STABILITY AND EXISTENCE OF SOLUTIONS OF TIME-IMPLICIT FINITE VOLUME SCHEMES FOR VISCOUS NONLINEAR CONSERVATION LAWS

JÜRGEN FUHRMANN AND HARTMUT LANGMACH

Date: September 25, 1998.

1991 *Mathematics Subject Classification.* 65E99, 65M12.

Key words and phrases. viscous conservation laws, finite volume methods, nonlinear parabolic PDEs, groundwater flow.

Email. fuhrmann@wias-berlin.de, langmach@wias-berlin.de.

Address. Weierstraß Institute for Applied Analysis and Stochastics, Mohrenstr. 39, 10117 Berlin, Germany.

WIAS preprint no. 437.

ABSTRACT. We introduce a time-implicite Voronoi box based finite volume discretization for the initial-boundary value problem of a scalar nonlinear viscous conservation law in a one, two- or threedimensional domain. Using notations from the theory of explicit finite volume methods for hyperbolic problems and results from the Perron-Frobenius theory of nonnegative matrices, we establish various existence, stability and uniqueness results for the discrete problem.

The class of schemes introduced covers as well hyperbolic problems as well as nonlinear diffusion problems.

To clarify our results, we provide numerical examples, and we show the practical relevance of our considerations with a groundwater flow example.

1. INTRODUCTION

The finite volume method is a well known tool for the discretization of partial differential equations which model diffusion and transport phenomena. The equations describing these phenomena are often derived by a limit process which starts from balance equations in representative elementary volumes. Finite volumes can be seen as a way to avoid this limit process and to obtain a discrete physical model directly. At the other hand, finite element methods focus on the approximation properties between the discrete and the continuous physical model.

Based on these quite philosophic considerations, we can ask now which qualitative properties of our physical process can be carried over to the discrete model. In our paper we will consider local and global mass conservation and maximum principles for the time implicite Euler discretization. We will show that, provided we choose the right way to approximate convective terms, these properties hold. Moreover, in various cases, we are able to prove the existence of discrete solutions and to provide L^∞ and L^1 stability results. L^∞ stability and the local maximum principle are what prevents the "wiggles" in the solution. These results are correlateded with those for linear problems one finds in [GR92, Sto86]. As the basic technical tool we use estimates obtained from the Perron-Frobenius theory of nonnegative matrices.

For the description of the discrete nonlinear operators, we generalize the flux function based ansatz commonly used in the theory of explicit finite volume schemes for hyperbolic conservation laws [Krö97].

Within this paper, we use a Voronoi box based vertex centered finite volume method, also called covolume or control volume method [Mac53, Hei87, GG96, BW93, Van96, J.B98]. Other choices are possible, we only mention here midpoint box based vertex centered finite volume methods [FKTT97] and cell centered methods [EGH97a, EGH97b].

The paper is organized as follows. In section 2, we state the boundary value problem considered. Section 3 is devoted to the description of the triangulation of the computational domain and the construction of the Voronoi boxes. Further we describe the constraints to the triangulation necessary for the definition of our finite volume scheme in section 4. Together with this definition, section 4 describes two ways to write the discretization schemes in operator form which give rise to results based either on L^∞ or on L^1 estimates. These stability, existence and uniqueness results are stated section 5. Section 6 describes how our results recover the well known theory in the linear case. Further, we discuss nonlinear diffusion and nonlinear hyperbolic problems and provide one-dimensional numerical examples for these equations which illustrate our findings. At the end of section 6 we show the practical relevance of our considerations for a groundwater flow example from [FK97]. For the convenience of the reader, the appendix contains necessary results from linear algebra, and summarizes the results of [Fuh98a].

2. The problem class considered

Let $[0, T]$ with $T > 0$ be a finite time interval and $\Omega \subset \mathbb{R}^d$ $(d = 1 \dots 3)$ a polygonal domain. We assume that Ω is partitioned into subdomains which represent several *materials* $m \in \mathcal{M}$ such that $\overline{\Omega} = \bigcup_{m \in \mathcal{M}} \overline{\Omega^m}$.

Assume further, that $\Gamma = \partial\Omega = \Gamma_N \cup \Gamma_E \cup \Gamma_D$.

Consider the following initial boundary value problem:

2.1. Problem. *Find a function* $u(x,t) : \Omega \times [0,T] \to \mathbb{R}$ *such that*

$$\begin{aligned} \partial_t b(x,u) + \nabla \cdot \mathbf{j}(x,u) = 0 \\ \mathbf{j}(x,u) = -k(x,u)\nabla u + \mathbf{q}(x,u) \end{aligned} \tag{2.2}$$

with homogeneous Neumann (no flow) boundary conditions on Γ_N*,*

$$\mathbf{j}(u) \cdot \mathbf{n} = 0, \tag{2.3}$$

natural (equilibrium flow) boundary conditions on Γ_E*,*

$$\mathbf{j}(u) \cdot \mathbf{n} = \mathbf{q}(u) \cdot \mathbf{n}, \tag{2.4}$$

Dirichlet boundary conditions on Γ_D*,*

$$u|_\Gamma = u_D, \tag{2.5}$$

and the initial value $u(x,0) = u_0(x)$ *in* Ω*.*

We assume that the dependency of the constitutive relationships on the space variable can be written as follows: $b(x,u) = b^m(u)$, $k(x,u) = k^m(u)$ and $\mathbf{q}(x,u) = \mathbf{q}^m(u)$ on Ω^m, $m \in \mathcal{M}$. In this sense we could speak about *coefficient jumps* within the problem.

3. Voronoi Tesselations

Simplicial partitions of heterogeneous domains. Let Σ be a *cell complex* consisting of abstract cells of dimensions $0, \dots, d$. More precisely, it consists of

- $\mathcal{N} = \mathcal{N}(\Sigma)$, the set of *nodes* – abstract cells of dimension 0
- $\mathcal{E} = \mathcal{E}(\Sigma)$, the set of *edges* – abstract cells of dimension 1
- $\mathcal{F} = \mathcal{F}(\Sigma)$, the set of *faces* – abstract cells of codimension 1
- $\mathcal{C} = \mathcal{C}(\Sigma)$, the set of *cells* – abstract cells of codimension 0

The codimension is defined as the difference between the space dimension and the dimension of an object. In dimensions less than three, some of the sets conicide in a natural way, namely, for $d = 1$, $\mathcal{N}(\Sigma) = \mathcal{F}(\Sigma)$ and $\mathcal{C}(\Sigma) = \mathcal{E}(\Sigma)$, and for $d = 2$, $\mathcal{E}(\Sigma) = \mathcal{F}(\Sigma)$.

We assume that the cells of Σ describe a conformal simplicial tesselation of Ω [Cia78], and we suppose that each simplex belongs to exacly one subdomain Ω^m.

This allows us to define a mapping, called *cell material*

$$\mathrm{mat} : \mathcal{C}(\Sigma) \to \mathcal{M}, \tag{3.1}$$

and we have

$$\overline{\Omega} = \bigcup_{\sigma \in \mathcal{C}(\Sigma)} \overline{\sigma}, \qquad \overline{\Omega^m} = \bigcup_{\sigma \in \mathcal{C}(\Sigma), \mathrm{mat}(\sigma) = m} \overline{\sigma} \tag{3.2}$$

We introduce the set of *boundary faces* by

$$\mathcal{B}(\Sigma) = \{ f \in \mathcal{F}(\Sigma) : f \subset \partial\Omega \}$$

For any boundary face we demand that it belongs exactly to one of of $\Gamma_N, \Gamma_E, \Gamma_D$, and we define

$$\begin{aligned}
\mathcal{B}_N(\Sigma) &= \{f \in \mathcal{F}(\Sigma) : f \subset \Gamma_N\} \\
\mathcal{B}_E(\Sigma) &= \{f \in \mathcal{F}(\Sigma) : f \subset \Gamma_E\} \\
\mathcal{B}_D(\Sigma) &= \{f \in \mathcal{F}(\Sigma) : f \subset \Gamma_D\}.
\end{aligned}$$

We assume a numbering of all the nodes of Σ. Let $N = |\mathcal{N}|$. In the sequal, we simply will identify a node with its number. This corresponds to an isomorphism between the space of piecewise linear functions on Σ and $\mathbb{R}^N$.

We have an obvious adjacency relation between different types of abstract cells, and to each abstract cell c, we associate the *neighbourhood* $\mathtt{nb}_{\mathcal{X}}(c)$ where $\mathcal{X}$ is one of $\mathcal{N}, \mathcal{E}, \mathcal{F}, \mathcal{C}$ or $\mathcal{B}$.

Thus, for a node $i \in \mathcal{N}(\Sigma)$, $\mathtt{nb}_{\mathcal{N}}(i) = \{j \in \mathcal{N}(\Sigma) | (i, j) \in \mathcal{E}(\Sigma)\}$, and $\mathtt{nb}_{\mathcal{B}}(i)$ is the set of boundary faces adjacent to a node i if it is situated on the boundary, or the empty set if it is interior.

Let $\mathtt{nb}_{\mathcal{B}_E}(i) = \mathtt{nb}_{\mathcal{B}}(i) \cap \mathcal{B}_E$ be the set of faces adjacent to i corresponding to the equilibrium boundary condition.

Further, $\mathcal{D} = \mathcal{N} \cap \Gamma_D$ denotes set of Dirichlet nodes, and $\mathcal{I} = \mathcal{N} \setminus \mathcal{D}$ the set of non-Dirichlet nodes.

