Eidgenössische
Technische Hochschule
Zürich

Ecole polytechnique fédérale de Zurich
Politecnico federale di Zurigo
Swiss Federal Institute of Technology Zurich

Multidimensional Scheme for the Shallow Water Equations. [1]

A.-T. Morel

Research Report No. 96-08
July 1996

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

[1] To appear in "Proceedings of the Hydrodynamics 96 Conference", Zürich, Sept. 9-13, 1996.

Multidimensional Scheme for the Shallow Water Equations. [1]

A.-T. Morel

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Research Report No. 96-08 July 1996

Abstract

The Method of Transport was originally developed for the Euler equation in 1993 by M. Fey. He introduced the physical property of infinitely many propagation directions into the numerical method. Here, we present the extension of this method to equations with inhomogeneous fluxes, such as the shallow water equations. For efficiency reasons and to reach higher order accuracy certain modifications had to be made to the method, whereby the multidimensional character will be kept. The resulting scheme can then be interpreted as a decomposition of the nonlinear equations into a system of linear advection equations with variable coefficients in conservative form.

Keywords: Shallow water equations, multidimensional schemes, method of transport, second order, correction terms.

Subject Classification: AMS(MOS) subject classifications (1991): 65M06, 65M25, 35L65, 35L70, 76M25.

[1] To appear in "Proceedings of the Hydrodynamics 96 Conference", Zürich, Sept. 9-13, 1996.

1 Introduction

The two-dimensional shallow water equations in conservation form read

$$\underline{U}_t + \nabla \cdot \underline{\underline{\mathcal{F}}} = 0, \qquad (1)$$

with

$$\underline{U} = \begin{pmatrix} h \\ h\,\underline{u} \end{pmatrix}$$

the state vector, where h is the total depth of the fluid and $\underline{u} = (u, v)^T$ the velocity vector. The divergence acts on the rows of the flux matrix $\underline{\underline{\mathcal{F}}}$ given by

$$\underline{\underline{\mathcal{F}}} = \underline{U}\,\underline{u}^T + \frac{h\,c^2}{2}\begin{pmatrix} \underline{0}^T \\ \underline{\underline{I}} \end{pmatrix},$$

where $c = \sqrt{g\,h}$ the celerity with g the constant of gravity and $\underline{\underline{I}}$ is the 2×2 identity matrix.

In Section 2, we present the main idea of the Method of Transport (MoT) for the shallow water equations. In Section 3, we derive a different formulation of the shallow water equations that indicates the possible decomposition. Error analysis shows that this system can be approximated to any order of accuracy by a number of linear advection equations in conservative form which can be solved independently. The idea of transport can also be applied to this type of equations. The extension to a high order scheme follows in a natural way as shown in Section 4. In Section 5, we present numerical results obtained with the developed scheme for free surface flow problem.

2 First Order Method of Transport

The Method of Transport is a finite volume method, where the update to the new timestep is done by adding incoming and subtracting outgoing flows with all the neighboring cells. The final scheme in conservation form reads

$$\underline{U}^{n+1}_{\Omega_i} = \underline{U}^n_{\Omega_i} - \frac{1}{|\Omega_i|}\sum_{j \neq i}(\underline{F}_{\Omega_i\Omega_j} - \underline{F}_{\Omega_j\Omega_i}), \qquad (2)$$

where $|\Omega_i|$ is the area of the cell. The contributions $\underline{F}_{\Omega_i\Omega_j}$ represent the quantity of information which flows from domain Ω_i into domain Ω_j. The contributions $\underline{F}_{\Omega_i\Omega_j}$ approximate the physical multidimensional flux $\underline{\underline{\mathcal{F}}}$.

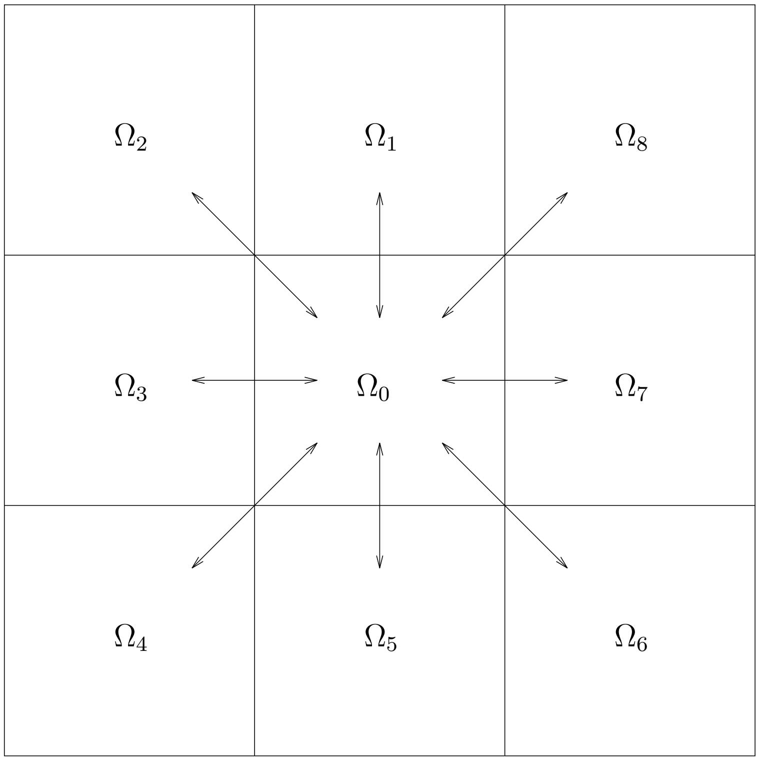

Figure 1: Interaction between the cell Ω_0 and its neighboring cells for a Cartesian grid.

In the case of the shallow water equations the flux is not homogeneous, i.e. $\underline{\underline{\mathcal{F}}}(\lambda\,\underline{U}) \neq \lambda\underline{\underline{\mathcal{F}}}(\underline{U})$, for all $\lambda \in \mathrm{I\!R}$. Nevertheless, it is possible to decompose the flux function into a linear combination of waves which are propagating with their corresponding characteristic speeds

$$\underline{\underline{\mathcal{F}}}\,\underline{n} = \sum_{i=1}^{3} \lambda_i\,\alpha_i\,\underline{r}_i.$$

