A positive spatial advection scheme on unstructured meshes for tracer transport

P. Wilders\textsuperscript{a,}\textsuperscript{*}, G. Fotia\textsuperscript{b}

\textsuperscript{a}Department of Information Technology and Systems, Delft University, P.O. Box 5031, 2600 GA Delft, Netherlands
\textsuperscript{b}Scientific Research Division, CRS4, VI Strada OVEST Z1. Macchiareddu, 09010 UTA Cagliari, Italy

Received 29 September 2000

Abstract

Modelling tracer transport (leading to a single linear advection–diffusion equation) for realistic data provides a challenging task with respect to the robustness of the underlying numerical procedures. In this paper, we contribute at this point by formulating a positive spatial advection scheme for unstructured triangular meshes. The proof of positivity is presented in detail, using an elementary classification. It is shown that with a careful reconstruction procedure and a moderate demand towards the grid a positive advection scheme is obtained. Next, a brief discussion is given on how we implement this scheme in combination with an implicit time-stepping procedure. As a numerical example, we discuss tracer transport in a strongly heterogeneous porous medium. © 2002 Elsevier Science B.V. All rights reserved.

1. Introduction

The study of contaminants or tracers in subsurface or surface environmental problems is an important task in environmental engineering. Numerical computations are important, among others because measurements are difficult to obtain. These computations are done in a wide variety of spatial domains with complex geometries and on time scales of several months or more. In [18] a flexible simulation tool has been proposed for tracer transport, based upon triangular finite volumes in combination with implicit time stepping. Here, the spatial advection scheme is based upon a variant of the JST-scheme [9]. This is a central scheme with artificial diffusion explicitly added. In this paper we enhance the robustness of the solver further by adding a spatial advection scheme based upon upwinding and by expanding this scheme towards hybrid grids.

In this paper, the chosen strategy for obtaining a positive advection scheme is limited reconstruction, e.g., [11], a popular technique, which in essence originates from the work of Van Leer, e.g., [6]. The technique is based upon the theorem that a second-order scheme can be obtained from a...
first-order scheme, by taking its numerical flux function and by replacing some basic unknowns by higher-order reconstructed values. This reconstruction involves some gradients and it is well-known that it is necessary to limit the gradients in order to prevent oscillatory behavior. For triangular grids the most popular way of limiting is strongly geometrical, a good example is the Barth/Jespersen limiter [2,1]. For structured grids promising results have been obtained with a procedure, which in essence dates back to Sweby [15]. Here, two gradients are weighted using a more analytical approach. For time accurate computations, it is preferable if the switch functions, used to limit, are smooth functions of its parameters. In literature only few remarks are made towards this issue and nearly all developed limiting functions are nondifferentiable. Our goal is to explore a differentiable limiting function. Against this background, we prefer to start from the Sweby-type procedure. In the first, more geometrical, procedure it is much more difficult to formulate smooth limiting functions.

We develop the new spatial advection scheme rather carefully for triangular grids and we will present all relevant details that are necessary for the proof of positivity. At this point our approach is basic and based upon a generalization of the notion of a $K$ matrix [17]. A similar approach has been chosen in [3]. As is often done, we compute the gradients that are necessary for reconstruction with the aid of the Green–Gauss procedure. A careful choice of the integration volumes needed in this procedure enables us to give an elementary proof of positivity. With this choice we follow [12] and, similar to this work, we end up with a modest requirement towards the grids that can be dealt with.

The outline of this paper is as follows: In Section 2 we present the governing equations. Section 3 is devoted to the formulation of a positive advection scheme via the notion of an advective $K$-approximation. In Section 4 the complete reconstruction procedure is described and it is proven that this reconstruction procedure leads to an advective $K$-approximation. Section 5 discusses some issues related to the time-integration procedure, in our case the trapezoidal rule. In Section 6 we propose a simple strategy for extending the spatial advection scheme towards hybrid grids. Section 7 presents a numerical example and, finally, we end with some remarks in Section 8.