Voronoi boxes and geometrical coefficients.

3.3. Definition. Let $\mathbf{x}_i$ denote the vector of the coordinates of the node $i \in \mathcal{N}$. The *Voronoi box* ω_i around i with respect to the point set $\mathbb{N}$ and domain Ω is the set

$$\omega_i = \{\mathbf{x} \in \Omega : |\mathbf{x} - \mathbf{x}_i| < |\mathbf{x} - \mathbf{x}_j| \quad \forall j \in \mathcal{N},\ j \neq i\}. \tag{3.4}$$

Further, we need the following data (see also figures 3.1,3.2): the *interior Voronoi box faces*

$$\gamma_{ij} = \overline{\omega_i} \cap \overline{\omega_j}, \quad i \neq j,$$

the outward unit normal to the Voronoi box boundary

$$\mathbf{n}_i$$

the outward unit normal to the interior Voronoi box face

$$\mathbf{n}_{ij}, \quad i \neq j,$$

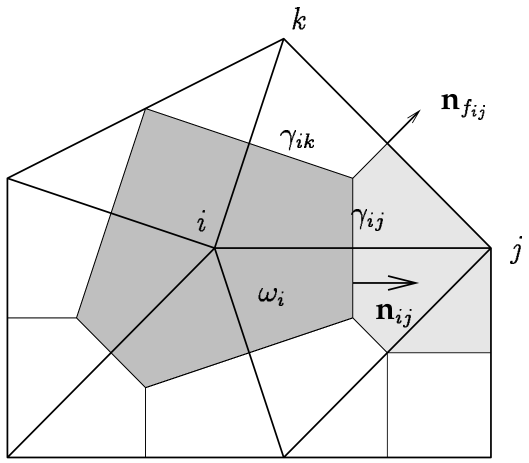

FIGURE 3.1. Voronoi box for the two-dimensional case

the *edge lengths*

$$h_{ij} = |\mathbf{x}_i - \mathbf{x}_j|, \quad i \neq j,$$

the *boundary Voronoi box faces* corresponding to a boundary face f and a node i

$$\gamma_{if} = \overline{\omega_i} \cap f \subset \overline{\omega_i} \cap \partial\Omega,$$

the outward unit normal to the boundary face face f

$$\mathbf{n}_f.$$

Further, we define

$$\omega_i^m = \omega_i \cap \Omega^m, \qquad \gamma_{ij}^m = \gamma_{ij} \cap \Omega^m, \qquad \gamma_{if}^m = \gamma_{if} \cap \overline{\partial\Omega^m}.$$

Constraints to the triangulation. The following condition we call *Delaunay property conforming to inner and outer boundaries*

3.5. Condition. *For any $m \in \mathcal{M}$, we demand*

(i). *If $|\overline{\omega_i^m} \cap \overline{\omega_j^m}| > 0$, then $(i,j) \in \mathcal{E}$*
(ii). *If $(i,j) \in \mathcal{E}$ and $\omega_i^m \neq \emptyset$, $\omega_j^m \neq \emptyset$, then $\overline{\omega_i^m} \cap \overline{\omega_j^m} \neq \emptyset$.*

In one space dimension, this property is always fulfilled. In two space dimensions, this demand is equivalent to

3.6. Condition.

(i). *For each interior face, the sum of the angles of the neigbouring triangles opposite to this face is less or equal than π.*
(ii). *Each angle opposite to an inner or outer boundary face is less or equal than $\pi/2$.*

3.7. *Corollary.*

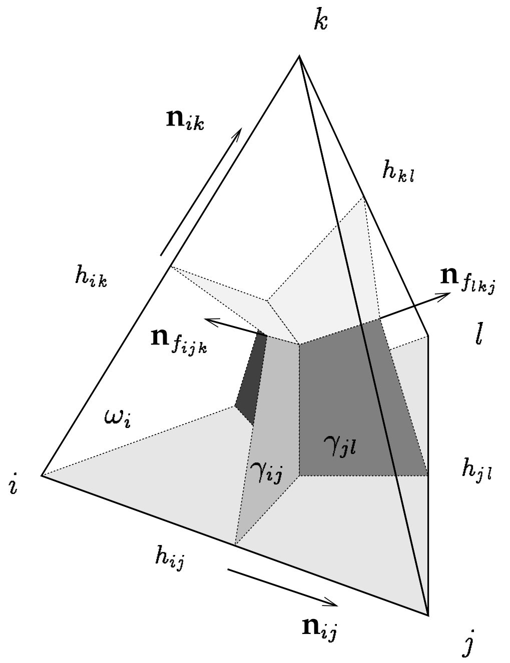

FIGURE 3.2. Intersection between Voronoi box and simplex for the three-dimensional case.

(i). Let i be an interior node. Then we have

$$\bigcup_{j\in \mathbf{nb}_{\mathcal{N}}(i)} \gamma_{ij} = \partial\omega_i \tag{3.8}$$

$$\sum_{j\in \mathbf{nb}_{\mathcal{N}}(i)} |\gamma_{ij}|\mathbf{n}_{ij} = 0 \tag{3.9}$$

(ii). Let i be a boundary node. Then we have

$$\bigcup_{j\in \mathbf{nb}_{\mathcal{N}}(i)} \gamma_{ij} + \bigcup_{f\in \mathbf{nb}_{\mathcal{B}}(i)} \gamma_{if} = \partial\omega_i \tag{3.10}$$

$$\sum_{j\in \mathbf{nb}_{\mathcal{N}}(i)} |\gamma_{ij}|\mathbf{n}_{ij} + \sum_{f\in \mathbf{nb}_{\mathcal{B}}(i)} |\gamma_{if}|\mathbf{n}_f = 0 \tag{3.11}$$

Proof. Equations (3.9) and (3.11) are consequences of Gauss' theorem. □

4. Nonlinear difference operators on heterogeneous unstructured meshes

The difference scheme. Assume a partition of the time interval $[0,T]$ with monotonically increasing times t^n and time steps $\tau^n = t^{n+1} - t^n$. Let $u_i^n = u(\mathbf{x}_i, t^n)$. Equation (2.2) gives for all n and $i \in \mathcal{N}$:

(4.1)

$$0 = \int_{t^n}^{t^{n+1}} \int_{\omega_i} \left(\frac{\partial b(x,u)}{\partial t} + \nabla\cdot\mathbf{j} \right) d\omega$$

$$= \int_{t^n}^{t^{n+1}} \left(\int_{\omega_i} \frac{\partial b(x,u)}{\partial t} d\omega + \int_{\partial\omega_i} \mathbf{j}\cdot\mathbf{n}_i d\gamma \right) d\tau$$

$\mathbf{n}_i$ being the unit outward normal of ω_i

$$= \int_{\omega_i} \left(b(x,u^{n+1}) - b(x,u^n) \right) d\omega +$$

$$+ \int_{t^n}^{t^{n+1}} \left(\sum_{j\in \mathbf{nb}_{\mathcal{N}}(i)} \int_{\gamma_{ij}} \mathbf{j}\cdot\mathbf{n}_{ij} d\gamma + \sum_{f\in \mathbf{nb}_{\mathcal{B}}(i)} \int_{\gamma_{if}} \mathbf{j}\cdot\mathbf{n}_f d\gamma \right) d\tau$$

$$= \sum_{m\in\mathcal{M}} \int_{\omega_i^m} \left(b^m(u^{n+1}) - b^m(u^n) \right) d\omega +$$

$$+ \sum_{m\in\mathcal{M}} \int_{t^n}^{t^{n+1}} \left(\sum_{j\in \mathbf{nb}_{\mathcal{N}}(i)} \int_{\gamma_{ij}^m} \mathbf{j}\cdot\mathbf{n}_{ij} d\gamma + \sum_{f\in \mathbf{nb}_{\mathcal{B}}(i)} \int_{\gamma_{if}^m} \mathbf{j}\cdot\mathbf{n}_f d\gamma \right) d\tau$$

Now we approximate the last condition by

$$\begin{aligned} &\sum_{m\in\mathcal{M}} |\omega_i^m| \left(b^m(u_i^{n+1}) - b^m(u_i^n) \right) + \\ &+ \tau^n \sum_{m\in\mathcal{M}} \left(\sum_{j\in \mathbf{nb}_{\mathcal{N}}(i)} g_{ij}^m(u_i^{n+1}, u_j^{n+1}) + \sum_{f\in \mathbf{nb}_{\mathcal{B}_{\mathcal{B}}}(i)} |\gamma_{if}^m| \mathbf{q}^m(u_i^{n+1})\cdot\mathbf{n}_f \right) = 0 \end{aligned} \tag{4.2}$$

g_{ij}^m is called the *numerical flux* between node i and node j. For simplicity of notation, we assume that all coefficients g_{ij}^m are defined over the whole domain. If an edge (i,j) lies outside of material m, i.e. $[x_i, x_j] \cap \bar{\Omega}^m = \emptyset$, we assume $g_{ij}^m = 0$.

The approximation step includes a quadrature in space for the calculation of the integrals over the Voronoi boxes, a quadrature in time for the calculation of the divergence integral and a quadrature in space for the calculation of the boundary integrals. Along Γ_N we used the homogeneous Neumann boundary condition (2.3), and along Γ_E, we inserted the equilibrium flow boundary condition (2.4), together with certain quadrature. The Dirichlet boundary condition we will insert algebraically below.

