{v}\nabla up d\Omega. \quad (48)$$

Typically, we take for p

$$p = \frac{h\xi}{2}\frac{\mathbf{v}}{\|\mathbf{v}\|}\nabla\varphi_i, i = 1, \ldots, n. \quad (49)$$

Substituting this and $u = \sum_{j=1}^{n+n_b} u_j\varphi_j$ into equations (47) and (48) leads to

$$\frac{\partial}{\partial t}\int_\Omega up d\Omega. = \frac{\partial}{\partial t}\frac{h\xi}{2}\frac{\mathbf{v}}{\|\mathbf{v}\|}\sum_{j=1}^{n+n_b} u_j\int_\Omega \varphi_j\nabla\varphi_i d\Omega, i = 1, \ldots, n, \quad (50)$$

$$-\int_\Omega \mathbf{v}\nabla up d\Omega = -\frac{h\xi}{2}\frac{\mathbf{v}\cdot\mathbf{v}}{\|\mathbf{v}\|}\sum_{j=1}^{n+n_b} u_j\int_\Omega \nabla\varphi_i\nabla\varphi_j d\Omega, i = 1, \ldots, n. \quad (51)$$

Applying Newton-Cotes formulas, the integrals give us extra terms for the element mass and stiffness matrix as:

$$M_{ij} = M_{ij} + \frac{h\xi}{2}\frac{\mathbf{v}}{\|\mathbf{v}\|}\frac{|\Delta|}{6}\begin{pmatrix} a_1^i \\ a_2^i \end{pmatrix} \quad (52)$$

$$S_{ij} = S_{ij} - \frac{h\xi}{2}\frac{\mathbf{v}\cdot\mathbf{v}}{\|\mathbf{v}\|}\frac{|\Delta|}{2}\left(a_1^i a_1^j + a_2^i a_2^j\right) \quad (53)$$

2.2.3 Perturbed SUPG

Similar as in the one-dimensional case, we can apply both a perturbation and the SUPG method. We now only look at multiplying $\varepsilon\text{div}(\nabla u)$ with $v + p$ and integrating this over the domain. In this way we have

$$\varepsilon\int_\Omega \text{div}(\nabla u)v d\Omega + \varepsilon\int_{d\Omega}\text{div}(\nabla u)p d\Omega. \quad (54)$$

The first integral already is completely determined in section 2.2.1 which means we only have to consider the second integral. Applying again the value for p as in (49) we obtain an integral containing a second derivative part times a first derivate part, so we do not apply any integration method. However the chosen elements are linear which means $\mathrm{div}(\nabla u)$ will be zero and thus the entire integral will be zero. We only will have left the original disturbed problem and we do not need to proceed further for now.

3 Discontinuity approach

3.1 Discontinuous Galerkin

In this section we take a look at the discontinuous Galerkin method, abbreviated by DG, for the scalar conservation law in one dimension. For more information we refer to [1]. The idea behind the discontinuous Galerkin method is to approximate each cell not by just one unknown but to approximate the value within each cell by some linear combination of polynomials of degree at most k for some $k \in \mathbb{N}$. Let us examine the following simple model

$$\begin{aligned} u_t + f(u)_x &= 0, && (55)\\ u(x,0) &= u_0(x), && (56) \end{aligned}$$

with periodic boundary conditions.

To derive the weak formulation, first partition the interval $(0,1)$ into $\{x_{j+1/2}\}_{j=0}^{N}$ and set $I_j = \left(x_{j-1/2}, x_{j+1/2}\right)$ and $\Delta_j = x_{j+1/2} - x_{j-1/2}$ for $j = 1, \ldots, N$. Denote by Δx the maximum element size, $\max_{1\le j\le N} \Delta_j$. Define V_h to be the following finite dimensional space

$$V_h = V_h^k \equiv \left\{ v \in L^1(0,1) : v|_{I_j} \in P^k(I_j), j = 1, \ldots, N \right\}, \tag{57}$$

so that V_h is the space of all functions v which are a piecewise polynomial on an interval I_j, of degree at most k. For each time $t \in (0,T)$ we want our approximation of u to be in V_h. In order to determine this approximation we first multiply (55) and (56) with some arbitrary smooth function v and integrate over the interval I_j to get

$$\begin{aligned} \int_{I_j} \frac{\partial}{\partial t} u(x,t)\, v(x)\, dx + \int_{I_j} \frac{\partial}{\partial x} f(u(x,t))\, v(x)\, dx &= 0, && (58)\\ \int_{I_j} u(x,0)\, v(x)\, dx &= \int_{I_j} u_0(x)\, v(x)\, dx. && (59) \end{aligned}$$

We then apply integration by parts on the second integral in (58) to remove the spatial derivative of the function $f(u(x,t))$

$$\begin{aligned} &\int_{I_j} \frac{\partial}{\partial t} u(x,t)\, v(x)\, dx - \int_{I_j} f(u(x,t)) \frac{\partial}{\partial x} v(x)\, dx \\ &+ f\left(u\left(x_{j+1/2},t\right)\right) v\left(x_{j+1/2}^-\right) - f\left(u\left(x_{j-1/2},t\right)\right) v\left(x_{j-1/2}^+\right) = 0, && (60)\\ &\int_{I_j} u(x,0)\, v(x)\, dx = \int_{I_j} u_0(x)\, v(x)\, dx. && (61) \end{aligned}$$