The α_i are the amplitudes of the waves propagating with speed λ_i, the eigenvalues of the Jacobian of $\underline{\underline{\mathcal{F}}}\,\underline{n}$. They are given by

$$\begin{aligned} \lambda_{1,3} &= \underline{u}\cdot\underline{n} \pm c, \\ \lambda_2 &= \underline{u}\cdot\underline{n}, \end{aligned}$$

where $\underline{n} = (n_1, n_2)^T$ is a unit vector. It follows from the inhomogeneity of the flux that the vectors $\underline{r}_i$ can not be the eigenvectors of the Jacobian as is

the case of the Euler equations. Note that the vectors $\alpha_i \underline{r}_i$ have no direct physical meaning. They represent neither a shock or a rarefaction wave. Hence, we look for some vectors $\underline{r}_i$ which are the eigenvectors of a matrix $\underline{\underline{M}}$ similar to the Jacobian and satisfying $\underline{\underline{M}}\,\underline{U} = \underline{\underline{\mathcal{F}}}(\underline{U})\,\underline{n}$.
Guided by the process used for the Euler equations [2], we define

$$\underline{R}_1 = \alpha_1\,\underline{r}_1 + \alpha_3\,\underline{r}_3 = h \begin{pmatrix} 1 \\ \underline{u} \end{pmatrix} \tag{3}$$

and

$$\underline{\underline{L}}\,\underline{n} = \alpha_1\,\underline{r}_1 - \alpha_3\,\underline{r}_3 = \frac{h\,c^2}{2} \begin{pmatrix} \underline{0}^T \\ \underline{\underline{I}} \end{pmatrix} \underline{n}. \tag{4}$$

Note that $\underline{\underline{L}}$ is a $(N+1)\times N$ matrix, with N the space dimension in our case 2. It turns out that $\underline{R}_2 = \alpha_2\,\underline{r}_2 = \underline{0}$ vanishes completely, i.e. there is no advection wave for the shallow water equations. Notice that the state vector $\underline{U}$ can be rewritten as

$$\underline{U} = \sum_{i=1}^{N+1} \alpha_i\,\underline{r}_i = \underline{R}_1.$$

We have decomposed $\underline{U}$ and $\underline{\underline{\mathcal{F}}}$ in a set of waves, which leads to a flux-vector splitting in one dimension for an inhomogeneous flux.

2.1 Contributions $\underline{F}_{\Omega_i\Omega_j}$

With the help of the coefficients $\underline{R}_1$, and $\underline{\underline{L}}$ the contributions $\underline{F}_{\Omega_i\Omega_j}$ can be split into two parts, corresponding to the $\underline{\mathcal{C}}^+$ and $\underline{\mathcal{C}}^-$ waves

$$\underline{F}_{\Omega_i\Omega_j} = \underline{F}^{c^+}_{\Omega_i\Omega_j} + \underline{F}^{c^-}_{\Omega_i\Omega_j}. \tag{5}$$

Each component can be written as the integrals of the waves generated by domain Ω_i into the domain Ω_j. The first contribution $\underline{F}^{c^+}_{\Omega_i\Omega_j}$ is represented by

$$\underline{F}^{c^+}_{\Omega_i\Omega_j} = \int_{\Omega_j} \underline{\mathcal{C}}^+_{\Omega_i}(\underline{x}, t_0 + \Delta t)\,d\underline{x}.$$

The wave $\underline{\mathcal{C}}^+$ describes the propagation of quantities $\underline{R}_1(\underline{U}(\underline{y}, t_0))$ with velocity $\underline{u} + c\,\underline{n}$. This propagation can be interpreted as an infinity of advections. The sum of all these advections is described by the integral over the unit sphere $S \in \mathbb{R}^N$. The wave $\underline{\mathcal{C}}^+$ is defined with help of the vector function

$$\underline{g}(\underline{y}, t_0, \underline{n}, \Delta t) = \underline{y} + \Delta t\,(\underline{u}(\underline{y}, t_0) + c(\underline{y}, t_0)\,\underline{n})$$

and the Dirac's delta distribution as

$$\underline{\mathcal{C}}^{+}_{\Omega_i}(\underline{x}, t_0 + \Delta t) = \frac{1}{|S|} \int_S \int_{\Omega_i} \underline{R}_1(\underline{U}(\underline{y}, t_0))\, \delta(\underline{x} - \underline{g}(\underline{y}, t_0, \underline{n}, \Delta t))\, d\underline{y}\, dS.$$

$|S|$ is the area of the sphere, dS is an area element and $\underline{n}$ is the outer normal to element dS with unit length. The factor $1/|S|$ has a normalization role, such that the quantity $\underline{R}_1$ distributed with the wave $\underline{\mathcal{C}}^+$ is invariant during the time evolution.
The second contribution has the form

$$\underline{F}^{c^-}_{\Omega_i \Omega_j} = \int_{\Omega_j} \underline{\mathcal{C}}^{-}_{\Omega_i}(\underline{x}, t_0 + \Delta t)\, d\underline{x}.$$

The wave $\underline{\mathcal{C}}^-$ is similar to the wave $\underline{\mathcal{C}}^+$. It represents the propagation of the quantity $\underline{\underline{L}}(\underline{U}(\underline{y}, t_0))\, \underline{n}$ with velocity $\underline{u} + c\, \underline{n}$

$$\underline{\mathcal{C}}^{-}_{\Omega_i}(\underline{x}, t_0 + \Delta t) = \frac{N}{|S|} \int_S \int_{\Omega_i} \underline{\underline{L}}(\underline{U}(\underline{y}, t_0))\, \underline{n}\, \delta(\underline{x} - \underline{g}(\underline{y}, t_0, \underline{n}, \Delta t))\, d\underline{y}\, dS.$$

The normalization factor N depends on the dimension of the space.
The above description of the contributions has to be applied on a discrete mesh. If $\underline{U}$ is assumed to be constant in each cell, this leads to the Method of Transport with infinitely many advection directions (MoT$_\infty$). For more details see [2].

2.2 Method of Transport Simple

In the MoT$_\infty$ the functions $\underline{\mathcal{C}}^+$ and $\underline{\mathcal{C}}^-$ are complicated and hence, computing the integrals for the contributions is very time consuming. A simplified version of the MoT$_\infty$ approximates the support of $\underline{\mathcal{C}}^+$ and $\underline{\mathcal{C}}^-$ by a rectangular and the functions by piecewise constant on rectangular subdomains. This method is called Method of Transport simple. The contributions $\underline{F}_{\Omega_i \Omega_j}$ have to be computed with

$$\underline{\mathcal{C}}^+(\underline{x}, t_0 + \Delta t) = \underline{R}_1(\underline{U}^n_{\Omega_i}) f^{c^+}(\underline{x}, t_0 + \Delta t)$$

$$\underline{\mathcal{C}}^-(\underline{x}, t_0 + \Delta t) = \underline{\underline{L}}(\underline{U}^n_{\Omega_i})\, \underline{f}^{c^-}(\underline{x}, t_0 + \Delta t).$$

The supports of the functions f^{c^+} and $\underline{f}^{c^-}$ and their partitions into subdomains are shown in Figure 2.