As mentioned above, there are various choices for the approximation of the time integrals, we took the one which gives us a fully implicite scheme, because we are interested in unconditional stability results.

Operator notation. We can rewrite the scheme (4.2) in operator form

$$B(u^{n+1}) - B(u^n) + \tau^n\big(A(u^{n+1}) + Q_E(u^{n+1})\big) = 0 \tag{4.3}$$

with a diagonal operators B and Q_E and an operator A admitting off diagonal nonlinearities. Let

$$\begin{aligned} b_i(u_i) &= \sum_{m\in\mathcal{M}} \frac{|\omega_i^m|}{|\omega_i|} b^m(u_i), \\ g_{ij}(u_i,u_j) &= \sum_{m\in\mathcal{M}} g_{ij}^m(u_i,u_j). \end{aligned} \tag{4.4}$$

Then

$$\begin{aligned} (B(u))_i &= |\omega_i| b_i(u_i) \\ (A(u))_i &= \sum_{j\in \mathrm{nb}_{\mathcal{N}}(i)} g_{ij}(u_i,u_j) \\ (Q_E(u))_i &= \sum_{f\in \mathrm{nb}_{\mathcal{B}_E}(i)} |\gamma_{if}^m| \mathbf{q}^m(u_i)\cdot \mathbf{n}_f \end{aligned} \tag{4.5}$$

This formulation will be the base of later considerations concerning L^1 estimates.

To obtain the maximum principle and L^∞ estimates, we will need to rewrite the system in the following way. Let

$$\begin{aligned} (Q_J(u))_i &= \sum_{j\in \mathrm{nb}_{\mathcal{N}}(i)} g_{ij}(u_i,u_i) \\ Q(u) &= Q_E(u) + Q_J(u) \end{aligned} \tag{4.6}$$

Assume that all functions g_{ij} are continuously differentiable. Define

$$k_{ij}(u_i,u_j) = \begin{cases} \frac{g_{ij}(u_i,u_j)-g_{ij}(u_i,u_i)}{u_i-u_j}, & u_i \neq u_j \\ \partial_2 g_{ij}(u_i,u_i), & u_i = u_j \end{cases} \tag{4.7}$$

and the operator $K(u): \mathbb{R}^N \to \mathbb{R}^N$ parametrized by $u \in \mathbb{R}^N$ by its entries

$$(K(u))_{ij} = \begin{cases} -k_{ij}(u_i,u_j), & (i,j)\in\mathcal{E} \\ \sum\limits_{l\in \mathrm{nb}_{\mathcal{N}}(i)} k_{il}(u_i,u_l), & i=j \\ 0, & \text{else.} \end{cases} \tag{4.8}$$

Here, ∂_1 and ∂_2 denote the partial derivatives of with respect to the first and second arguments, respectively.

Then $A(u) = K(u)u + Q_J(u)$, and equation (4.3) is equivalent to

$$B(u^{n+1}) - B(u^n) + \tau^n\big(K(u^{n+1})u^{n+1} + Q(u^{n+1})\big) = 0 \tag{4.9}$$

All the operator names introduced here we will keep throughout the rest of the paper.

Dirichlet boundary conditions. By fixing the values in the Dirichlet boundary nodes, we can define the Dirichlet problem algebraically in the same way as in [Fuh98a]. The partitioning $\mathcal{N} = \mathcal{D} \cup \mathcal{I}$ for any vector u induces a partitioning $u = \begin{pmatrix} u_{\mathcal{D}} \\ u_{\mathcal{I}} \end{pmatrix}$ We write

$$\begin{cases} B(u^{n+1}) - B(u^n) + \tau^n\big(A(u^{n+1}) + Q_E(u^{n+1})\big) & = 0 \\ u|_{\mathcal{D}} & = u_d \end{cases} \tag{4.10}$$

for demanding

$$u_{\mathcal{D}}^{n+1} = u_d$$
$$B_{\mathcal{I}}(u_{\mathcal{I}}^{n+1}) - B_{\mathcal{I}}(u_{\mathcal{I}}^n) + \tau^n\big(A_{\mathcal{I}}(u^{n+1}) + Q_{E\mathcal{I}}(u^{n+1})\big) = 0$$

where for an operator $X : \mathbb{R}_{\mathcal{N}} \to \mathbb{R}_{\mathcal{N}}$ we denote by $X_{\mathcal{I}}$ the operator defined by the subset $(X)_i$ with $i \in \mathcal{I}$.

For a matrix K, we get a block structure

$$\begin{pmatrix} K_{\mathcal{D}\mathcal{D}} & K_{\mathcal{D}\mathcal{I}} \\ K_{\mathcal{I}\mathcal{D}} & K_{\mathcal{I}\mathcal{I}}. \end{pmatrix}$$

We say that u^n is a solution of the Dirichlet problem

$$\begin{cases} B(u^{n+1}) + \tau^n\big(K(u^{n+1})u^{n+1} + Q(u^{n+1})\big) & = B(u^n) \\ u^{n+1}|_D & = u_d \end{cases} \tag{4.11}$$

if

$$\begin{aligned} u_D^{n+1} &= u_d \\ B_{\mathcal{I}}(u_{\mathcal{I}}^{n+1}) + K_{\mathcal{I}\mathcal{I}}(u^{n+1})u_{\mathcal{I}}^{n+1} + K_{\mathcal{I}\mathcal{D}}(u^{n+1})u_d + Q_{\mathcal{I}}(u^{n+1}) &= B_{\mathcal{I}}(u_{\mathcal{I}}^n). \end{aligned} \tag{4.12}$$

Conditions and definitions.

4.13. Condition. *We assume the following properties for the numerical fluxes* (4.2)

(i). *Smoothness:*

$$g_{ij}^m(u, v) \in C^1(\mathbb{R} \times \mathbb{R}), b^m \in C^1(\mathbb{R}) \tag{4.14}$$

(ii). *Mass conservation:*

$$g_{ij}^m(u, v) = -g_{ji}^m(v, u) \tag{4.15}$$

(iii). *Flux consistency:*

$$g_{ij}^m(u, u) = |\gamma_{ij}^m| \mathbf{n}_{ij} \cdot \mathbf{q}^m(u) \tag{4.16}$$

(iv). *Diffusion consistency: If* $\mathbf{q} = 0$ *then for each pair* u, v, *there exists* $\lambda \in [u, v]$ *with*

$$g_{ij}^m(u, v) = \frac{|\gamma_{ij}^m|}{h_{ij}} k^m(\lambda)(u - v) \tag{4.17}$$

4.18. *Remark.* Conditions (4.15) and (4.16) have been taken from [Krö97]. The diffusion consistency (4.17) is motivated by the mean value theorem for the integral of k and naturally does not appear in the theory of hyperbolic problems. However it fits perfectly into the framework of this theory. The smoothness properties which are a stronger demand than in [Krö97] (where only Lipschitz continuity is demandend) are necessary for the mean value theorem B.1 and equation (4.7).

The name "mass conservation property" is motivated by theorem 5.4 and

4.19. *Corollary.* If the scheme (4.2) fulfills the conditions 4.13, we have

$$\sum_{i\in\mathcal{N}}(A(u))_i = 0 \tag{4.20}$$

Proof. Obviously, condition (4.15) is inherited by g_{ij} from g^m_{ij}.

$$\begin{aligned}
\sum_{i\in\mathcal{N}}(A(u)_i) &= \sum_{i\in\mathcal{N}}\sum_{j\in\mathrm{nb}_\mathcal{N}(i)} g_{ij}(u_i,u_j)\\
&= \sum_{i\in\mathcal{N}}\Big(\sum_{\substack{j\in\mathrm{nb}_\mathcal{N}(i)\\ j<i}} g_{ij}(u_i,u_j) + \sum_{\substack{j\in\mathrm{nb}_\mathcal{N}(i)\\ j>i}} g_{ij}(u_i,u_j)\Big)\\
&= \sum_{i\in\mathcal{N}}\Big(\sum_{\substack{j\in\mathrm{nb}_\mathcal{N}(i)\\ j<i}} g_{ij}(u_i,u_j) + \sum_{\substack{j\in\mathrm{nb}_\mathcal{N}(i)\\ j>i}} -g_{ji}(u_j,u_i)\Big)\\
&= \sum_{i\in\mathcal{N}}\sum_{\substack{j\in\mathrm{nb}_\mathcal{N}(i)\\ j<i}} g_{ij}(u_i,u_j) + \sum_{j\in\mathcal{N}}\sum_{\substack{i\in\mathrm{nb}_\mathcal{N}(j)\\ j<i}} -g_{ij}(u_i,u_j)\\
&= 0
\end{aligned}$$

because the last summations both sum over the directed edges $(i,j)\in\mathcal{E}$ with $i>j$. ☐

4.21. Definition. The scheme (4.2) is called *weakly isotone* if the triangulation fulfills condition 3.5 and if for all $u\in\mathbb{R}_\mathcal{N}$

$$\begin{aligned} b_i'(u) &\geq 0\\ k_{ij} &\geq 0 \end{aligned} \tag{4.22}$$

4.23. *Corollary.* If the scheme fulfills the conditions 4.13 and is weakly isotone, then B is an isotone diagonal mapping, and $K(u)\in\mathbb{Z}^0_r(N)$ is a Z-matrix with row sum zero continuously depending on u.