### 2. Governing equations

We consider a scalar conservation law of the form

$$
\varphi \frac{\partial c}{\partial t} + \nabla \cdot [vf(c) - D\nabla c] = 0, \quad x \in \Omega \subset \mathbb{R}^2, \quad t > 0.
$$

(2.1)

In the case of a tracer (or single phase miscible displacement) there holds

$$
f(c) = c.
$$

(2.2)

In this contribution the emphasis is on the advective terms, we assume that

$$
D = 0.
$$

(2.3)

Moreover, the coefficient $\varphi(\varphi > 0)$ and $v$ are assumed to be time-independent, with $v$ satisfying

$$
\nabla \cdot v = 0.
$$

(2.4)

The domain $\Omega$ is covered with a triangular grid with triangles $\Delta$. The area of $\Delta$ is denoted with $|\Delta|$, the boundary $\partial \Delta$ is composed of three edges $e$ of length $|e|$. A cell-centered finite volume
approach is adopted. Integration of (2.1) over $\Delta$, applying Green’s formula in combination with appropriate integration rules leads to

$$\varphi_{\Delta} \frac{dc_{\Delta}}{dt} + \sum_{e \in \partial \Delta} h_{e,\Delta}(c_{e}^{L}, c_{e}^{R}) = 0.$$  \hspace{1cm} (2.5)

Here, $(\cdot)_{\Delta}$ stands for taking centroid values, $c_{e}^{L}$ and $c_{e}^{R}$ are estimated values of $c$ on the edge $e$ and $h$ is a numerical flux function, for which, following [4], we take the Enquist–Osher function. In the present case (2.2) (a single linear flux) we obtain

$$h_{e,\Delta}(c_{e}^{L}, c_{e}^{R}) = \frac{U_{e,\Delta} + |U_{e,\Delta}|}{2} c_{e}^{L} + \frac{U_{e,\Delta} - |U_{e,\Delta}|}{2} c_{e}^{R}.$$  \hspace{1cm} (2.6)

Here, $U_{e,\Delta} = v_{e} \cdot n_{e,\Delta}$ with $v_{e}$ the edge value of the velocity $v$ and $n_{e,\Delta}$ the (infinitesimal) outward normal on $e$ with length $|e|$. The numerical flux on edges is evaluated using the edge molecule depicted in Fig. 1. This enables a flexible edge-based datastructure with an efficient edge-based loop for computing the numerical fluxes.

3. Positive advection

Taking (2.5) together for all triangles $\Delta$, we end up with the semidiscrete system

$$L \frac{dc}{dt} = - F(c,v).$$  \hspace{1cm} (3.1)

The matrix $L$ is a diagonal matrix containing the cell values of the porosity $\varphi$ multiplied with the area of the cell.

We refer to the rhs of (3.1) as an *advective $K$-approximation* if there exists a matrix function $\tilde{Q} = (\tilde{q}_{jk})$, $\tilde{q}_{jk} = \tilde{q}_{jk}(c,v)$, with

$$\tilde{q}_{jk} \leq 0, \quad j \neq k \quad \text{and} \quad \sum_{k} \tilde{q}_{jk} = 0 \quad \text{for all} \ c,$$

$$\hspace{1cm} (3.2)$$

Fig. 1. Edge molecule for the discretization of the advective and viscous flux on $e$. 

such that
\[ F_j(c, v) = \sum_k \hat{q}_{jk} c_k \quad \text{for all } j \text{ in the interior.} \quad (3.3) \]

Note that a discussion of the boundary treatment is omitted. The terminology is consistent with [17].

If the entries \( \hat{q}_{jk} \) are independent of \( c \), then the (constant) matrix \( \hat{Q} \) is a \( K \)-matrix. The relation (3.2), (3.3) are strong demands, implying some significant properties. To mention a few: a steady state is monotonic if the input is monotonic [14], local minima are nondecreasing and local maxima are nonincreasing [8], a nonnegative input implies a nonnegative solution [7,10].

With reference to the numbering in Fig. 1, the choice \( c_L = c_1, c_R = c_2 \) in (2.6) leads to the classical first-order upwind scheme. It is known, e.g., [11], that the use of limited linear reconstructed values for \( c_L \) and \( c_R \) leads to an improved accuracy. In the case of (2.6) a sufficient condition for an advective \( K \)-approximation is that for all triangles \( \Delta \) and associated edges \( e \in \partial \Delta \) there holds
\[ \frac{c_R - c_1}{c_2 - c_1} \geq 0 \quad (3.4) \]
and
\[ [c^L - c_1 = 0] \text{ or } \left[ \frac{c^L - c_1}{c_1 - c_k} \geq 0 \text{ for at least one } k \in \{2,3,4\} \right]. \quad (3.5) \]