The functions v are only defined within each interval I_j and therefore we use $x_{j+1/2}^-$ to indicate the point $x_{j+1/2}$ approached from the left, and $x_{j-1/2}^+$ to indicate the point $x_{j-1/2}$ approached from the right. Next we replace our smooth functions v by test functions $v_h \in V_h$ and replace the exact solution u by the approximation u_h. The function u_h now is discontinuous at the points $x_{j+1/2}$ for $j = 1, \ldots, N$. Therefore, we need to replace the function $f\left(u\left(x_{j+1/2},t\right)\right)$ with a numerical analogue that depends on both $x_{j+1/2}^-$ and $x_{j+1/2}^+$. For this we define the function h by

$$h(u)_{j+1/2}(t) = h\left(u\left(x_{j+1/2}^-, t\right), u\left(x_{j+1/2}^+, t\right)\right), \tag{62}$$

where this function h is free to choose by the user. We consider two possible choices for h

- The Godunov flux:

$$h^G(a,b) = \begin{cases} \min_{a\le u\le b} f(u), & \text{if } a \le b, \\ \max_{a\ge u\ge b} f(u), & \text{if } a > b; \end{cases}$$

- Upwind flux:

$$h^{UW}(a,b) = f(a).$$

As basis functions we take the Legendre polynomials P_l so that we can exploit their L^2-orthogonality in order to get a diagonal mass matrix

$$\int_{-1}^{1} P_l(x) P_m(x)\, dx = \left(\frac{2}{2l+1}\right)\delta_{lm}, \tag{63}$$

where δ_{lm} denotes the Kronecker delta function. We also note that have the equalities $P_l(1) = 1$ and $P_l(-1) = (-1)^l$. Taking $v_h(x) = \varphi_j^m(x)$ and $u_h(x,t) = \sum_{l=0}^{k} u_j^l \varphi_j^l$ with $\varphi_j^l = P_l\left(\frac{2(x-x_j)}{\Delta_j}\right)$. We substitute these basis functions in (60) and (61) to obtain

$$\begin{aligned} &\left(\frac{1}{2m+1}\right)\frac{\partial}{\partial t}u_j^m(t) - \frac{1}{\Delta_j}\int_{I_j} f\left(u_h(x,t)\right)\frac{\partial}{\partial x}\varphi_j^m(x)dx \\ &+\frac{1}{\Delta_j}\left\{h\left(u_h(x_{j+1/2})\right)(t) - (-1)^m h\left(u_h(x_{j-1/2})\right)(t)\right\} &= 0 \qquad (64) \\ &u_j^0 = \frac{2m+1}{\Delta_j}\int_{I_j} u_0(x)\varphi_j^m(x)dx \qquad (65) \\ &\forall j \in \{1,\dots,N\},\ m \in \{0,\dots,k\} \end{aligned}$$

3.2 Application

We will apply discontinuous Galerkin first to the simplified conservation equation

$$u_t + u_x = 0, \tag{66}$$

with some initial condition $g(x)$. For the function h we use the upwind flux to prevent spurious oscillations. We will use constant and linear polynomials which gives us $k = 1$. Furthermore we partition the interval $(0,1)$ int $N \in \mathbb{N}$ equally spaced elements and thus get we have $(0,1) = (x_{j-1/2}, x_{j+1/2})_{j=1}^{N}$. Since each interval Δ_j has the same length, we have $\Delta x = \frac{1}{N} = \Delta_j,\ \forall j$.
Equations (64) and (65) now simplify to

$$\begin{aligned} &\left(\frac{1}{2m+1}\right)\frac{\partial}{\partial t}u_j^m(t) - \frac{1}{\Delta x}\sum_{l=0}^{k} u_j^l \int_{I_j} \varphi_j^l \frac{\partial}{\partial x}\varphi_j^m(x)dx \\ &+\frac{1}{\Delta x}\left\{\sum_{l=0}^{k} u_j^l\varphi_j^l(x_{j+1/2}) - (-1)^m\sum_{l=0}^{k} u_{j-1}^l\varphi_{j-1}^l(x_{j-1/2})\right\} &= 0, \qquad (67) \\ &\frac{2m+1}{\Delta x}\int_{I_j} u_0(x)\varphi_j^m(x)dx &= u_j^0. \qquad (68) \end{aligned}$$

For the mass matrix we simply have

$$M_{ml} = \frac{1}{2m+1}\delta_{ml}. \tag{69}$$

Due to the spatial integral we obtain the following stiffness matrix

$$\begin{aligned} S_{ml} &= \frac{1}{\Delta x}\int_{I_j} \varphi_j^l \frac{\partial}{\partial x}\varphi_j^m dx \\ &\Rightarrow \\ S &= \begin{bmatrix} 0 & 0 \\ \frac{2}{\Delta x} & 0 \end{bmatrix}. \end{aligned} \tag{70}$$

Next we have two matrices, A and B, corresponding to the flux of the current cell and the previous cell. We have $A\mathbf{u}_j + B\mathbf{u}_{j-1}$:

$$\frac{1}{\Delta x}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\mathbf{u}_j + \frac{1}{\Delta x}\begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}\mathbf{u}_{j-1}, \quad \mathbf{u}_j = \begin{pmatrix} u_j^0 \\ u_j^1 \end{pmatrix} \tag{71}$$

Next we apply a simple Euler forward scheme, where $\dot{\mathbf{u}}$ stands for the value at the next time step, for the time derivative to get

$$M\dot{\mathbf{u}}_j = (M - \Delta t A + \Delta t S)\,\mathbf{u}_j - \Delta t B\mathbf{u}_{j-1} \tag{72}$$

3.3 Shock Detection

When taking a discontinuous initial condition when using the discontinuous Galerkin method we have wiggles near discontinuities. In practice this may lead to nonlinear instabilities and to nonphysical solutions like negative concentrations or pressures. Since limiting reduces the quality of the solution in smooth regions, we try to determine only these regions where limiting is needed. Therefore we need to be able to detect discontinuities.