Proof. Everything follows directly from the defining equations (4.5), (4.7), (4.8). ☐

4.24. Definition. The scheme (4.2) is called *isotone* if the triangulation fulfills condition 3.5 and if for all $u\in\mathbb{R}^N$,

$$\begin{aligned} b_i'(u) &\geq b^- > 0\\ \partial_1 g_{ij} &\geq 0\\ \partial_2 g_{ij} &\leq 0 \end{aligned} \tag{4.25}$$

4.26. *Corollary.* If the scheme fulfills the conditions 4.13 and is isotone, then B is an isotone diagonal homeomorphism, and $A(u)$ is Gâteaux differentiable with a Gâteaux derivative $A'(u)\in\mathbb{Z}^0_c(N)$ continuously depending on u. Further, in this case it is weakly isotone as well.

Proof. For B, this is obvious. The sign pattern of A' is a consequence of (4.25), and the column sum zero property comes from corollary 4.19.

If the scheme is isotone, g_{ij} it is nonincreasing in the second argument and we have $(g_{ij}(u,v)-g_{ij}(u,u))(u-v)\geq 0$. But we have the same sign for the quotient

$$\frac{g_{ij}(u,v)-g_{ij}(u,u)}{v-u}$$

which just defines k_{ij}. ☐

4.27. Definition. A node i is called an equilibrium node with respect to the scheme (4.2) if $(Q(u))_i = 0$.

4.28. *Corollary.* If the scheme (4.2) fulfills the conditions 4.13, a node is an equilibrium node if at least one of the following conditions are fulfilled:

(i). It belongs to the interior of some material.
(ii). It is situated at the boundary equipped with equilibrium flow boundary conditions, and adjacent to only one material.
(iii). It is situated at an interior material boundary or an outer boundary with equilibrium boundary conditions, but all convective terms q^m in the adjacent materials are equal.

Proof. This is a consequence of (3.9) and (3.11). □

5. PROPERTIES OF THE DISCRETIZATION

Mass conservation. Assert no flow boundary conditions on the whole boundary: $\Gamma = \Gamma_N$. Then for a solution of problem 2.1 we have

$$\int_\Omega \nabla \cdot \mathbf{j} d\omega = 0 \tag{5.1}$$

and thus

$$\begin{aligned} 0 &= \int_{t^n}^{t^{n+1}} \int_\Omega \frac{\partial b(x,u)}{\partial t} d\omega \\ &= \int_\Omega \left(b(x, u^{n+1}) - b(x, u^n) \right) d\omega \end{aligned} \tag{5.2}$$

which means *global mass conservation*. As we have chosen the global mass

$$m^n = \sum_{i \in \mathcal{N}} |\omega_i| b_i(u_i^n) = 1^T B(u^n) \tag{5.3}$$

as an approximation to $\int_\Omega b(x, u^n) d\omega$, we get

5.4. Theorem. *With no flow boundary conditions* (2.3)*, the scheme* (4.2) *with conditions 4.13 is globally mass conservative.*

Proof. We have

$$\begin{aligned} m^{n+1} - m^n &= 1^T B(u_i^{n+1}) - 1^T B(u_i^n) \\ &= \sum_{i \in \mathcal{N}} (A(u))_i = 0 \end{aligned}$$

because of (4.20). □

Local maximum principle. The following theorem corresponds to the main result of [FK97], similar considerations one finds in [Fro98] and in [Sto86]. This property is closely related to the fact that $K(u) \in \mathbb{Z}_r^0(N)$.

5.5. Theorem. *If the scheme* (4.2) *is weakly isotone, it admits a local maximum principle in any equilibrium node i with $b_i'(u_i) > 0$ and $k_{ij}(u_i, u_j) > 0$ for $j \in \mathrm{nb}_\mathcal{N}(i)$ in the sense that, provided, u^{n+1} is a solution of* (4.2)*, we have*

$$u_i^{n+1} \le \max(\{u_i^n\} \cup \{u_j^{n+1}\}_{j \in \mathrm{nb}_\mathcal{N}(i)}) \tag{5.6}$$

Proof. We derive the result based on equation (4.9). The mean value theorem B.1 yields a nonnegative diagonal operator $\tilde{B}(u^{n+1}, u^n)$ which together with $(Q(u))_i = 0$ allows to rewrite the ith equation in (4.9) as

$$\tilde{b}_i(u_i^{n+1}, u_i^n)(u_i^{n+1} - u_i^n) + \tau^n \sum_{j \in \mathrm{nb}_\mathcal{N}(i)} k_{ij}(u_i^{n+1}, u_j^{n+1})(u_i^{n+1} - u_j^{n+1}) = 0 \tag{5.7}$$

Let $d = \tilde{b}_i + \tau^n \sum_{j \in \mathrm{nb}_{\mathcal{N}}(i)} k_{ij}$. Then we have

$$u_i^{n+1} = \frac{\tilde{b}}{d} u_i^n + \sum_{j \in \mathrm{nb}_{\mathcal{N}}(i)} \frac{\tau^n k_{ij}}{d} u_j^{n+1} \tag{5.8}$$

with all coefficients in the right hand side being nonnegative and their sum not greater than 1. Thus we can take the maximum at the right hand side. □

5.9. *Remark.* From equation (5.8) we easily can derive a maximum principle in the degenerate case $b_i'(u_i) \geq 0$.

Existence, L^∞-stability, uniqueness, nonnegativity. Here, we refer to variant (4.9) of (4.2).

5.10. Theorem. *Let the scheme* (4.2) *be weakly isotone and fulfill the conditions 4.13.*

(i). *If $(b^m)' > 0$ and $Q(u)$ is bounded in the L^∞-norm, the time step problems* (4.11) *have a solution u^{n+1}.*

(ii). *If $Q = 0$, the time step problems have a solution u^{n+1} and the scheme is unconditionally stable in the sense that*

$$||u^{n+1}||_\infty \leq ||u^n||_\infty \leq ||u^0||_\infty + ||u_d||_\infty \tag{5.11}$$

(iii). *If $(b^m)' > 0$ and $Q = 0$, and if u^0 and u_d are nonnegative, then all u^n are nonnegative*

Proof. The first two parts follow directly from theorem C.1,C.5. The nonnegativity comes from the mean value theorem which allows to rewrite (4.9) as

$$\tilde{B}(u^{n+1}, u^n)u^{n+1} + K(u^{n+1})u^{n+1} = \tilde{B}(u^{n+1}, u^n)u^{n+1}$$

and the fact that $B(u^{n+1}, u^n) + K(u^{n+1})$ is a M-Matrix. □

Existence, uniqueness, stability and dissipativity in the L^1 norm. This result is based on version (4.3) of (4.2).

5.12. Theorem. *Let the scheme 4.2 be isotone.*

(i). *If $Q_E(u)$ is bounded,the time step problems* (4.10) *have a solution u^{n+1}.*

(ii). *For Dirichlet and homogeneous Neumann boundary conditions the solutions of the time step problems* (4.10) *are unique.*

(iii). *If we assume homogeneous Neumann boundary conditions and $A(B^{-1}(0)) = 0$, the scheme* (4.3) *is unconditionally stable in the discrete L^1-norm:*

$$||B(u^{n+1})||_1 \leq ||B(u^n)||_1 \leq ||B(u^0)||_1 \tag{5.13}$$

(iv). *If we assume homogeneous Neumann boundary conditions, the scheme is dissipative in the discrete L^1-norm in the sense* [GR92] *that for any u_0, v_0, we have the estimate*

$$||B(u^{n+1}) - B(v^{n+1})||_1 \leq ||B(u^n) - B(v^n)||_1 \leq ||B(u^0) - B(v^0)||_1 \tag{5.14}$$

.

(v). *For $Q_E = 0$ and $A(0) = 0$ if u^0 is nonnegative, this is valid for all u^n.*

Proof. The first part, the stability estimate and the uniqueness statement are a direct consequence of C.9 and C.12, respectively. For the dissipativity statement, using the mean value theorem B.1, rewrite (4.3) as

$$B(u^{n+1}) - B(v^{n+1}) + \tau^n \tilde{A}_B(u^{n+1}, v^{n+1})(B(u^{n+1}) - B(v^{n+1})) = B(u^n) - B(v^n)$$

with $A_B(u) = A(B^{-1}(u))$ using the identity $A_B(B(u)) = A(u)$ and the fact that $A_B(u)$ by isotonicity of B has the same properties as $A(u)$. The result follows from the equivalent equation

$$B(u^{n+1}) - B(v^{n+1}) = (I + \tau^n \tilde{A}_B)^{-1}(B(u^n) - B(v^n))$$

and theorem A.9. The nonnegativity result follows from rewriting the scheme based on the mean value theorem B.1 as

$$\begin{aligned}(\tilde{B}(u^{n+1}, u^n))_{\mathcal{I}} u_{\mathcal{I}}^{n+1} + \tau^n(\tilde{A}(u^{n+1}, 0))_{\mathcal{I}\mathcal{I}} u_{\mathcal{I}}^{n+1} + \tau^n(\tilde{A}(u^{n+1}, 0))_{\mathcal{I}\mathcal{D}} u_d \\ = (\tilde{B}(u^{n+1}, u^n))_{\mathcal{I}} u_{\mathcal{I}}^n\end{aligned} \tag{5.15}$$

and the fact that $\tilde{B}(u^{n+1}, u^n) + \tau^n \tilde{A}(u^{n+1}, 0)$ is an M-Matrix. □

6. Examples

For a more intuitive understanding, videos of the results in the MPEG format of the numerical simulations can be downloaded from the web page [Fuh98b].