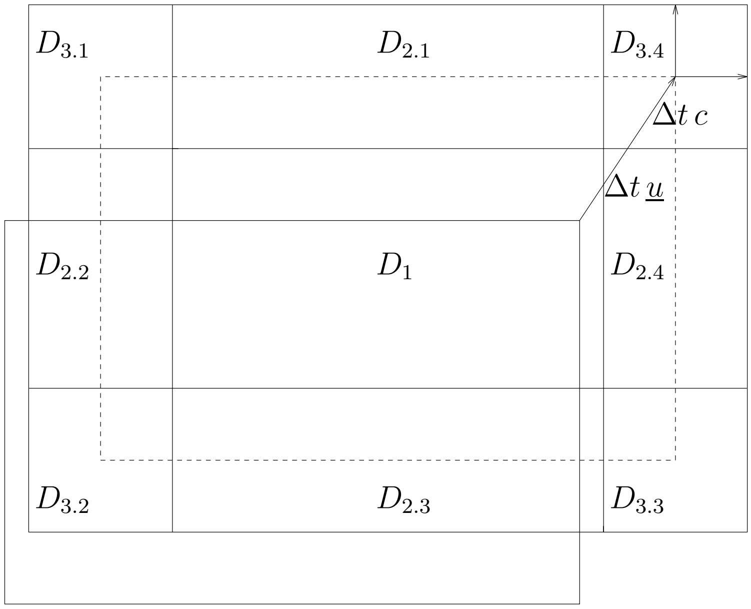

Figure 2: Decomposition of the support of the function f^{c^+} and $\underline{f}^{c^-}$.

Using the notation $D_i = \cup_j D_{i.j}$ the piecewise constant function f^{c^+} and the vector $\underline{f}^{c^-} = (f_1^{c^-}, f_2^{c^-})^T$ are given by

$$f^{c^+}(\underline{x}, t_0 + \Delta t) = \begin{cases} 1 \text{ if } \underline{x} \in D_1 \\ 1/2 \text{ if } \underline{x} \in D_2 \\ 1/4 \text{ if } \underline{x} \in D_3 \\ 0 \text{ elsewhere} \end{cases},$$

$$f_1^{c^-}(\underline{x}, t_0 + \Delta t) = \begin{cases} 1/2 \text{ if } \underline{x} \in D_{2.4} \\ -1/2 \text{ if } \underline{x} \in D_{2.2} \\ 1/4 \text{ if } \underline{x} \in D_{3.3} \cup D_{3.4} \\ -1/4 \text{ if } \underline{x} \in D_{3.1} \cup D_{3.2} \\ 0 \text{ elsewhere} \end{cases},$$

and analog for $f_2^{c^-}$.
The function f^{c^+} can be interpreted as the sum of four translations of the original cell. The translations can be described as

$$f^a(\underline{x}, \underline{a}, t_0 + \Delta t) = \int_{\Omega_i} \delta(\underline{y} + \Delta t\, \underline{a}(\underline{y}, t_0) - \underline{x})\, d\underline{y}$$

and

$$f^{c^+}(\underline{x}, t_0 + \Delta t) = \frac{1}{4} \sum_{i=1}^{4} f^a(\underline{x}, \underline{u} + c\,\underline{\tilde{n}}_i, t_0 + \Delta t)$$

with

$$\underline{\tilde{n}}_i \in \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$$

for $i = 1, \ldots, 4$. The function $\underline{f}^{c^-}$ can also be written with the help of f^a as

$$\underline{f}^{c^-}(\underline{x}, t_0 + \Delta t) = \frac{1}{4} \sum_{i=1}^{4} \underline{\tilde{n}}_i\, f^a(\underline{x}, \underline{u} + c\,\underline{\tilde{n}}_i, t_0 + \Delta t).$$

The contribution $\underline{F}_{\Omega_i \Omega_j}$ can then be computed as

$$\underline{F}_{\Omega_i \Omega_j} = \int_{\Omega_j} \frac{1}{4} \sum_{i=1}^{4} (\underline{R}_1 + \underline{\underline{L}}\,\underline{\tilde{n}}_i)\, f^a(\underline{x}, \underline{u} + c\,\underline{\tilde{n}}_i, t_0 + \Delta t)\, d\underline{x}. \tag{6}$$

This decomposition will be further developed to reach second order.

3 Decomposition of the Equations

In Section 2, the contributions are decomposed into two waves, $\underline{\mathcal{C}}^+$ and $\underline{\mathcal{C}}^-$. These waves are related to critical waves. It is the aim of this section to decompose the shallow water equations in a similar fashion. In [4] the decomposition is done for the Euler equations.

3.1 Decomposition in Infinitely Many Advection Equations

Using the coefficients $\underline{R}_1$ and $\underline{\underline{L}}$ from (3) and (4), $\underline{\underline{\mathcal{F}}}$ can be written as

$$\underline{\underline{\mathcal{F}}}(\underline{U}) = \underline{R}_1\, \underline{u}^T + c\,\underline{\underline{L}}.$$

The propagation of the quantity $\underline{R}_1$ with the velocity $\underline{u} + c\,\underline{n}$ is a translation by $\underline{u}$ combined with an expansion $c\,\underline{n}$. For each $\underline{n}$ we can interpret the behavior of $\underline{R}_1$ as a transport process described by

$$\underline{\phi}_1(\underline{n}) := (\underline{R}_1)_t + \nabla \cdot (\underline{R}_1 (\underline{u} + c\,\underline{n})^T).$$

Since the critical waves move in all directions, we have to split $\underline{R}_1$ and propagate it in all directions. With the identities

$$\frac{1}{|S|}\int_S dS = 1 \quad \text{and} \quad \frac{1}{|S|}\int_S \underline{n}\, dS = 0,$$

where S is the unit sphere in $\mathbb{R}^N$ and $|S|$ its area, and the state vector $\underline{U}$ rewritten as

$$\underline{U} = \frac{1}{|S|}\int_S \underline{R}_1\, dS,$$

the integral of $\underline{\phi}_1(\underline{n})$ over the unit sphere becomes

$$\frac{1}{|S|}\int_S \underline{\phi}_1(\underline{n})\, dS = \underline{U}_t + \nabla\cdot(\underline{U}\,\underline{u}^T) = 0.$$

However, this is not the left-hand side of the shallow water equations. The missing term in the flux matrix can be associated with the $\underline{\mathcal{C}}^-$ wave. The vector $\underline{\underline{L}}\,\underline{n}$ is also transported with the velocity $\underline{u} + c\,\underline{n}$. The corresponding transport terms are

$$\underline{\phi}_2(\underline{n}) := (\underline{\underline{L}}\,\underline{n})_t + \nabla\cdot(\underline{\underline{L}}\,\underline{n}(\underline{u} + c\,\underline{n})^T).$$

Clearly

$$\frac{1}{|S|}\int_S \underline{\underline{L}}\,\underline{n}\, dS = 0 \text{ and } \frac{1}{|S|}\int_S \underline{n}\,\underline{n}^T\, dS = \frac{1}{N}\underline{\underline{I}}.$$