Since the derivation of this method does not depend on a specific dimension, we will derive everything for multi dimensions and indicate cells by Ω instead of I, which is only convenient for dimension one. Let us look at a given problem on a certain cell Ω_j. We partition the boundary $\partial\Omega_j$ of this cell into two parts. Part one is where the flow enters and therefore $(v \cdot n) < 0$. Denote this by $\partial\Omega_j^-$. The other part is where is outflow, $(v \cdot n) > 0$, which we denote by $\partial\Omega_j^+$. According to [4] smooth solutions of hyperbolic conservation laws show strong superconvergence phenomena at outflow boundaries such that

$$\frac{1}{|\partial\Omega_j^+|}\int_{\partial\Omega_j^+} (Q_j - q)\ d\Gamma = \mathcal{O}(h^{2k+1}). \tag{73}$$

Here, $|\partial\Omega_j^+|$ denotes the length/area of $\partial\Omega_j^+$. Note that q is the exact solution of the equation and Q_j is the discontinuous Galerkin value of q on Ω_j. We will use this information to detect locations of shocks. Therefore, we consider a jump in Q_j across $\partial\Omega_j^-$ and we examine

$$\mathbf{I}_j = \int_{\partial\Omega_j^-} (Q_j - Q_{nbj})\ d\Gamma = \int_{\partial\Omega_j^-} (Q_j - q)\ d\Gamma + \int_{\partial\Omega_{nbj}^+} (q - Q_{nbj})\ d\Gamma. \tag{74}$$

In this equation Q_{nbj} stands for a neighboring element of Ω_j with common boundary $\partial\Omega_{j,nbj}$. We know from equation (73) that the second integral in this equation is $\mathcal{O}(h^{2k+2})$. Furthermore we know that the first integral across the inflow boundary is $\mathcal{O}(h^{k+2})$ so $\mathbf{I}_j$ is $\mathcal{O}(h^{k+2})$ for smooth solutions on $\partial\Omega_j^-$. However, if q is discontinuous near $\partial\Omega_j$ then $q-Q_j$ and/or $q-Q_{nbj}$ will be $\mathcal{O}(1)$. Using this information we construct a discontinuity detector by normalizing $\mathbf{I}_j$ to some convergence rate and the solution Ω_j:

$$\mathcal{I}_j = \frac{\left|\int_{\partial\Omega_j^-} (Q_j - Q_{nbj})\ d\Gamma\right|}{h^{(k+1)/2}|\partial\Omega_j^-|\|Q_j\|} \tag{75}$$

We know that in smooth regions $\mathbf{I}_j \to 0$, and so $\mathcal{I}_j \to 0$ as well, if $h \to 0$ or $k \to \infty$. However near a discontinuity we have $\mathcal{I}_j \to \infty$. The discontinuity detection scheme we use is then

$$\begin{cases} \mathcal{I}_j > 1 & \Rightarrow \quad q \text{ is discontinuous,} \\ \mathcal{I}_j < 1 & \Rightarrow \quad q \text{ is smooth.} \end{cases} \tag{76}$$

3.4 Application to a 1D scalar conservation equation

We will apply this shock detector to equation (67) to test the effectiveness of this strategy. We note the following:

- The boundary $\partial\Omega_j^-$ consists of one point, $x_{j-1/2}$;
- For the norm, we take the element average;
- The term $|\partial\Omega_j^-|$ drops out since it consists of only one point;
- The integral reduces to $Q_j(x_{j-1/2}^-) - Q_{j-1}(x_{j-1/2}^+)$;

For the element average we get the following

$$\begin{aligned} \|Q_j\| &= \frac{1}{\Delta_j}\int_{x_{j-1/2}}^{x_{j+1/2}} Q_j(x)dx \\ &= \frac{1}{\Delta_j}\int_{x_{j-1/2}}^{x_{j+1/2}} u_j^0 + u_j^1\frac{2(x-x_j)}{\Delta_j}dx \\ &= \frac{1}{\Delta_j}\int_{x_{j-1/2}}^{x_{j+1/2}} u_j^0 dx \\ &= u_j^0 \end{aligned} \tag{77}$$

So taking $k = 1$, the norm reduces to u_j^0.

As a first test we take the following initial condition

$$u(x,0) = \begin{cases} 5, & \text{if } x < 0.5, \\ 1, & \text{otherwise.} \end{cases} \tag{78}$$

After 100 time steps we get the results in Figure 1. The + signs indicate where the shock detection method indicates a discontinuity. This figure clearly shows us that taking a discontinuous initial condition will lead to wiggles, and furthermore it shows us that the shock

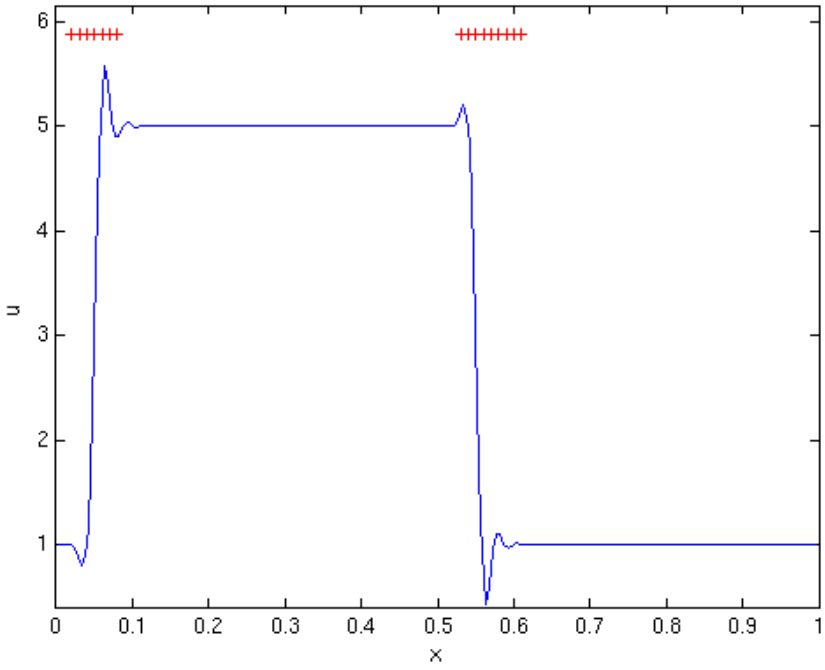

Figure 1: Solution of (67) with initial condition (78) after 100 time steps.