Linear diffusion. Regard the linear heat conduction equation

$$u_t - \nabla \cdot (k(x)\nabla u) = 0 \tag{6.1}$$

with piecewise constant heat conductivity k. Let

$$g_{ij}^m(u, v) = \frac{|\gamma_{ij}|}{h_{ij}} k^m (u - v). \tag{6.2}$$

For this equation, Neumann boundary conditions (2.3) and equilibrium flow boundary conditions (2.4) coincide. (6.2) is a commonly used ansatz for the finite volume discretization, which dates back at least to [Mac53]. In two space dimensions, it is equivalent to finite elements, but in three space dimensions, this equivalence does not hold. [Hac89, GG96]. The resulting scheme is isotone, and all inner nodes are equilibrium nodes, notably those at material boundaries. Thus also in these nodes, the maximum principle is valid.

Linear hyperbolic problems. Regard the linear transport equation

$$u_t + \nabla \cdot (\mathbf{q}(x)u) = 0 \tag{6.3}$$

with piecewise constant convection $\mathbf{q}$. We regard the Engquist-Osher scheme [EO81].

$$g_{ij}^m(u, v) = \frac{|\gamma_{ij}|}{h_{ij}} (q_{ij}^+ u - q_{ij}^- v) \tag{6.4}$$

where

$$q_{ij}^m = h_{ij} \mathbf{q}_{ij}^m \cdot \mathbf{n}_{ij} \tag{6.5}$$

$$q_{ij}^{m+(-)} = \begin{cases} q_{ij}^m, & q_{ij}^m > (<) 0 \\ 0, & q_{ij}^m < (>) 0. \end{cases}$$

For this type of equation, the equilibrium flow boundary condition results in a maximum principle at the boundary nodes. We will see this illustrated in the nonlinear example. The scheme is isotone, and at material boundaries we do not have a maximum principle, which physically is quite reasonable, as we observe a "congestion" of the transported "material", when the speed of transport given by the absolute value of q changes. In practical calculations, for this kind of problems explicit higher order finite volume schemes [Krö97] are preferred.

Linear convection-diffusion problems. Regard the convection-diffusion equation

$$u_t - \nabla\cdot(k(x)\nabla u - \mathbf{q}(x)u) = 0 \tag{6.6}$$

with piecewise constant coefficients. We can define an upwind scheme for this problem in the following way using q_{ij}^m from (6.5):

$$g_{ij}^m(u,v) = \frac{|\gamma_{ij}|}{h_{ij}} k^m \left(U\big(\frac{q_{ij}^m}{k^m}\big)u - U\big(-\frac{q_{ij}^m}{k^m}\big)v \right) \tag{6.7}$$

If

$$U(\xi) = \frac{1}{2}|\xi| - \frac{1}{2}\xi + 1 = \begin{cases} 1-\xi, & \xi \le 0 \\ 1, & \xi > 0, \end{cases}$$

we get the simple upwind scheme. For $U(\xi) = B(\xi)$ with the Bernoulli function

$$B(\xi) = \frac{\xi}{e^{\xi} - 1}$$

we get the scheme independently described by [AS55, Il'69, SG69].

In both cases, in the limit $k \to 0$, we get the Engquist-Osher scheme. If q is nonzero, with respect to the maximum principle, we get similar effects as in the purely hyperbolic case.

Nonlinear diffusion. Regard the nonlinear heat conduction equation

$$u_t - \nabla\cdot(k(x,u)\nabla u) = 0 \tag{6.8}$$

Here, we can introduce at least two schemes which fulfil the diffusion consistency condition:

$$g_{ij}^m(u,v) = \frac{|\gamma_{ij}|}{h_{ij}} k^m\big(\frac{u+v}{2}\big)(u-v) \tag{6.9}$$

and for $k^m(u) = (K^m(u))'$,

$$g_{ij}^m(u,v) = \frac{|\gamma_{ij}|}{h_{ij}} (K^m(u) - K^m(v)) \tag{6.10}$$

The scheme (6.10) is isotone, while (6.9) is an example for a scheme which is weakly isotone, but not isotone. Of course, the isotonicity property is highly desirable from the theoretical point of view - it allows us to establish uniqueness and the L^1-estimates, but in many practical cases, $K^m(u)$ is not at hand, so that we are forced to stick to the weakly isotone variant.

As in the linear diffusion case, all inner nodes are equilibrium nodes, and noflow (2.3) and equilibrium flow (2.4) boundary conditions coincide.

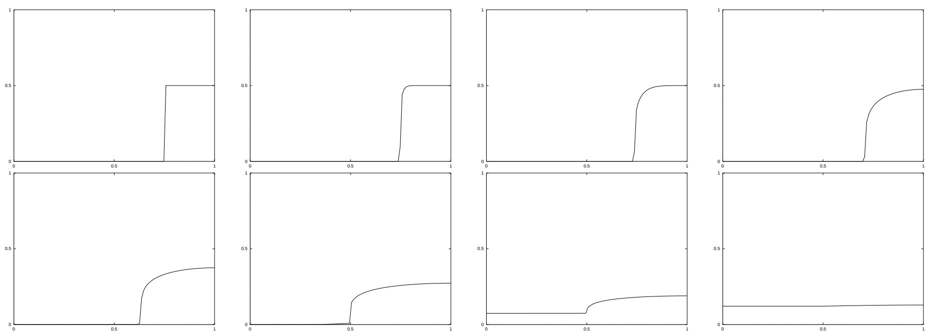

FIGURE 6.1. Solutions of (6.11) with $\beta^1 = 2, \beta^2 = 6, t = 0, 10^k, k = 1\dots7$.

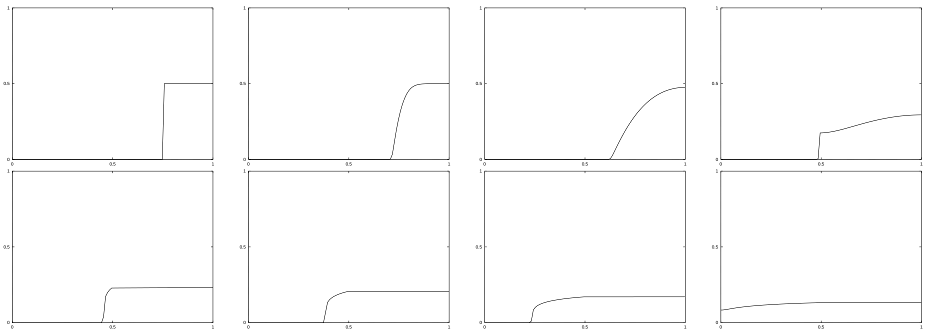

FIGURE 6.2. Solutions of (6.11) with $\beta^1 = 6, \beta^2 = 2, t = 0, 10^k, k = 1 \ldots 7$.

A special case of this type of equations is the porous medium equation

$$u_t - \nabla \cdot (\beta(x) u^{\beta(x)-1} \nabla u) = 0 \tag{6.11}$$

For constant β, it has a well known analytical solution with a finite support, and numerical examples in [Fuh97a] show that this type of fully implicite schemes is able to reproduce this property in a straightforward way. For the behaviour of other solution schemes and an adaptive method, see [Bän95].

We sove (6.11) numerically in a domain splitted into two materials: $\overline{\Omega} = [0,1] = \overline{\Omega^1} \cup \overline{\Omega^2}$ with $\Omega^1 = (0, 0.5)$, $\Omega^2 = (0.5, 1)$. It is discretized by an equidistant 100 node mesh. The initial value is

$$u_0(x) = \begin{cases} 0, & x \leq 0.75 \\ 0.5, & x > 0.75 \end{cases}$$

We use equilibrium flow boundary conditions (2.4). If not stated otherwise, these data will be the same for also for the next example.

Figures 6.1 and 6.2 show the behaviour of the solution depending on different combinations of the exponent. We see the finite support, and we also see the diffusive behaviour. Further, there is no local maximum of the solution at the inner boundary.

Nonlinear hyperbolic conservation laws. Regard

$$u_t + \nabla \cdot \mathbf{q}(x, u) = 0 \tag{6.12}$$

with the same domain, initial and boundary conditions as above. Here again, the nonlinear variant of the Enguist-Osher scheme is the simplest choice to gain isotonicity. In this case it takes the form

$$g^m_{ij}(u, v) = q^{m+}_{ij}(u) - q^{m-}_{ij}(v)$$

with $q^m_{ij}(u) = \mathbf{n}_{ij} \cdot \mathbf{q}^m(u)$, $q^{m+}_{ij}(u) = q^m_{ij}(0) + \int_0^u \max(q^m_{ij}(s), 0) ds$ and $q^{m-}_{ij}(u) = \int_0^u \min(q^m_{ij}(s), 0) ds$. This scheme again is isotone, and all nodes in the interior of the material domains and at the equilibrium flow boundary are equilibrium nodes.

A special case of this type of equations is

$$u_t - \nabla \cdot u^{\beta(x)} = 0 \tag{6.13}$$

For $\beta = 2$, it coincides up to a factor with Burger's equation.