To get consistency with the shallow water equations (1), we take N times $\underline{\phi}_2$. Then the equations

$$\frac{1}{|S|}\int_S \underline{\phi}_a(\underline{n})\, dS = \underline{U}_t + \nabla\cdot\underline{\underline{\mathcal{F}}} = 0, \tag{7}$$

with

$$\underline{\phi}_a(\underline{n}) := \underline{\phi}_1(\underline{n}) + N\underline{\phi}_2(\underline{n})$$

recover the original nonlinear system. $\underline{\phi}_a(\underline{n})$ is the combination of the $\underline{\mathcal{C}}^+$ and $\underline{\mathcal{C}}^-$ waves. Observe that this factor N for the $\underline{\mathcal{C}}^-$ wave had already been derived to make the first order scheme consistent with the shallow water flux. Using

$$\underline{R}_a(\underline{n}) := \underline{R}_1 + N\,\underline{\underline{L}}\,\underline{n},$$

we get for $\underline{\phi}_a$

$$\underline{\phi}_a(\underline{n}) = (\underline{R}_a)_t + \nabla \cdot (\underline{R}_a(\underline{u} + c\,\underline{n})^T).$$

Hence the state vector is represented by

$$\underline{U} = \frac{1}{|S|} \int_S \underline{R}_a(\underline{n})\, dS. \tag{8}$$

In the next subsection we shall make use of this representation to create a numerical scheme. The similar decomposition for the Euler equations is described in [4].

3.2 Decomposition in Finitely Many Advection Equations

The disadvantages of the formulations (7) and (8) are that the state vector $\underline{U}$ is represented by an integral and infinitely many advection equations have to be solved. The integral will now be replaced by a finite sum of k terms. (8) becomes

$$\underline{U} = \frac{1}{k}\sum_{i=1}^{k} \underline{R}_a(\underline{n}_i) = \frac{1}{k}\sum_{i=1}^{k}\left(\underline{R}_1 + N\,\underline{\underline{L}}\,\underline{n}_i\right) \tag{9}$$

and (7)

$$\frac{1}{k}\sum_{i=1}^{k} \underline{\phi}_a(\underline{n}_i) = \underline{U}_t + \nabla \cdot \underline{\underline{F}} = 0. \tag{10}$$

In order that (9) and (10) hold exactly, the $\underline{n}_i$ have to satisfied certain conditions. From equation (9) follows

$$\sum_{i=1}^{k} \underline{n}_i = 0 \tag{11}$$

and from (10)

$$\frac{N}{k}\sum_{i=1}^{k} \underline{n}_i\,\underline{n}_i^{\,T} = \underline{\underline{I}}. \tag{12}$$

Collecting these results we can approximate the shallow water equations (10) by a combination of k advection equations if (11) and (12) holds.

The condition (11) and (12) do not define $\underline{n}_i$ uniquely. Here we consider in particular the four vectors aligned on the horizontal and vertical axis, which are

$$\underline{n}_i \in \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix} \right\} \tag{13}$$

for $i = 1, \ldots, 4$. Note that this choice of is not related to the dimensional splitting approach. In general the final propagation $\underline{u} + c\,\underline{n}_i$ is not aligned with the coordinate axes. These $\underline{n}_i$ are also a natural way to approximate the characteristic cone. The vectors $\underline{n}_i$ can be interpreted as the support points for a quadrature rule to integrate the characteristic cone.

The choice of the $\underline{n}_i$ influences the functions f^a used in (6). Here we will check the union of their supports, shown in Figure 3, which have to approximate the support of the exact waves $\underline{\mathcal{C}}^+$ and $\underline{\mathcal{C}}^-$.

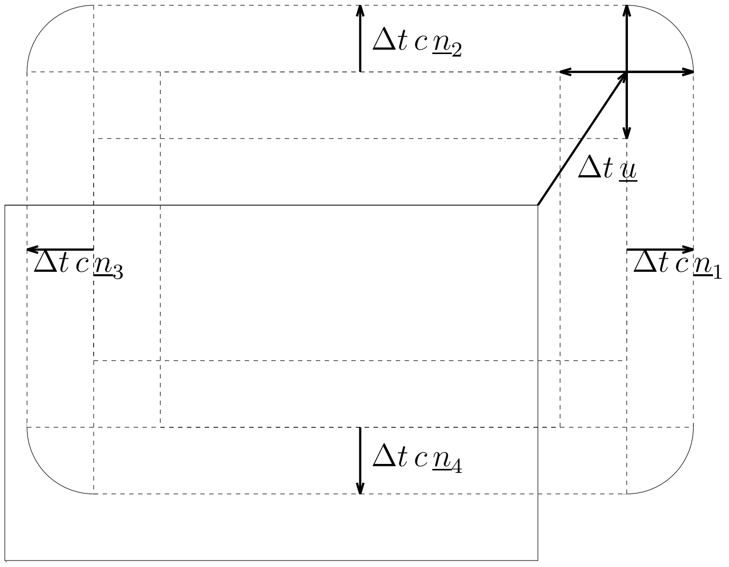
$\Delta t\, c\, \underline{n}_2$
$\Delta t\, \underline{u}$
$\Delta t\, c\, \underline{n}_3$
$\Delta t\, c\, \underline{n}_1$
$\Delta t\, c\, \underline{n}_4$

Figure 3: Support of the contributions $\underline{F}_{\Omega_i\Omega_j}$ for the unit vectors defined in (13) and for the MoT$_\infty$.

We can see that the support of the waves $\underline{\mathcal{C}}^+$ and $\underline{\mathcal{C}}^-$ does not completely overlap with the one of the MoT$_\infty$. The corners are not recovered.

Another possible choice is given by the unit vectors lying on the diagonal, but in this case strips along the edges of the exact support are not recovered.