detection method quite accurate indicates the places where limiting is needed. However this shock detection method also has a drawback. When we take an initial condition with some large, but still smooth slopes, the detector may indicate that slopes are discontinuities and so a limiter method may adjust the coefficients where it is not needed. Figure 2 is a nice example of this issue. In this example we have chosen the following initial condition

$$u(x,0) = 5 + \sin(12\pi x). \tag{79}$$

It is quite clear that the large slopes are indicated as shocks. However at first it seems rather strange that only the bottom peaks are indicated as discontinuities, and not the upper peaks as well. This is due to the definition of the indicator $\mathcal{I}_j$. In this definition we divide by the norm of Q_j and in 1D this is the same as the absolute value of the zeroth coefficient. At the top peaks this value is 1.5 times as large as at the bottom peaks, so the indicator value is here 1.5 times as low as at the bottom peaks. This factor leads to not detecting shocks at the top peaks.

3.5 Limiter in 1D

Now that we have a method to determine shocks, we need a method to limit near these shocks. We implement the limiter discussed in [3]. The idea behind this limiter is to start limiting in each cell where it is needed from the top-coefficient and stop if the limited value is the same as the coefficient. For this we first define the minmod function by

$$\text{minmod}(a,b,c) = \begin{cases} \text{sgn}(a)\min(|a|,|b|,|c|) & \text{if } \text{sgn}(a) = \text{sgn}(b) = \text{sgn}(c), \\ 0 & \text{otherwise,} \end{cases} \tag{80}$$

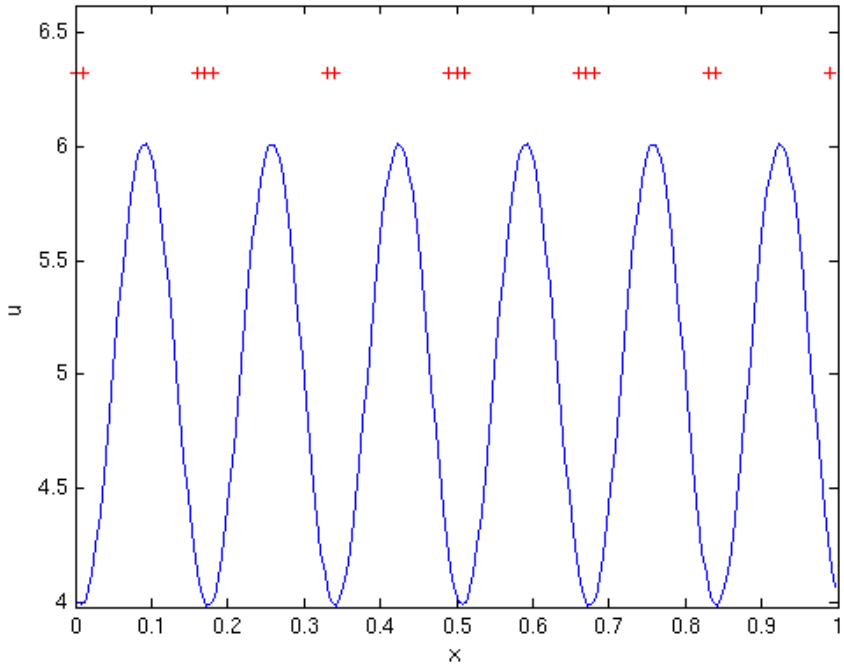

Figure 2: Solution of (67) with initial condition (79) after 100 time steps.

where sgn(a) naturally stands for the sign of a. Next we need to take values for a, b and c. We take

$$\tilde{u}_j^l = \text{minmod}(u_j^l, u_{j+1}^{l-1} - u_j^{l-1}, u_j^{l-1} - u_{j-1}^{l-1}). \tag{81}$$

We start in each cell at $l = k$ and stop the limiting procedure if either $\tilde{u}_j^l = u_j^l$ or $l = 0$.

3.5.1 Overview

We give a brief overview of how the different methods in the chapter are used. We present this in a pseudo-code algorithm as in algorithm 3.1. In this algorithm the solution u is stored as a matrix. The columns indicate the values for the cells. The rows indicate the coefficients of the different orders. By $u(i,:)$ we mean all coefficients of element i.

Algorithm 3.1 Pseudo-code algorithm for 1D shock detection and limiting

for $j = 1$ to n **do**
  $u(0,:) = u(n,:)$
  $u(n+1,:) = u(1,:)$
  $u_j = 0$;
  $u_{j-1} = 0$;
  for $l = 0$ to order **do**
    $u_j = u_j + (-1)^l u(j,l)$
    $u_{j-1} = u_{j-1} + u(j-1,l)$
  end for
  $normu_j = \text{ norm}(u(j,:))$
  $straal = (h/2)^{((order+1)/2)}$
  $I_j = \text{ abs}(u_j - u_{j-1})/(straal \cdot normu_j)$
  if $I_j > 1$ **then**
    for $l =$ order to 0 **do**
      $repl = \text{ minmod}(u(j,l), u(j+1,l-1) - u(j,l-1), u(j,l-1) - u(j-1,l-1))$
      if $repl \neq u(j,l)$ **then**
        $u(j,l) = repl$
      else
        break
      end if
    end for
  end if
end for

4 Results

4.1 1D scalar advection

We implement the one dimensional scalar advection equation using 100 elements. We take the initial condition to be

$$u(x,0) = 5 + \sin(2\pi x). \tag{82}$$

In Figure 3 we see the plots of our results. The first plot is a plot of the initial condition and the second plot shows the result after 100 time steps. The solid blue line indicates the approximated solution using the finite element method. The red dotted line shows the exact solution.