The behaviour of the solutions is illustrated by the figures 6.3 and 6.4. We observe that at the material boundary, our equilibrium condition from definition 4.27 is violated, if we have a jump in q. The consequence is that in fact, for $\beta^1 > \beta^2$ and $u < 1$, we get a new local maximum. At the other hand, the equilibrium flow

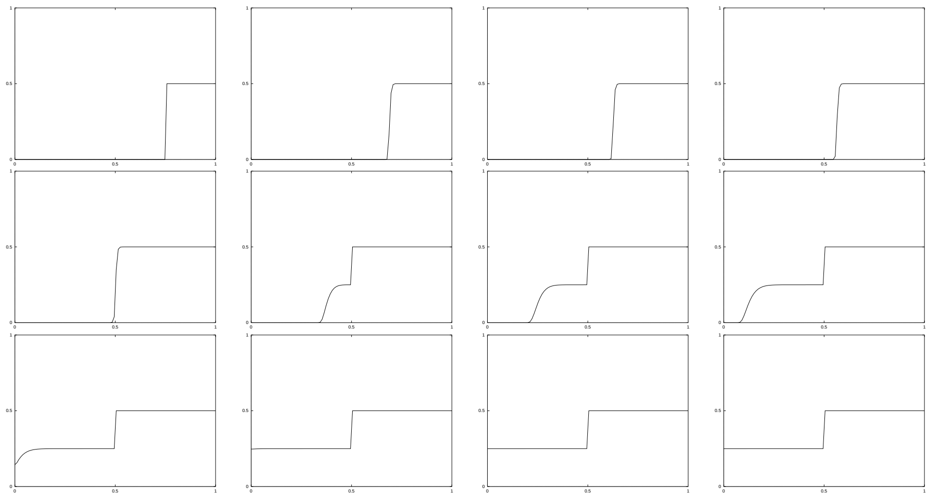

FIGURE 6.3. Solutions of (6.13) with $\beta^1 = 2, \beta^2 = 4, t = 0.5 \cdot k, k = 0 \dots 11$.

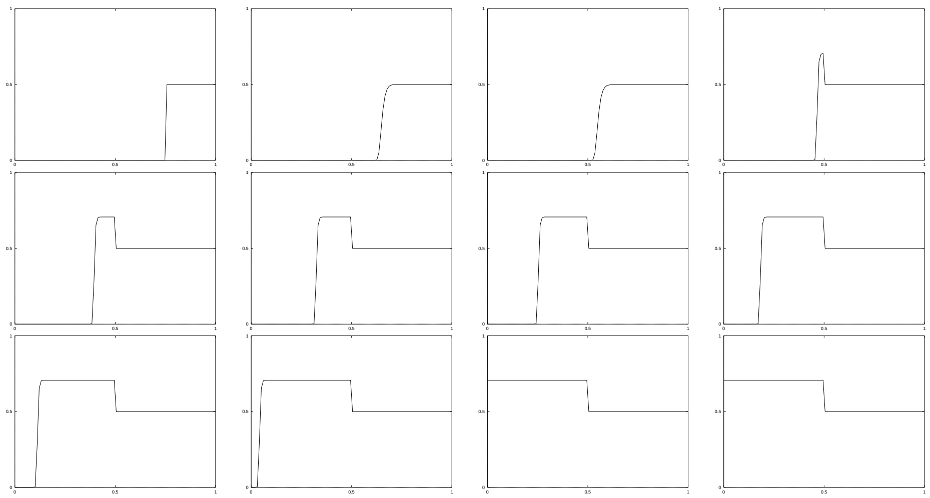

FIGURE 6.4. Solutions of (6.13) with $\beta^1 = 4, \beta^2 = 2, t = 0.2 \cdot k, k = 0 \dots 11$.

boundary conditions assure that we do not get any new maxima or minima at the boundary.

Richards Equation. This equation [Ric31] describes saturated-unsaturated fluid transport in porous media.

$$\frac{\partial \theta(x,u)}{\partial t} - \nabla K(x,u)(\nabla u - \gamma) = 0$$

Here, θ is the saturation, that is the part of the pore space which is filled with the fluid. u is the capillary pressure, that is the difference between the athmospheric pressure and the pressure of the fluid. γ is the gravity.

Main parameters for the soil are the values for the residual saturation $\theta_r(x)$, the size of the pore space $\theta_s(x)$, and the permeability $K_s(x)$ A common heuristic ansatz for the nonlinear functions is the van Genuchten-Mualem model [vG80] which is parametrized by $\alpha(x)$ and $n(x)$:

$$
\begin{aligned}
\theta(x,u) &= \theta_r(x) + \theta_0(x,u)(\theta_s(x) - \theta_r(x))\\
K(x,u) &= K_s(x)K_0(x,u)\\
K_0(x,u) &= \begin{cases} 1, & u \geq 0\\ \left(1-(-\alpha u)^{n-1}\chi(u)^{\frac{1-n}{n}}\right)^2 \chi(u)^{\frac{1-n}{2n}}, & u \geq 0 \end{cases}\\
\theta_0(x,u) &= \begin{cases} 1, & u \geq 0\\ \chi(u)^{\frac{1-n}{n}}, & u \geq 0 \end{cases}
\end{aligned}
$$

where

$$\chi(u) = (1 + (-\alpha p))^n$$

As in section 2, we assume the soil parameters to be piecewise constant.

At the first glance one would try the obvious ansatz for the flux function

$$g_{ij}^m(u,v) = \frac{|\gamma_{ij}|}{h_{ij}} \frac{k^m(u)+k^m(v)}{2}(u - v - \gamma\cdot\mathbf{n}_{ij}) \tag{6.14}$$

If one verifies this against the considerations made so far in this paper, one recognizes that this ansatz is neither isotone nor weakly isotone. The local maximum principle is violated. So one rather should use a stable variant like

$$g_{ij}^m(u,v) = \frac{|\gamma_{ij}|}{h_{ij}} \frac{k^m(u)+k^m(v)}{2}(u - v) - \begin{cases} k^m(u)\gamma\cdot\mathbf{n} & ,\gamma\cdot\mathbf{n}\geq 0\\ k^m(v)\gamma\cdot\mathbf{n} & ,\gamma\cdot\mathbf{n} < 0. \end{cases} \tag{6.15}$$

In [FK97], concerning the maximum principle, the same results have been obtained for a finite volume discretization which slightly differently treats material discontinuities.

Theorem 5.10 in the case of a homogeneous material gives us existence and L^∞ stability for equilibrium flow boundary conditions. Further, any homogeneous node is an equilibrium node, and thus we have a local maximum principle.

In the sequel, we provide two numerical examples which compare the scheme (6.14) with (6.15).

One-dimensional column. The first example is inspired by [FK97], example 1, which resembles a column inflow experiment in a column with two different material layers. The values for the van Genuchten parameters provided here have been fitted by Jan Christian Kaiser from the WASY GmbH.

Let $\Omega = [0, 15] \subset \mathbb{R}_1$ be discretized with 100 grid points. We take an initial value[1] $u_0 = -10.197$. We suppose Dirichlet boundary conditions $u = 0$ at $z = 15$ and $u = -10.197$ at $z = 0$. Regard the following material parameters:

domain	K_s	n	α	θ_s	θ_r
$[0, 7.5]$	$6.06\cdot 10^{-8}$	1.92	50	1	0.083
$[7.5, 15]$	$6.06\cdot 10^{-7}$	1.92	50	1	0.083

Figure 6 clearly shows the difference between both ansatzes. There is *one* non-equilibrium node which naturally violates the local maximum principle. This phenomenon is similar to that in the nonlinear hyperbolic case. Physically, this is a ponding effect where water enters enters material with a more narrow pore space. Using the stable scheme (6.15), we can resemble this effect very well. If, at the other

[1] A more natural choice would be a hydrostatic distribution which is linear in x, but we wanted to stay close to [FK97]

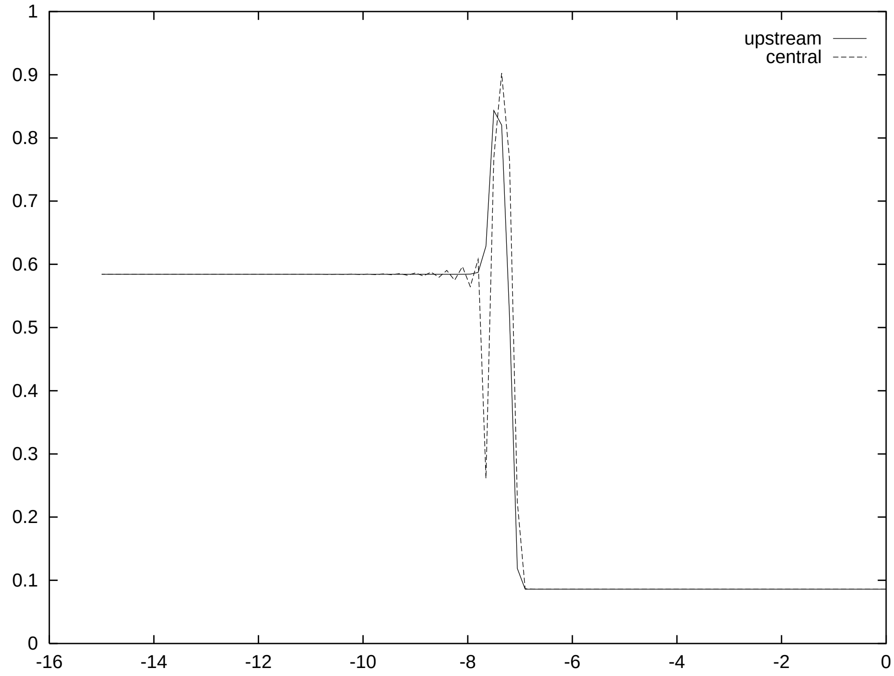

FIGURE 6.5. Solution of the column problem with flux ansatz 6.15 and 6.14, respectively.