This problem can be improved by replacing the vectors $\underline{n}_i$ by the $\underline{\tilde{n}}_i$

$$\underline{\tilde{n}}_i \in \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\} \tag{14}$$

for $i = 1, \ldots, 4$. The support generated by these $\underline{\tilde{n}}_i$ is identical to the support of the functions f^{c^+} and $\underline{f}^{c^-}$ of the Method of Transport simple, (see Figure 2), but the $\underline{\tilde{n}}_i$ do not satisfy (12) since they are not unit vectors.
We already introduced the scaling factor N to set consistency in (12). If we defined $\underline{R}_a$ in a more general way as

$$\underline{\tilde{R}}_a(\underline{\tilde{n}}_i) := \underline{R}_1 + \omega_i \, \underline{\underline{L}} \, \underline{\tilde{n}}_i, \tag{15}$$

where

$$\omega_i := \frac{N}{\underline{\tilde{n}}_i^T \underline{\tilde{n}}_i},$$

we get exactly the shallow water equations

$$\frac{1}{k} \sum_{i=1}^{k} \underline{\tilde{\phi}}_a(\underline{\tilde{n}}_i) = \underline{U}_t + \nabla \cdot \underline{\underline{F}} = 0, \tag{16}$$

$\underline{\phi}_a$ becomes

$$\underline{\tilde{\phi}}_a(\underline{\tilde{n}}_i) := (\underline{\tilde{R}}_a(\underline{\tilde{n}}_i))_t + \nabla \cdot (\underline{\tilde{R}}_a(\underline{\tilde{n}}_i) \, (\underline{u} + c \, \underline{\tilde{n}}_i)^T). \tag{17}$$

3.3 High Order Resolution

To solve the equations (16) for one time step, we linearize $\underline{\tilde{\phi}}_a(\underline{\tilde{n}}_i)$ and set to zero each component of the sum. The decomposition process now becomes obvious. At a given time, t_0, we eliminate the time dependency of $\underline{u}$ and c by freezing the time so that

$$\underline{a}(\underline{x}, \underline{\tilde{n}}_i) := \underline{u}(\underline{U}(\underline{x}, t_0)) + \underline{\tilde{n}}_i \, c(\underline{U}(\underline{x}, t_0))$$

becomes a function of $\underline{x}$ only. Thus, we obtain a set of linear advection equations of the form

$$\underline{\tilde{\tilde{\phi}}}_a(\underline{\tilde{n}}_i) := (\underline{\tilde{R}}_a(\underline{\tilde{n}}_i))_t + \nabla \cdot (\underline{\tilde{R}}_a(\underline{\tilde{n}}_i) \, \underline{a}(\underline{x}, \underline{\tilde{n}}_i)^T) = 0. \tag{18}$$

Summing up the solutions of (18) for $i = 1, \ldots, k$ leads to an approximate solution of (16):

$$\frac{1}{k}\sum_{i=1}^{k} \underline{\widetilde{\widetilde{\phi}}}_a(\underline{\widetilde{n}}_i) = 0. \tag{19}$$

The time evolution of the exact solution can be approximated by the average of the solutions of the decomposed equations. For a general nonlinear system, this approximation is only of first order.
To find a more accurate approximation we replace (15) by

$$\underline{\widetilde{R}}_a(\underline{\widetilde{n}}_i) = \underline{R}_1 + \omega_i\,(\underline{\underline{L}} + \underline{\underline{K}})\,\underline{\widetilde{n}}_i \tag{20}$$

such that the solution of (19) with (20) coincides with the solution of (1) up to second order. $\underline{\underline{K}}$ is a correction matrix

$$\underline{\underline{K}}(\underline{x}, t_0) = \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \\ k_{31} & k_{32} \end{pmatrix},$$

which is determined by an error analysis. We will see that the correction terms are of order $O(\Delta t)$. It turns out that the corrected equations have the same structure as before. This idea can be generalized to a higher order method.

3.4 Correction Terms

The correction terms are computed by comparing the Taylor expansion of the solution of (1) with the Taylor expansion of (19) with (20). Here, we explicitly carry out the computation for the first component h, the same strategy has to be applied to the other components. The Taylor series of the solution of (1) at $(\underline{x}, t_0 + \Delta t)$ is

$$h(\underline{x}, t_0 + \Delta t) = h(\underline{x}, t_0) + \Delta t\, h_t(\underline{x}, t_0) + \frac{\Delta t^2}{2}\, h_{tt}(\underline{x}, t_0) + O(\Delta t^3). \tag{21}$$

From now on, we will omit the argument $(\underline{x}, t_0)$. Using the conservation of mass and momentum (1), the time derivatives in (21) can be replaced by the spatial derivatives

$$\begin{aligned} h(\underline{x}, t_0 + \Delta t) ={}& h - \Delta t\,(h\,u)_x - \Delta t\,(h\,v)_y + \frac{\Delta t^2}{2}\,((h\,u^2 + \frac{h\,c^2}{2})_x + (h\,u\,v)_y)_x \\ &+ \frac{\Delta t^2}{2}\,((h\,u\,v)_x + (h\,v^2\frac{h\,c^2}{2})_y)_y + O(\Delta t^3). \end{aligned} \tag{22}$$

Although the correction terms depend on the vectors $\underline{\widetilde{n}}_i$, they can be computed for any choice of $\underline{\widetilde{n}}_i$ that fulfill the consistency relations (11) and (12). The solution $\widetilde{h}$ of (19) is the average of the k solutions of (18). The Taylor expansion for its first component is

$$h_i(\underline{x}, t_0 + \Delta t) = h_i + \Delta t\, (h_i)_t + \frac{\Delta t^2}{2}\, (h_i)_{tt} + O(\Delta t^3), \tag{23}$$

where the derivatives are given by

$$\begin{aligned}(h_i)_t =& -((h_i + \omega_i\, k_{11}\, \widetilde{n}_{i,1} + \omega_i\, k_{12}\, \widetilde{n}_{i,2})\, (u + c\, \widetilde{n}_{i,1}))_x \\ & -((h_i + \omega_i\, k_{11}\, \widetilde{n}_{i,1} + \omega_i\, k_{12}\, \widetilde{n}_{i,2})\, (v + c\, \widetilde{n}_{i,2}))_y\end{aligned}$$

and

$$\begin{aligned}(h_i)_{tt} =& -((h_i + \omega_i\, k_{11}\, \widetilde{n}_{i,1} + \omega_i\, k_{12}\, \widetilde{n}_{i,2})\, (u + c\, \widetilde{n}_{i,1}))_{tx} \\ & -((h_i + \omega_i\, k_{11}\, \widetilde{n}_{i,1} + \omega_i\, k_{12}\, \widetilde{n}_{i,2})\, (v + c\, \widetilde{n}_{i,2}))_{ty} \\ =& -((h_i)_t\, (u + c\, \widetilde{n}_{i,1}))_x - ((h_i)_t\, (v + c\, \widetilde{n}_{i,2}))_y.\end{aligned}$$

The last equality follows from the linearity of (18), where u, v, c are functions of $\underline{x}$ only and the correction coefficients and their derivatives are first order terms in Δt, but they appear in the second derivative $(h_i)_{tt}$, with third order influences on the solution $\widetilde{h}$. $(h_i)_{tt}$ can be rewritten as

$$\begin{aligned}(h_i)_{tt} =& ((((h_i + \omega_i\, k_{11}\, \widetilde{n}_{i,1} + \omega_i\, k_{12}\, \widetilde{n}_{i,2})\, (u + c\, \widetilde{n}_{i,1}))_x \\ & + ((h_i + \omega_i\, k_{11}\, \widetilde{n}_{i,1} + \omega_i\, k_{12}\, \widetilde{n}_{i,2})\, (v + c\, \widetilde{n}_{i,2}))_y)\, (u + c\, \widetilde{n}_{i,1}))_x \\ & + ((((h_i + \omega_i\, k_{11}\, \widetilde{n}_{i,1} + \omega_i\, k_{12}\, \widetilde{n}_{i,2})\, (u + c\, \widetilde{n}_{i,1}))_x \\ & + ((h_i + \omega_i\, k_{11}\, \widetilde{n}_{i,1} + \omega_i\, k_{12}\, \widetilde{n}_{i,2})\, (v + c\, \widetilde{n}_{i,2}))_y)\, (v + c\, \widetilde{n}_{i,2}))_y.\end{aligned}$$