The proof of this statement has in essence been given in [12]. As a consequence of (2.4) we assume that
\[ \sum_{e \in \partial \Delta} U_{e, \Delta} = 0, \quad (3.6) \]
which implies
\[ \sum_{e \in \partial \Delta} h_{e, \Delta}(c_1, c_1) = 0. \quad (3.7) \]

Thus,
\[ \sum_{e \in \partial \Delta} h_{e, \Delta}(c^L, c^R) = \sum_{e \in \partial \Delta} \zeta^R_{e, \Delta} + \sum_{e \in \partial \Delta} \zeta^L_{e, \Delta} \quad (3.8) \]
with
\[ \zeta^R_{e, \Delta} = h_{e, \Delta}(c^L, c^R) - h(c^L, c_1), \quad \zeta^L_{e, \Delta} = h_{e, \Delta}(c^L, c_1) - h_{e, \Delta}(c_1, c_1). \quad (3.9) \]

A straightforward computation shows that
\[ \zeta^R_{e, \Delta} = \begin{cases} -U_{e, \Delta} \frac{c^R - c_1}{c_2 - c_1} (c_1 - c_2), & U_{e, \Delta} < 0, \\ 0 & \text{elsewhere,} \end{cases} \quad (3.10) \]
\[ \zeta^L_{e, \Delta} = \begin{cases} U_{e, \Delta} \frac{c^L - c_1}{c_1 - c_k} (c_1 - c_k), & U_{e, \Delta} \geq 0, \quad c^L - c_1 \neq 0, \\ 0 & \text{elsewhere.} \end{cases} \quad (3.11) \]

From (3.8), (3.10) and (3.11) it follows directly that (3.2) and (3.3) are implied by (3.4) and (3.5).
Equivalent with (3.4) is
\[
\frac{c_e^L - c_1}{c_2 - c_1} \leq 1
\] (3.12)
for all triangles \(\Delta\) and \(e \in \partial \Delta\), because \(c_e^R = c_e^L\) with \(\Delta'\) a neighboring triangle (note that in the local numbering system of \(e \in \partial \Delta\) the indices 1 and 2 change order).

4. Reconstruction

Let \(P_a\) and \(P_b\) be two points in \(\mathbb{R}^2\). With \(t_{ab}\) we denote the vector from \(P_a\) to \(P_b\) and with \(n_{ab}\) the normal, pointing to the right, on the segment between \(P_a\) and \(P_b\). The length of the vectors \(t_{ab}\), \(n_{ab}\) is equal to the length of the segment \(P_aP_b\). In the following we use the numbering from Fig. 1. We assume that the triangulation is such that (called a TVD triangulation in [12])
\[
n_{13}.t_{1e} \geq 0 \quad \text{and} \quad n_{41}.t_{1e} \geq 0. \tag{4.1}
\]
Here, \(t_{1e}\) denotes the vector pointing from cell center \(P_1\) to the midpoint of the edge. Note that (4.1) must hold for each triangle \(\Delta\) and associated edge \(e \in \partial \Delta\). For Fig. 1 this means that a similar property holds on the other side of the edge \(e(e \in \partial \Delta')\).

Let \(\nabla_1\) be an approximation of the gradient of \(c\) in cell center \(P_1\). We define
\[
\delta_{e1} = \nabla_1 . t_{1e}, \quad \delta_{21} = c_2 - c_1, \tag{4.2}
\]
\[
c_e^L = c_1 + \psi \delta_{e1}. \tag{4.3}
\]
For the computation of \(\nabla_1\) we apply Green–Gauss reconstruction on the triangle \(P_1P_3P_4\). Some computations show that
\[
\delta_{e1} = [(n_{41}.t_{1e})(c_1 - c_3) + (n_{13}.t_{1e})(c_1 - c_4)]/\Delta_{134} \tag{4.4}
\]
with \(\Delta_{134} = t_{13}.n_{14}\). Due to the counterclockwise ordering there holds \(\Delta_{134} > 0\) (\(\Delta_{134}\) is twice the area of \(P_1P_3P_4\)).