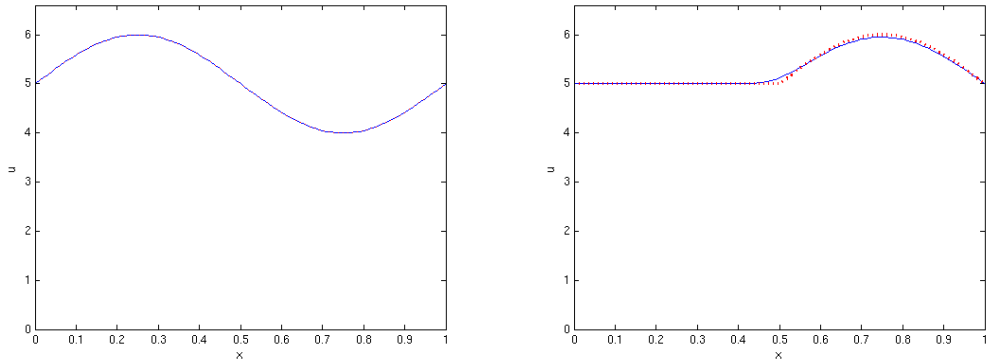

Figure 3: Initial condition (82) (left) and solution of (6) after 100 time steps.

It is quite common to get a discontinuous initial condition. To investigate what our method does with a discontinuous initial condition we take the following

$$u(x,0) = \begin{cases} 5, & x <= 0.5, \\ 0, & x > 0.5. \end{cases} \tag{83}$$

In Figure 4 we see the plots of the results. The first plot is again just a plot of the initial condition. The second plot shows the exact and approximated solution after 10 time steps. We clearly see that the approximation produces wiggles and smearing.
To cope with this discontinuity we look at the SUPG method in Figure 5. We see that we get only one peak and less smearing demonstrating the effectiveness of the SUPG method. A disadvantage of the SUPG method is that this method is applied on the whole mesh and not only where a discontinuity arises. This leads to more numerical diffusion then necessary, as illustrated in Figure 6 this is illustrated. The SUPG with a perturbation is shown in Figures 7 and 8. In the first figure we applied a perturbation as in equation (12) with initial condition (82) and ran it for 100 time steps. We applied in the right plot the SUPG method as well. Note that applying SUPG is not usefull because there is no difference with not applying SUPG. In Figure 8 we repeated this test but now with initial condition (83) and only run for 20 time steps. We now see two results. First of all, the slope is somewhat less smeared out when applying SUPG. But the kink is less rugged.

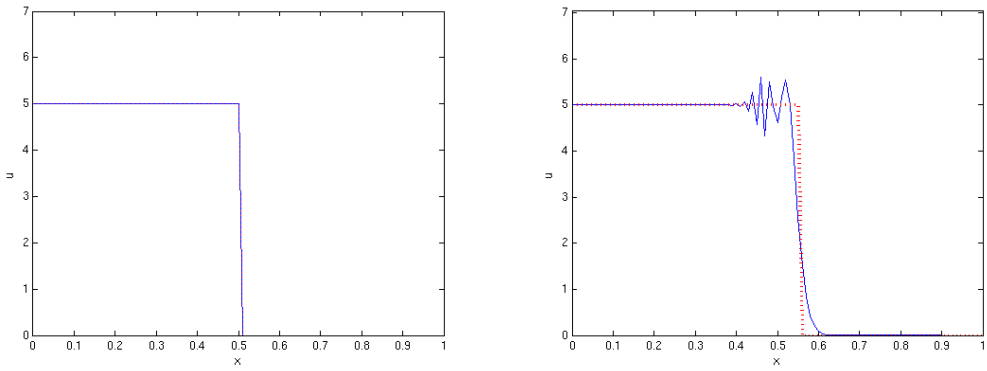

Figure 4: Initial condition (83) (left) and solution of (6) after 10 time steps.

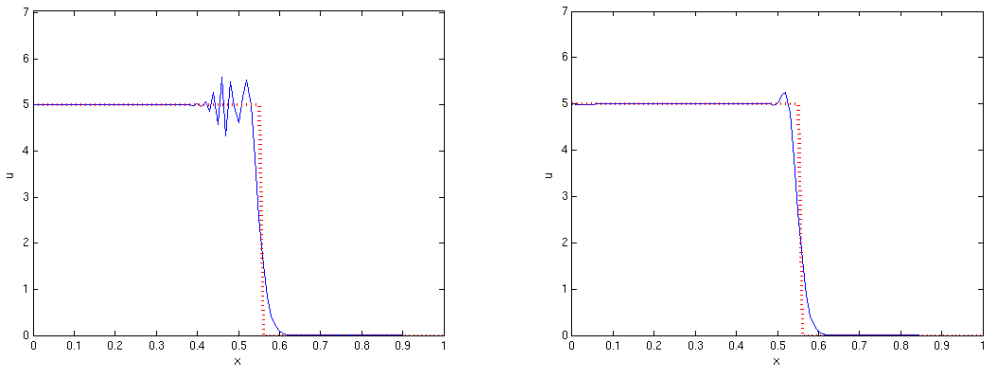

Figure 5: Solution of (6) with initial condition (83) after 10 time steps without and with SUPG.