From the generalization of (12) for non-unit vectors follows

$$\frac{1}{k} \sum_{i=1}^{k} \omega_i\, \widetilde{n}_{i,1}^2 = \frac{1}{k} \sum_{i=1}^{k} \omega_i\, \widetilde{n}_{i,2}^2 = 1.$$

If we sum up the expansion (23), the solution $\widetilde{h}$ is

$$\begin{aligned}\widetilde{h} =& h - \Delta t\, (h\, u + k_{11}\, c)_x - \Delta t\, (h\, v + \; k_{21}\, c)_y \\ & + \frac{\Delta t^2}{2}\, (u\, (h\, u + \; k_{11}\, c)_x + c\, (\omega\, h\, c + k_{11}\, u)_x + u(h\, v + k_{12}\, c)_y + c(k_{11}\, v)_y)_x \\ & + \frac{\Delta t^2}{2}\, (v\, (h\, u + k_{11}\, c)_x + c\, (k_{12}\, u)_x + v(h\, v + k_{12}\, c)_y + c(\omega\, h\, c + k_{12}\, v)_y)_y + O(\Delta t^3)\end{aligned} \tag{24}$$

where

$$\omega = \frac{1}{k}\sum_{i=1}^{k} \underline{n}_{i,1}^2 = \frac{1}{k}\sum_{i=1}^{k} \underline{n}_{i,2}^2.$$

For all sets of unit vectors satisfying (11) and in particular for the vectors (13) , ω is always one half. For the set of vectors (14), ω is one. It is also possible to combine the two type of vectors, then the ω corresponding to these eight vectors has a value of three quarters.
We want (24) and (22) to be equal up to second order, so we have to fix k_{11} and k_{12} so that

$$\Delta t(k_{11}\, c)_x + \frac{\Delta t^2}{2}\,(h\,u\,u_x + (1 - \frac{3\,\omega}{2})\,h_x\,c^2 + h\,u_y\,v)_x = O(\Delta t^3)$$

and

$$\Delta t(k_{12}\, c)_y + \frac{\Delta t^2}{2}\,(h\,v\,v_y + (1 - \frac{3\,\omega}{2})\,h_y\,c^2 + h\,u\,v_x)_y = O(\Delta t^3).$$

Following the same argument for the moments $h\,u$ and $h\,v$, we can compute the other components and will find that the correction matrix is given by

$$\begin{aligned}
k_{11} &= -\frac{\Delta t}{2\,c}((1 - \frac{3\,\omega}{2})\,h_x\,c^2 + h\,u\,u_x + h\,u_y\,v)\\
k_{12} &= -\frac{\Delta t}{2\,c}(h\,u\,v_x + (1 - \frac{3\,\omega}{2})\,h_y\,c^2 + h\,v\,v_y)\\
k_{21} &= -\frac{\Delta t}{2\,c}((\frac{5}{4} - \frac{3\,\omega}{2})\,h_x\,c^2\,u + (\frac{1}{2} - \omega)\,h\,c^2\,u_x + h\,u^2\,u_x + \frac{1}{4}\,h_y\,c^2\,v + h\,u\,u_y\,v + \frac{1}{2}\,h\,c^2\,v_y)\\
k_{22} &= -\frac{\Delta t}{2\,c}(h\,u^2\,v_x) + (1 - \frac{3\,\omega}{2})\,h_y\,c^2\,u - \omega\,h\,c^2\,u_y + h\,u\,v\,v_y)\\
k_{31} &= -\frac{\Delta t}{2\,c}((1 - \frac{3\,\omega}{2})\,h_x\,c^2\,v + h\,u\,u_x\,v - \omega\,h\,c^2\,v_x + h\,u_y\,v^2)\\
k_{32} &= -\frac{\Delta t}{2\,c}((\frac{5}{4} - \frac{3\,\omega}{2})\,h_y\,c^2\,v + (\frac{1}{2} - \omega)\,h\,c^2\,v_y + h\,v^2\,v_y + \frac{1}{4}\,h_x\,c^2\,u + \frac{1}{2}h\,c^2\,u_x + h\,u\,v\,v_x).
\end{aligned}$$

4 Higher Order Scheme

In the previous section, we decomposed the shallow water equations in a set of linear advection equations with variable coefficients. Now we want to present a high order numerical scheme to solve these equations. Therefore, we consider the two-dimensional advection equation in conservative form

$$u_t + \nabla \cdot (u\,\underline{a}^T) = 0, \tag{25}$$

where $\underline{a} = \underline{a}(\underline{x}) = (a, b)^T$. We want to extend the multidimensional Method of Transport to higher order. For the scalar equation the scheme (2) becomes

$$u_{\Omega_i}^{n+1} = u_{\Omega_i}^{n} - \frac{1}{|\Omega_i|} \sum_{j \neq i} (F_{\Omega_i\Omega_j} - F_{\Omega_j\Omega_i}),$$

where the contributions $F_{\Omega_i\Omega_j}$ are defined as

$$F_{\Omega_i\Omega_j} = \int_{\Omega_j} \mathcal{U}(\underline{x}, t_0 + \Delta t)\, d\underline{x}.$$

The wave $\mathcal{U}$ describes the transport of u from the computational cell Ω_i to any point $\underline{x}$ in space.
The advection equation (25) can be rewritten as

$$u_t + (\nabla u) \cdot \underline{a} = -u\, (\nabla \cdot \underline{a}^T). \tag{26}$$

It follows that the evolution of u in (26) along the characteristic curve $\underline{z}(\tau)$ satisfies

$$\frac{d}{dt} u(\underline{z}(t), t) = -u\, (\nabla \cdot \underline{a}^T),$$

where $\underline{z}(\tau)$ is defined by

$$\dot{\underline{z}}(\tau) = \underline{a}(\underline{z}(\tau)), \quad \underline{z}(t) = \underline{\xi}. \tag{27}$$

We can express the transport of the quantity u along the characteristic curve $\underline{z}(t_0 + \Delta t, \underline{\xi})$, the solution of (27), by using

$$\mathcal{U}(\underline{x}, t_0 + \Delta t) = \int_{\Omega_i} u(\underline{\xi}) \delta(\underline{z}(t_0 + \Delta t, \underline{\xi}) - \underline{x})\, d\underline{\xi}.$$