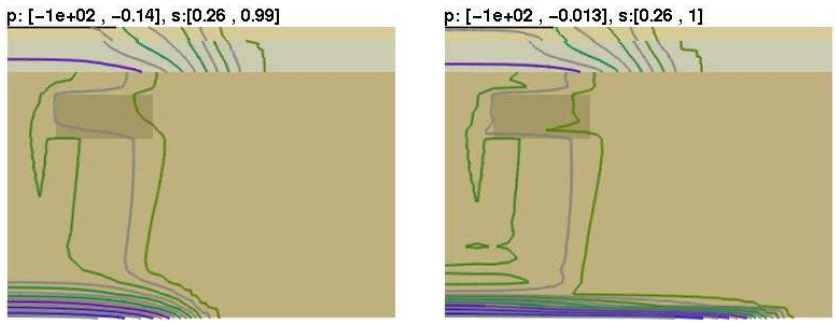

FIGURE 6.6. Solution of the 2D problem with flux ansatz 6.15 and 6.14, respectively.

hand, we use the unstable scheme (6.14), we get "wiggles" which physically don't make any sense. Quite the same result we find for this problem in [FK97].

2D flow. This example is nearly exactly example 2 from [FK97], where we measured the geometry from the the drawings, and created our own grid which is coarser than in their paper.

Let $\Omega = [0, 8] \times [0, 6.5] \subset \mathbb{R}^2$ discretized by a 22×30 grid with lines fitted to the material distribution. Every rectangle in the grid is subdivided into two triangles so that we can use the discretization procedure described in section 4.

Let $u_0 = -1000$ and define an inflow of $\mathbf{j} \cdot \mathbf{n} = 2.3 \cdot 10^{-7}$ on $[0, 2.25] \times 6.5$. At all other parts of the boundary, we assume no flow boundary conditions. We assume the following material distribution:

domain	K_s	n	α	θ_s	θ_r
$[0,8]\times[6.2,6.5]$	$9.15\cdot 10^{-5}$	1.98	3.34	0.37	0.27
$[0,8]\times[5.5,6.2]$	$5.45\cdot 10^{-5}$	1.63	3.63	0.35	0.28
$[0,8]\times[0,5.5]$	$4.8\cdot 10^{-5}$	5	3.45	0.32	0.26
$[1,3]\times[4,5]$	$4.8\cdot 10^{-4}$	5	3.45	0.32	0.26

Again, we are able to establish results similar to those in [FK97]. Some of the artefacts in the unstable example are such that it is not possible to establish from visual evidence that they are part of the flow physics or introduced by the scheme.

Appendix A. Linear Algebra

A good reference for the following results still is [Var62]. Let $\mathbb{M}(n)$ be the set of all real $n\times n$-matrices.

A.1. Definition. For $u=(u_i)_{i=1\dots n}\in\mathbb{R}^n$, let

$$(A.2)\qquad \begin{aligned} ||u||_1 &= \sum_{i=1}^n |u_i| \\ ||u||_\infty &= \max_{i=1}^n |u_i| \end{aligned}$$

A.3. *Remark.* Let $A\in\mathbb{M}(n)$. Then

$$(A.4)\qquad \begin{aligned} ||A||_1 &= \sup_{0\neq u\in\mathbb{R}^n}\frac{||Au||_1}{||u||_1} = \max_{j=1}^n\sum_{i=1}^n |a_{ij}| = ||A^T||_\infty \\ ||A||_\infty &= \sup_{0\neq u\in\mathbb{R}^n}\frac{||Au||_\infty}{||u||_\infty} = \max_{i=1}^n\sum_{j=1}^n |a_{ij}| = ||A^T||_1 \end{aligned}$$

Further, we introduce some notations which are not usual, but we think they are useful in the context of this paper. We note that often there isn't paid attention to the difference between row-wise and column-wise diagonal dominance.

A.5. Definition. Define the following subsets of $\mathbb{M}(n)$

$\mathbb{Z}(n)\subset\mathbb{M}(n)$	nonnegative main diagonal and nonpositive off-diagonal elements (often denoted as $Z^{n\times n}$)
$\mathbb{Z}_c^+(n)\subset\mathbb{Z}(n)$	columnwise weakly diagonally dominant
$\mathbb{Z}_c^{++}(n)\subset\mathbb{Z}_c^+(n)$	columnwise strictly diagonally dominant
$\mathbb{Z}_c^0(n)\subset\mathbb{Z}_c^+(n)$	column sum zero
$\mathbb{Z}_r^+(n)\subset\mathbb{Z}(n)$	row wise weakly diagonally dominant
$\mathbb{Z}_r^0(n)\subset\mathbb{Z}_r^+(n)$	row sum zero
$\mathbb{Z}_r^{++}(n)\subset\mathbb{Z}(n)$	row wise strictly diagonally dominant
$\mathbb{D}(n)\subset\mathbb{M}(n)$	diagonal
$\mathbb{D}^+(n)\subset\mathbb{D}(n)\cap\mathbb{Z}_r^+(n)\cap\mathbb{Z}_c^+(n)$	nonnegative diagonal
$\mathbb{D}^{++}(n)\subset\mathbb{D}^+(n)\cap\mathbb{Z}_r^{++}(n)\cap\mathbb{Z}_c^{++}(n)$	positive diagonal

For a subset $X\subset\{1\dots n\}$ we write $\mathbb{M}(X)$, $\mathbb{Z}(X)$ etc. to denote the sets of matrices corresponding to this subset.

$A=(a_{ij})_{i,j=1\dots n}\in\mathbb{M}(n)$ is called *positive* ($A>0$) if $a_{ij}>0, i,j=1\dots n$ and nonnegative ($A\geq 0$) if $a_{ij}\geq 0, i,j=1\dots n$

A.6. Definition. A Matrix $A \in \mathbb{Z}(n)$ is called (nonsingular) *M-matrix* if $A^{-1} \geq 0$.

A.7. Theorem. *Let* $A \in \mathbb{Z}_r^+(n)$, $D \in \mathbb{D}^{++}(n)$. *Then*
(i). $DA \in \mathbb{Z}_r^+(n)$
(ii). $A + D \in \mathbb{Z}_r^{++}(n)$ *is an M-matrix*
(iii). $||(I + A)^{-1}||_\infty \leq 1$

Proof. See [Fuh98a]. □

A.8. *Remark.* The third part of the theorem comes from the M-criterion of [GR92, RST96], both refer to [Boh81].

A.9. Theorem. *Let* $A \in \mathbb{Z}_c^+(n)$, $D \in \mathbb{D}^{++}(n)$. *Then*
(i). $AD \in \mathbb{Z}_c^+(n)$
(ii). $A + D \in \mathbb{Z}_c^{++}(n)$ *is an M-matrix*
(iii). $||(I + A)^{-1}||_1 \leq 1$

Proof. Apply theorem (A.7) to A^T. □

Appendix B. A mean value theorem

Here, we state a mean value theorem which several times is applied in this paper. We are indebted to L. Recke for the hint to use this type of mean value problem within our considerations.

B.1. Theorem. *Let* $A : \mathbb{R}^n \to \mathbb{R}^n$ *be Gâteaux differentiable for any* $u \in \mathbb{R}^n$ *with the Gâteaux derivative* $A'(u)$. *Assume* $A' \in C(\mathbb{R}^n, \mathbb{M}(n))$. *Then for any* $u, v \in \mathbb{R}^n$, *the operator* $\tilde{A}(u, v)$ *defined by*

$$\tilde{A}(u, v) = \int_0^1 A'(v + \theta(u - v))d\theta \tag{B.2}$$

satisfies $\tilde{A} \in C(\mathbb{R}^n \times \mathbb{R}^n, \mathbb{M}(n))$ *and*

$$A(u) - A(v) = \tilde{A}(u, v)(u - v). \tag{B.3}$$

If $A'(u)$ *independently of* u *belongs to any of the subsets of* $M(n)$ *from definiton A.5, then* $\tilde{A}(u, v)$ *does.*

Proof. The existence of $\tilde{A}$ satisfying (B.3) is proven in [OR70]. The nonzero pattern, the sign pattern and the row/column sum conditions are straightforward. □

Appendix C. Existence and Uniqueness results for systems of nonlinear algebraic equations

For convenience of the reader, we collect here the results of [Fuh98a]. For the notation of a Dirchlet problem, we also refer to that paper, or to section 4.