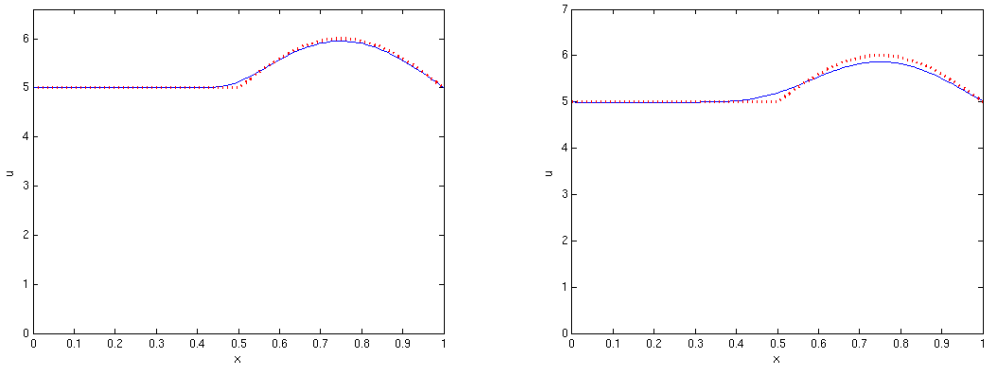

Figure 6: Solution of (6) with initial condition (82) after 100 time steps without and with SUPG.

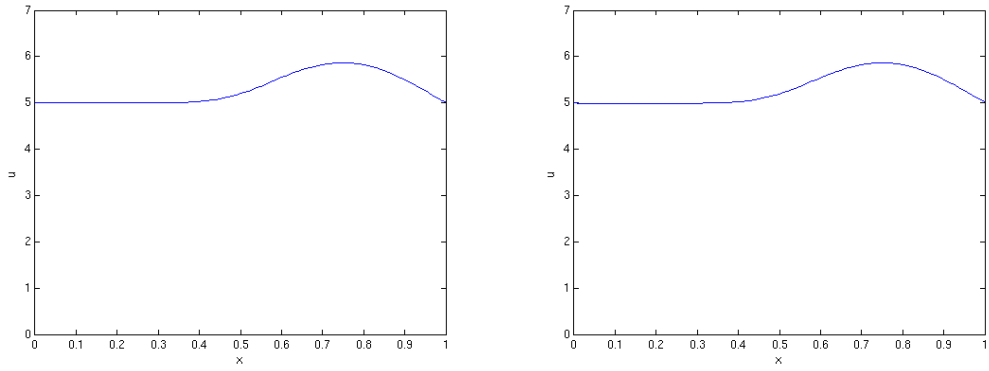

Figure 7: Solution of (12) with initial condition (82) after 100 time steps without and with SUPG.

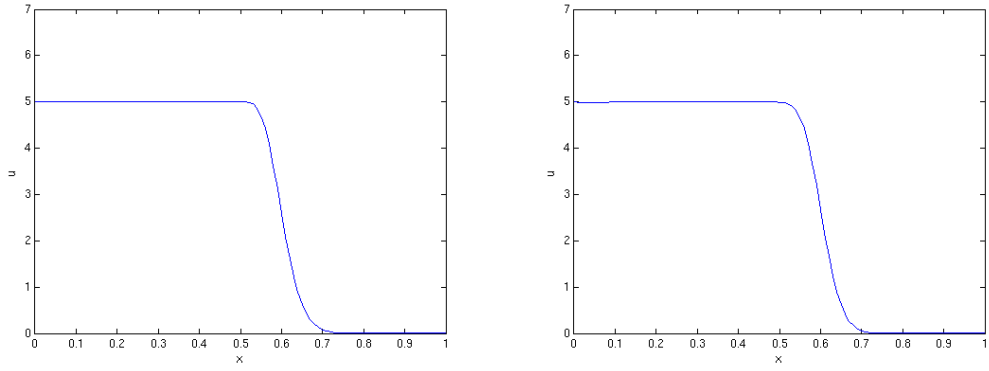

Figure 8: Solution of (12) with initial condition (83) after 20 time steps without and with SUPG.

Finally in Figure 9 we see the result of applying a variable speed. We needed 500 time steps to get some clear results. This can be explained by the fact that the final time step is small, 0.01838.

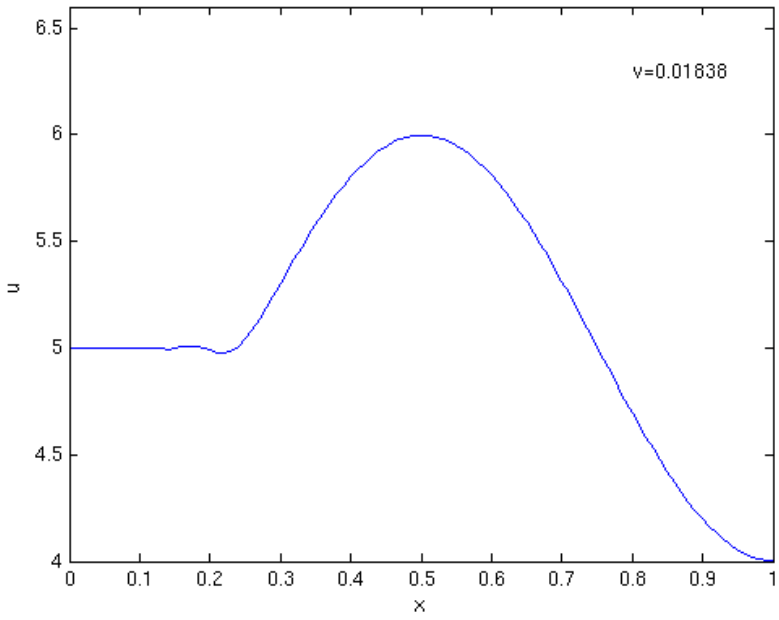

Figure 9: Solution of (22) with initial condition (82) after 500 time steps.