The integration of the Dirac's delta distribution is not trivial, due to the nonlinear argument. Assuming the map defined in (27) to be bijective, then the variable transformation

$$\underline{v}(\underline{\xi}, \underline{x}) := \underline{z}(t_0 + \Delta t, \underline{\xi}) - \underline{x}$$

has an inverse $\underline{s}(\underline{v}, \underline{x})$, i.e. $\underline{v} \circ \underline{s} = Id$, which allows the elimination of the Dirac's delta distribution. Thus the computation of the contributions $F_{\Omega_i\Omega_j}$ becomes

$$F_{\Omega_i\Omega_j} = \int_{\Omega_j} u(\underline{s}(0, \underline{x}))\, \frac{1}{\det(J)}\, d\underline{x},$$

where $J := d\underline{z}/d\underline{\xi}$ is the Jacobian of the mapping in (27).
To get a first order approximation of the characteristic curve we take a linear reconstruction for $\underline{a}(\underline{z})$

$$\underline{a}(\underline{z}) = \underline{a} + \underline{\underline{A}}\,\underline{z}, \tag{28}$$

where the matrix $\underline{\underline{A}}$ is defined as

$$\underline{\underline{A}} = \begin{pmatrix} a_x & a_y \\ b_x & b_y \end{pmatrix}.$$

If $\underline{\underline{A}}$ is constant, the solution of the corresponding linear differential equation

$$\dot{\underline{z}}(\tau) = \underline{a} + \underline{\underline{A}}\,\underline{z}, \quad \underline{z}(t_0) = \underline{\xi}$$

is given by

$$\underline{z}(\tau) = -\underline{\underline{A}}^{-1}\,\underline{a} + e^{\underline{\underline{A}}\tau}(\underline{\underline{A}}^{-1}\,\underline{a} + \underline{\xi}).$$

By taking the Taylor expansion of $\underline{z}(\tau)$

$$\underline{z}(\tau) = \underline{\xi} + \tau\,(\underline{a} + \underline{\underline{A}}\,\underline{\xi}) + \frac{\tau^2}{2}(\underline{\underline{A}}\,\underline{a} + \underline{\underline{A}}^2\,\underline{\xi}) + O(\tau^3),$$

we get for $\underline{z}(t_0 + \Delta t, \underline{\xi})$ as

$$\underline{z}(t_0 + \Delta t, \underline{\xi}) = \underline{\xi} + \Delta t\,(\underline{a} + \underline{\underline{A}}\,\underline{\xi}) + \frac{\Delta t^2}{2}(\underline{\underline{A}}\,\underline{a} + \underline{\underline{A}}^2\,\underline{\xi}) + O(\Delta t^3). \tag{29}$$

Substituting (29) into $\underline{v}(\underline{\xi}, \underline{x})$, the equations

$$\underline{v}(\underline{\xi}, \underline{x}) = 0$$

are easily solved and the solution is

$$\underline{s}(\underline{0}, \underline{x}) = (\underline{\underline{I}} + \Delta t\,\underline{\underline{A}} + \frac{\Delta t^2}{2}\underline{\underline{A}}^2)^{-1}(\underline{x} - \Delta t\,\underline{a} - \frac{\Delta t^2}{2}\underline{\underline{A}}\,\underline{a})$$

Using the Neumann series we get as second order approximation

$$\begin{aligned} \underline{s}(\underline{0}, \underline{x}) &= (\underline{\underline{I}} - \Delta t\,\underline{\underline{A}} + \frac{\Delta t^2}{2}\underline{\underline{A}}^2)(\underline{x} - \Delta t\,\underline{a} - \frac{\Delta t^2}{2}\underline{\underline{A}}\,\underline{a}) \\ &= \underline{x} - \Delta t\,\underline{\underline{A}}\,\underline{x} - \Delta t\,\underline{a} + \frac{\Delta t^2}{2}\underline{\underline{A}}\,\underline{a} + \frac{\Delta t^2}{2}\underline{\underline{A}}^2\underline{x} + O(\Delta t^3). \end{aligned}$$

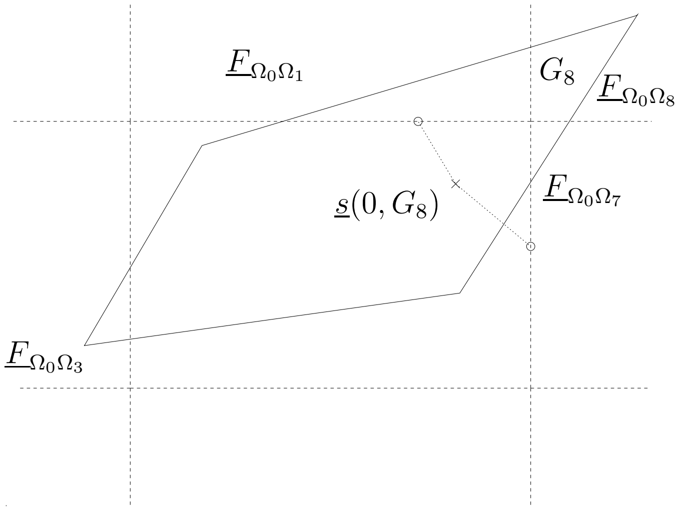
$\underline{F}_{\Omega_0\Omega_1}$
G_8
$\underline{F}_{\Omega_0\Omega_8}$
$\underline{F}_{\Omega_0\Omega_7}$
$\underline{s}(0, G_8)$
$\underline{F}_{\Omega_0\Omega_3}$

Figure 4: Sketch of transformation (28). The solid line represents the forward transformation of the original cell. The dotted line denotes the backward transformation of G_8 into the original cell.

By the variable substitution $\underline{x} = \underline{s}(0, \underline{y})$ we get for the contributions

$$F_{\Omega_i\Omega_j} = \int_{\underline{s}(0,G_j)} u(\underline{x}, t_0)\, d\underline{x},$$

where $\underline{s}(0, G_j) \subset \Omega_i$ describes the inverse of the domain G_j, which is sketched in Figure 4.

The domains of integration can be triangles, quadrilaterals, pentagons or hexagons. Therefore we reconstruct $u(\underline{x}, t_0)$ linearly and use a quadrature rule to get the contributions.

We can check that with the contributions $F_{\Omega_i\Omega_j}$ defined above, the Taylor expansion of the numerical solution corresponds up to second order with the Taylor expansion of the exact solution. Further explanation about the scheme can be found in [4] and [5].

5 Numerical Results

This method is implemented on a parallel machine with domain decomposition strategy. It performs very well on the Intel paragon.

We compare the numerical simulation of a supercritical expansion of water in a channel with the measurement of Hager and Mazumder [6].