C.1. Theorem. *Let* $K \in C(\mathbb{R}^n, \mathbb{M}(n))$ *and* $Q \in C(\mathbb{R}^n, \mathbb{R}^n)$. *For any* $u \in \mathbb{R}^n$ *assume*

$$K(u) \in \mathbb{Z}_r^+(n) \tag{C.2a}$$

and

$$||Q(u)||_\infty \leq Q^+ \tag{C.2b}$$

Further, let $B \in C^1(\mathbb{R}^n, \mathbb{R}^n)$ *be a diagonal operator with*

$$||B'(u)||_\infty \geq B^- > 0. \tag{C.2c}$$

Then for any $u_0 \in \mathbb{R}^n$, for any (possibly empty) subset $D \subset \{1 \dots n\}$ and any $u_d \in \mathbb{R}_D$, the Dirichlet problem

$$\begin{aligned} B(u) + K(u)u &= Q(u) + B(u_0) \\ u|_D &= u_d \end{aligned} \tag{C.3}$$

has a solution $\hat{u}$ with

$$||\hat{u}_I||_\infty \leq \frac{Q^+}{B^-} + ||u_{0,I}||_\infty + ||u_d||_\infty, \tag{C.4}$$

assuming $||u_d||_\infty = 0$ for $D = \emptyset$.

C.5. Theorem. *Let $K \in C(\mathbb{R}^n, \mathbb{M}(n))$. For any $u \in \mathbb{R}^n$ assume*

$$K(u) \in \mathbb{Z}_r^+(n) \tag{C.6a}$$

Further, let $B \in C^1(\mathbb{R}^n, \mathbb{R}^n)$ be a diagonal operator with

$$B'(u) \geq B^- \geq 0. \tag{C.6b}$$

Then for any $u_0 \in \mathbb{R}^n$, for any (possibly empty) subset D and any $u_d \in \mathbb{R}_D$, the Dirichlet problem

$$\begin{cases} B(u) + K(u)u &= B(u_0) \\ u|_D &= u_d \end{cases} \tag{C.7}$$

has a solution $\hat{u}$ with

$$||\hat{u}_I||_\infty \leq ||u_{0,I}||_\infty + ||u_d||_\infty \tag{C.8}$$

assuming $||u_d||_\infty$ for $D = \emptyset$.

C.9. Theorem. *Let $A \in C(\mathbb{R}^n, \mathbb{R}^n)$ have a Gâteaux derivative $A'(u) \in \mathbb{Z}_c^+(n)$ continuously depending on u. Further, let $B \in C^1(\mathbb{R}^n, \mathbb{R}^n)$ be an isotone diagonal homeomorphism. Let $Q \in C(\mathbb{R}^n, \mathbb{R}^n)$ be bounded with $||Q(u)||_1 \leq Q^+$. Then for $u_0 \in \mathbb{R}^n$, for any (possibly empty) subset $D \subset \{1 \dots n\}$, and $u_d \subset \mathbb{R}_D$, the Dirichlet problem*

$$\begin{cases} B(u) + A(u) &= Q(u) + B(u_0) \\ u|_D = u_d \end{cases} \tag{C.10}$$

has a solution $\hat{u}$ which in the case $D = 0$ admits the estimate

$$||B(u)||_1 \leq ||B(u_0)||_1 + Q^+ + ||A\big(B^{-1}(0)\big)||_1. \tag{C.11}$$

C.12. Theorem. *Assume that the conditions of theorem C.11 are fulfilled, that $Q = 0$ and that for any $u \in \mathbb{R}^n$, $B'(u) \in \mathbb{D}^{++}(n)$. Then the solution of* (C.10) *is unique.*

References

[AS55] D. N. ALLEN AND R. V. SOUTHWELL, *Relaxation methods applied to determine the motion, in two dimensions, of a viscous fluid past a fixed cylinder*, Quart. J. Mech. and Appl. Math., 8 (1955), 129–145.

[Bän95] E. BÄNSCH, *Numerical experiments with adaptivity for the porous medium equation*, Acta Math. Univ. Comenianae, LXIV (1995) 2, 157–172.

[Boh81] E. BOHL, *Finite Modelle gewöhnlicher Randwertaufgaben*, Teubner, Stuttgart, 1981.

[BW93] L. A. BAUGHMAN AND N. J. WALKINGTON, *Co-volume methods for degenerate parabolic problems*, Numer. Math., 64 (1993) 1, 45–67.

[Cia78] P. G. CIARLET, *The finite-element-method for elliptic problems*, North Holland, Amsterdam, 1978.

[EGH97a] R. EYMARD, T. GALLOUËT, AND R. HERBIN, *Finite volume methods*, Tech. Rep. Prépublication no 97-19, LATP, UMR 6632, Marseille, 1997. to appear in Handbook of Numerical Analysis, P.G. Ciarlet, J.L. Lions, eds.

[EGH97b] R. EYMARD, M. GUTNIC, AND D. HILHORST, *The finite volume method for the Richards equation*, tech. rep., Université de Paris-Sud, Orsay, France, 1997. in preparation.

[EO81] B. ENGQUIST AND S. OSHER, *One-sided difference approximations for nonlinear conservation laws*, Math. Comp., 36 (1981) 154, 321–351.

[FK97] P. A. FORSYTH AND M. C. KROPINSKI, *Monotonicity considerations for saturated–unsaturated subsurface flow*, SIAM Journal on Scientific Computing, 18 (1997) 5, 1328–1354.

[FKTT97] P. FROLKOVIČ, P. KNABNER, C. TAPP, AND K. THIELE., *Adaptive finite volume discretization of density driven flows in porous media.*, tech. rep., IAM Uni Erlangen, 1997. Preprint no. 220; Proc.INRIA Rocquencourt: Transport de contaminants multiespeces en milieux poreux 2. - 5. 6. 1997.

[Fro98] P. FROLKOVIČ, *Maximum principle and local mass balance for numerical solutions of transport equation coupled with variable density flow*, Acta Mathematica Universitatis Comenianae, 67 (1998) 1, 137–157.

[Fuh97a] J. FUHRMANN, *Numerical solution schemes for nonlinear diffusion problems based on newton's method*, in ALGORITMY '97, 14th Conference on Scientific Computing West Tatra Mountains - Zuberec, Slovakia, September 2-5, 1997, Proceedings of contributed papers and posters, K. M. A. Handlovičovaá, M.Komorníková, ed., Slovak Technical University, Bratislava, 1997.

[Fuh97b] J. FUHRMANN, *On numerical solution methods for nonlinear parabolic problems*, in Modeling and Computation in Environmental Sciences. Proceedings of the First GAMM-Seminar at ICA Stuttgart, October 12-13,1995, R. Helmig, W. Jäger, W. Kinzelbach, P. Knabner, and G. Wittum, eds., vol. 59 of Notes on Numerical Fluid Mechanics, Braunschweig, 1997, Vieweg, 170–180.

[Fuh98a] J. FUHRMANN, *Existence and uniqueness of solutions of certain systems of algebraic equations with off diagonal nonlinearity*, tech. rep., WIAS Berlin, 1998. Preprint No. 425, submitted.

[Fuh98b] — —, *Numerical solutions of viscous conservation laws*, `http://www.wias-berlin.de/porous-media-adaptive/conlaw.html`, 1998.

[GG96] H. GAJEWSKI AND K. GÄRTNER, *On the discretization of van Roosbroeck's equations with magnetic field*, Z. Angew. Math. Mech., 76 (1996) 5, 247–264.

[GR92] C. GROSSMANN AND H.-G. ROOS, *Numerik partieller Differentialgleichungen*, Teubner Studienbücher Mathematik, Teubner, Stuttgart, 1992.

[Hac89] W. HACKBUSCH, *On first and second order box schemes*, Computing, 41 (1989) 4, 277–296.

[Hei87] B. HEINRICH, *Finite difference methods on irregular networks*, vol. 33 of Mathematical Research, Akademie-Verlag, Berlin, 1987. A generalized approach to second order elliptic problems.

[Il'69] A. M. IL'IN, *A difference scheme for a differential equation with a small parameter multiplying the second derivative*, Matematičeskije zametki, 6 (1969) 2, 237–248.

[J.B98] J.BEY, *Finite-Volumen- und Mehrgitterverfahren für elliptische Randwertprobleme*, Advances in Numerical Mathematics, Teubner-Verlag, Stuttgart, 1998.

[Krö97] D. KRÖNER, *Numerical Schemes for Conservation Laws*, Advances in Numerical Mathematics, Wiley Teubner, 1997. ISBN 3-519-02720-8.

[Mac53] R. H. MACNEAL, *An asymmetrical finite difference network*, Quart. Math. Appl., 11 (1953), 295–310.

[OR70] J. M. ORTEGA AND W. C. RHEINBOLDT, *Iterative solution of nonlinear equations in several variables*, Academic Press, New York, 1970.

[Ric31] L. A. RICHARDS, *Capillary conduction of liquids through porous mediums*, Physics, 1 (1931), 318–333.

[RST96] H.-G. ROOS, M. STYNES, AND L. TOBISKA, *Numerical Methods for Singularly Perturbed Differential Equations*, no. 24 in Springer Series in Computational Mathematics, Springer, 1996.

[SG69] D. L. SCHARFETTER AND H. K. GUMMEL, *Large signal analysis of a silicon Read diode*, IEEE Transactions on Electron Devices, 16 (1969), 64–77.

[Sto86] G. STOYAN, *On maximum principles for monotone matrices*, Lin.Alg.Appl, 78 (1986), 147–161.

[Van96] R. VANSELOW, *Relations between FEM and FVM applied to the Poisson equation*, Computing, 57 (1996) 2, 93–104.

[Var62] R. S. VARGA, *Matrix iterative analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1962.

[vG80] M. T. VAN GENUCHTEN, *A closed-form equation for predicting the hydraulic conductivity of unsaturated soils*, Soil Sci. Soc. Amer. J., 44 (1980), 892–898.