4.2 2D scalar advection

We implement the two dimensional scalar advection equation using 30 elements in each direction so in total this leads to 1800 elements. The initial condition is given as

$$u(x, y, 0) = 5 + \sin(2\pi x)\sin(2\pi y). \tag{84}$$

In figure 10 we plot the results. The first plot is a plot of the initial condition and the second plot shows the result after 100 time steps. It looks like the solution is being blown up. However this is caused by the angle at which we look at the surface plot.

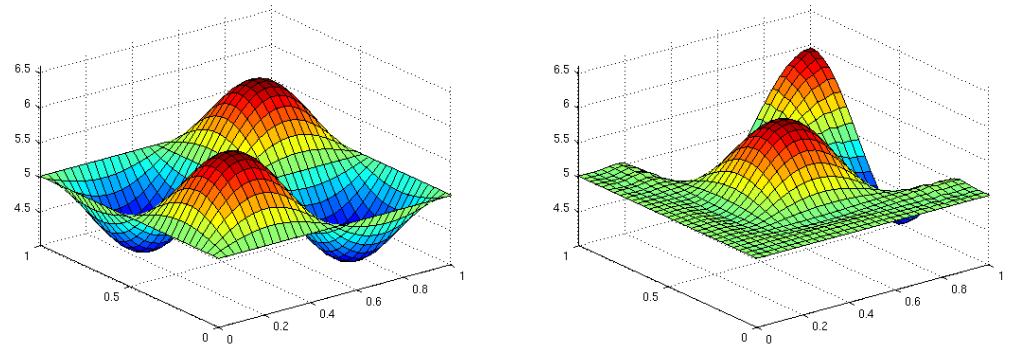

Figure 10: Initial condition (84) (left) and solution of (38) after 50 time steps.

As in the one dimensional case, we look at what happens when we have a discontinuous initial condition. We take

$$u(x, y, 0) = \begin{cases} 5, & x + y <= 1, \\ 0, & \text{otherwise.} \end{cases} \tag{85}$$

Figure 11 shows the results where we used initial condition (85) and took 20 time steps. We

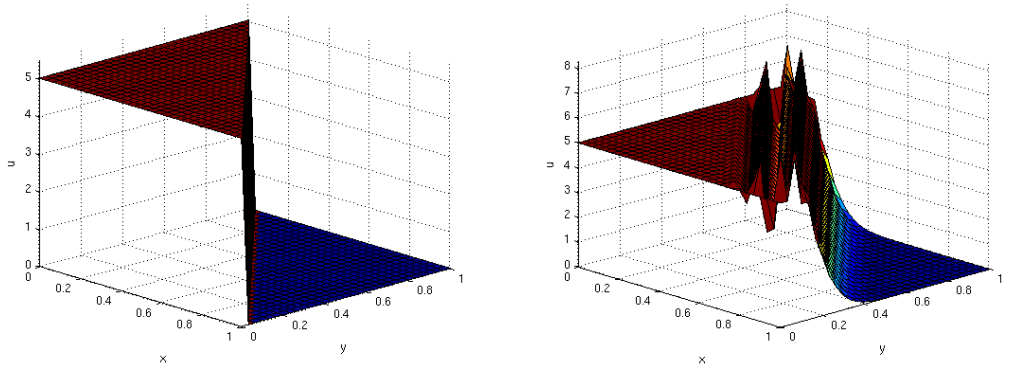

Figure 11: Initial condition (85) (left) and solution of (38) after 20 time steps.

see in this figure that, as in the one dimensional case, the approximation produces wiggles. Therefore we apply the SUPG method, see Figure 12. As in the one dimensional case we see less smearing and still one small peak, but no more wiggles. Since we can't make a clear

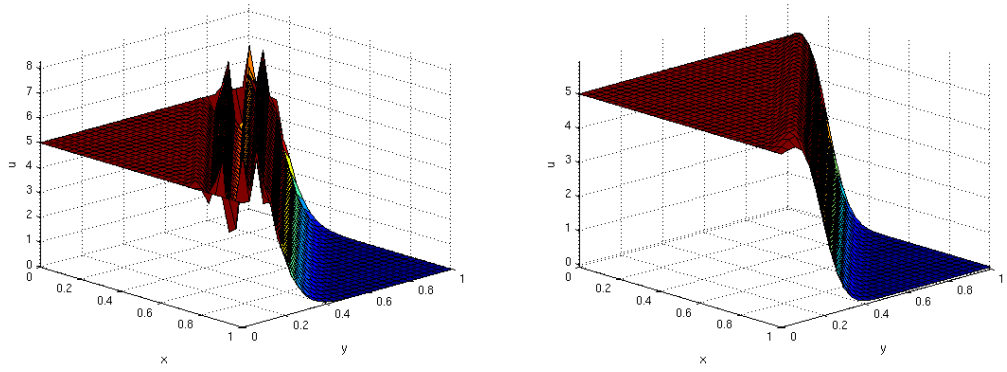

Figure 12: Solution of (38) with initial condition (85) after 20 time steps without and with SUPG.

surface plot of the approximated and exact solution in one plot we will show the results of applying a small perturbation without comparison. Figure 13 shows the result of applying a small perturbation using the SUPG method. We see that we get more smearing than in figure 10.