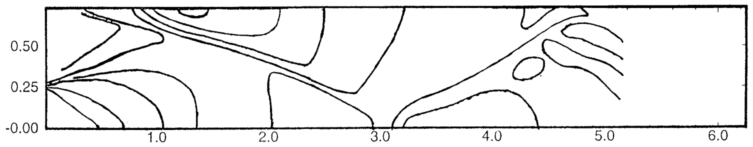

Figure 5: Experimental results of Hager and Mazumder [6], 10 contour lines of h for an abrupt expansion in a channel.

The channel is 8 m long and 1.5 m wide, the opening is one third of the total width and the water streams in with height of 96 mm and a Froude number of 2. The fact that water flows into a dry bed, does not cause problems. The first order scheme is a non negative scheme and the second order can be construct with the same property.
The space discretization uses a Cartesian grid with 240 × 45 points and a time step of $\Delta t = 10^{-2}$, which corresponds to a CFL number of 0.8. The computation is done till a steady state is reached. Here the pictures are shown at time $T = 5$ sec.
The stationary solution shows the same structure as the measurements, but in the simulation all lines are shifted downstream. The inclusion of friction corresponding to the bed shear stress to the model, leads to the correct solution (see Figure 6).
The friction can be added to the equations by a source term

$$\underline{S}(\underline{U}) = \begin{pmatrix} 0 \\ -g\,h\,S_{f_x} \\ -g\,h\,S_{f_y} \end{pmatrix},$$

where S_{f_x} and S_{f_y} are the slopes of the energy grade lines in the x and y directions respectively. The values are given by the steady state friction formulae $S_{f_x} = (n^2\, u\, \sqrt{u^2 + v^2})/h^{4/3}$ in which n is Manning's roughness coefficient. In each step, we use an operating splitting, i.e. we first solve the homogeneous equations (1) and then the ordinary differential equations $\underline{U}_t = \underline{S}(\underline{U})$ with a second order Runge-Kutta scheme. Since in this example the source term is not stiff, the operating splitting causes no problems.

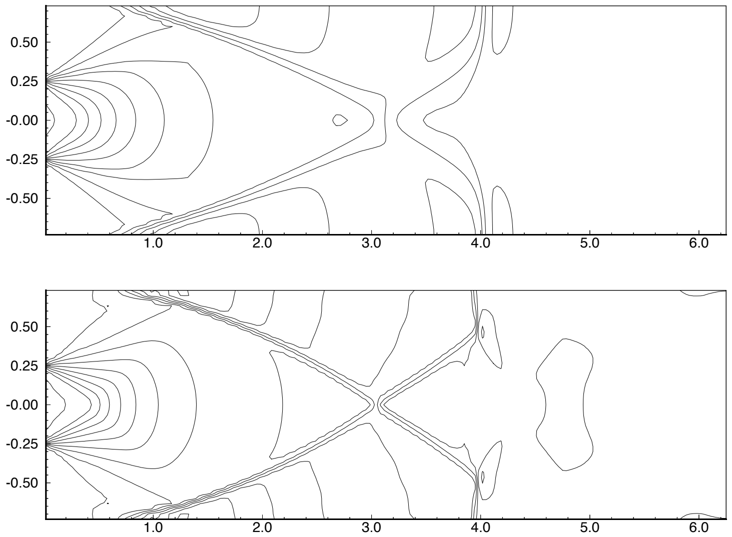

Figure 6: Abrupt expansion in a channel. 10 contour lines of h for the first order solution (upper figure) and the second order solution (lower figure).

The curved shock structure in the lower picture is well captured and coincides with the measurements. However, by adding viscosity to the model, we could possibly get better numerical results.

Ackowledgments

The project at SAM is partly funded by the Ercoftac fellowship program. The author would like to thank both Prof. R. Jeltsch and Dr. M. Fey for their constant support.

References

[1] S. J. Billett. A Class of Upwind Methods for Conservation Laws, Ph. D Thesis, Cranfield University, 1994.

[2] M. Fey. Ein echt mehrdimensionales Verfahren zur Lösung der Eulergleichungen, Dissertation, ETH Zürich, 1993.

[3] M. Fey and A.-T. Morel. Multidimensional method of transport for the shallow water equations, Research Report 95-05, Seminar für Angewandte Mathematik, ETH Zürich, 1995.

[4] M. Fey, R. Jeltsch and A.-T. Morel. Multidimensional schemes for nonlinear systems of hyperbolic conservation laws, to appear in the Proceeding of the 16th Biennial conference on numerical analysis, Dundee, 1995.

[5] M. Fey. Decomposition of the multidimensional Euler equations into advection equations, Research Report 95-14, Seminar für Angewandte Mathematik, ETH Zürich, 1995.

[6] W. H. Hager and S. K. Mazumder. Supercritical flow at abrupt expansions. *Proc. Instn Civ. Engrs Wat., Marit. and Energy*, 1992, 96, Sept., 153-166.

Research Reports

No.	Authors	Title
96-08	A.-T. Morel	Multidimensional Scheme for the Shallow Water Equations
96-07	M. Feistauer, C. Schwab	On coupled problems for viscous flow in exterior domains
96-06	J.M. Melenk	A note on robust exponential convergence of finite element methods for problems with boundary layers
96-05	R. Bodenmann, H.J. Schroll	Higher order discretisation of initial-boundary value problems for mixed systems
96-04	H. Forrer	Boundary Treatment for a Cartesian Grid Method
96-03	S. Hyvönen	Convergence of the Arnoldi Process when applied to the Picard-Lindelöf Iteration Operator
96-02	S.A. Sauter, C. Schwab	Quadrature for *hp*-Galerkin BEM in $\mathbb{R}^3$
96-01	J.M. Melenk, I. Babuška	The Partition of Unity Finite Element Method: Basic Theory and Applications
95-16	M.D. Buhmann, A. Pinkus	On a Recovery Problem
95-15	M. Fey	The Method of Transport for solving the Euler-equations
95-14	M. Fey	Decomposition of the multidimensional Euler equations into advection equations
95-13	M.D. Buhmann	Radial Functions on Compact Support
95-12	R. Jeltsch	Stability of time discretization, Hurwitz determinants and order stars
95-11	M. Fey, R. Jeltsch, A.-T. Morel	Multidimensional schemes for nonlinear systems of hyperbolic conservation laws
95-10	T. von Petersdorff, C. Schwab	Boundary Element Methods with Wavelets and Mesh Refinement
95-09	R. Sperb	Some complementary estimates in the Dead Core problem
95-08	T. von Petersdorff, C. Schwab	Fully discrete multiscale Galerkin BEM
95-07	R. Bodenmann	Summation by parts formula for noncentered finite differences
95-06	M.D. Buhmann	Neue und alte These über Wavelets
95-05	M. Fey, A.-T. Morel	Multidimensional method of transport for the shallow water equations