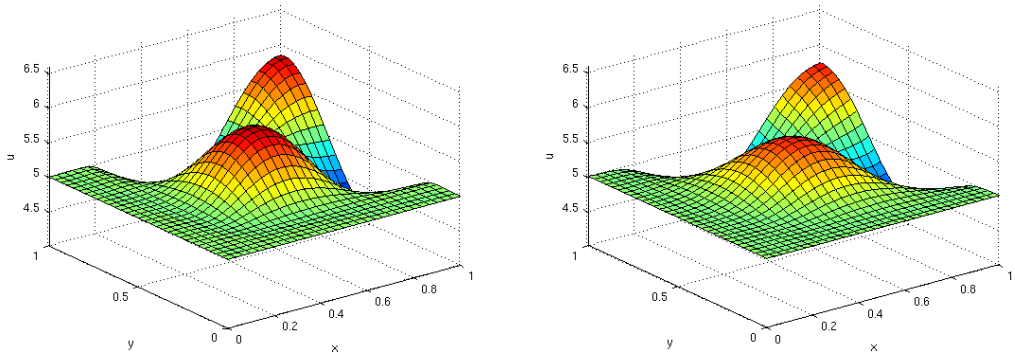

Figure 13: Solution of (44) with initial condition (85) after 50 time steps without and with SUPG.

4.3 DG approach

We apply this method to the simple scalar conservation law as in equation (67) and view the two separate cases of section 3.4. For the discontinuous initial condition we see the results in Figure 14. We see that the limiter works quite well. It is expected that with this limiter

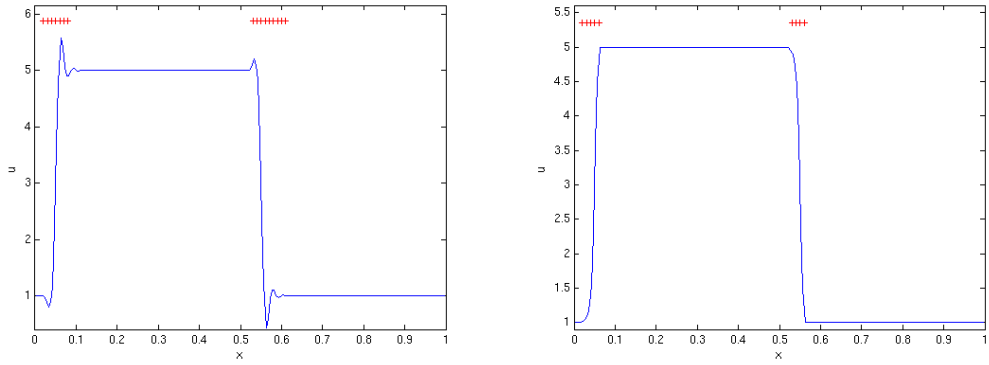

Figure 14: Solution of (67) with initial condition (78) after 100 time steps without and with limiting.

there will still be some smearing, but the wiggles are completely gone and the smearing is within the boundaries. Figure 15 illustrates nicely the drawback of this method. In this figure we clearly see that the bottom peaks are limited and therefore these peaks look somewhat flattened out.

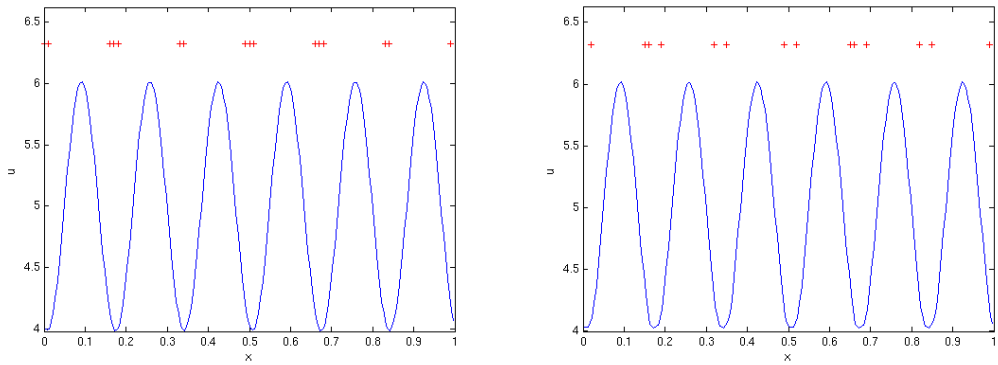

Figure 15: Solution of (67) with initial condition (79) after 100 time steps without and with limiting.

5 Conclusions and further research

In section 4 we have seen the results of applying SUPG using finite elements and a limiter and slope detector paired with the discontinuous Galerkin method. Although we know how we can apply finite elements in two dimensions, we do not know this for discontinuous Galerkin yet. One first research direction lies in applying disconinuous Galerkin to two dimensions. Furthermore, we have seen in Figure 2 that the slope detector sometimes wrongly indicates steep slopes as discontinuities. It would be nice if we could reduce occurence of this.

Additionally, we need to start looking at the model for the waterflow and how to apply the methods derived in this literature study. It would also be advantageous to obtain some realistic data for this model and examining the configuration of the implementation in the building of the EWI faculty.

In summary, this project will look at the following directions in the remainder of this master thesis.

- Expand discontinuous Galerkin to two dimensions;
- Improve the slope detector to treat steep slopes discontinuities;
- Derive the model for the water flow;
- Apply the derived methods to this model;
- Obtain realistic data for the model;
- Improve the implementation;

References

[1] Bernardo Cockburn. An introduction to the discontinuous galerkin method for convection-dominated problems. Technical report, SChool of Mathematics, University of Minnesota, Minneapolis, 1997.

[2] F. Vermolen J. van Kan, A. Segal. *Numerical Methods in Scientific Computation.* VSSD, 2008.

[3] Lilia Krivodonova. *Limiters for high-order discontinuous Galerkin methods.* Elsevier B.V., 2007.

[4] N. Chevaugeon J.E. Flaherty L. Krivodonova, J. Xin. *Shock detection and limiting with discontinuous Galerkin methods for hyperbolic conservation laws.* Elsevier B.V., 2003.