We consider a limiting function \(\psi\) of the form
\[
\psi = \psi(r), \quad r = \frac{\delta_{21}}{\delta_{e1}}. \tag{4.5}
\]
From (4.3) it easily follows that
\[
\frac{c_e^L - c_1}{c_2 - c_1} = \frac{\psi}{r}, \tag{4.6}
\]
From (3.5), (3.12), (4.3), (4.4) and (4.6) in combination with the mesh restriction (4.1) it follows that
\[
0 \leq \psi \leq r \tag{4.7}
\]
is a sufficient condition for an advective \(K\)-approximation.
An example of a limiting function \( \psi \) satisfying (4.7) is the modified Van Leer limiter used in [3]. However, we like to have a differentiable limiter. A good candidate is the \( R - 1 \) limiter \( (R - \kappa \text{ limiter with } \kappa = 1) \) from [20]:

\[
\psi = 2 \frac{(r + |r|)r}{(1 + r)^2}.
\]

(4.8)

It is easy to check for (4.7). On the interval \([0,1]\) \( \psi \) is close to \( r \), except near \( r = 0 \).

For equidistant meshes it is known, e.g., [20], that condition (4.7) can be relaxed and that (4.8) does not present an optimal compromise between accuracy and monotonicity. It is an open question whether one can relax (4.7) for unstructured meshes [3,10].

Substitution of (4.3), (4.4) in (2.5), (2.6) shows that the semi-discrete system (3.1) takes the form

\[
L \frac{dc}{dt} = -Qc, \quad Q = (q_{jk}), \quad \tilde{Q} = Q_1 + Q_\psi.
\]

(4.9)

\( Q_1 \) is the (constant) matrix associated with the first-order upwind scheme. \( Q_\psi \) corresponds with the limited anti-diffusion and has entries depending upon \( c \) (through the limiting function \( \psi \) only). To be precise \( Q_\psi \) is the matrix that one reads of directly from the second term on the rhs of (4.3) in combination with (4.4). For example, from the term \( \psi((n_4, \tau_e))(c_1 - c_3) \) the part \( \psi((n_4, \tau_e)) \) is shifted to the matrix elements of \( Q_\psi \). Assuming \( \psi = 1 \) turns \( Q_\psi \) into a constant matrix associated with a fully decentral scheme (due to the bias of triangle \( P_1P_3P_4 \)). Note that \( Q \) and \( \tilde{Q} \) differ in general. Here, \( \tilde{Q} \) is the matrix constructed in Section 3. In our approach the matrix \( \tilde{Q} \) is only used to study the scheme analytically.

5. Time integration

The trapezoidal rule is used for the time integration of the semidiscr
tere system (3.1). This leads to

\[
\frac{1}{\tau_n} Lc^{n+1} + \frac{1}{2} F^{n+1} = b^n, \quad b^n = \frac{1}{\tau_n} Lc^n - \frac{1}{2} F^n.
\]

(5.1)

Here, \( \tau_n \) denotes the time step. We call (5.1) positive if [10]

\[
c^n \geq 0 \implies c^{n+1} \geq 0.
\]

(5.2)

The linear invariance property of the employed numerical fluxes leads to \( F(\chi c + \beta, \nu) = \chi F(c, \nu) \). This implies that (5.2) is equivalent with the absence of general under- and overshoots [7].

There exists a matrix function \( \tilde{Q} \) with property (3.2) such that \( F = \tilde{Q}c \). As a consequence, (5.1) is equivalent with

\[
\tilde{A}c^{n+1} = b^n, \quad \tilde{A} = \frac{1}{\tau_n} L + \frac{1}{2} \tilde{Q}^{n+1}.
\]

(5.3)

Property (3.2) implies (independently of \( \tau_n \)):

\[
\tilde{a}_{jk} \leq 0, \quad j \neq k \quad \text{and} \quad \sum_k \tilde{a}_{jk} > 0, \quad \text{for all } c.
\]

(5.4)
From this it easily follows that
\[ c^n \geq 0 \quad \Rightarrow \quad b^n \geq 0 \] (5.5)
is a sufficient condition for positivity.

There holds
\[ b^n = \tilde{B} c^n, \quad \tilde{B} = \frac{1}{\tau_n} L - \frac{1}{2} \tilde{Q}^n. \] (5.6)

Using (3.2) this expression shows that (5.5) is satisfied for \( \tau_n \) sufficiently small. One can work out this further to obtain a condition relating the time step \( \tau_n \) to the main diagonal elements of \( \tilde{Q} \). This condition can be formulated as a restriction of the Courant number \( \eta \). For example, in the case of the first-order upwind scheme this procedure leads to \( \eta \leq 2 \), e.g., [20]. Our numerical experiments indicate that computations with larger Courant numbers are visible as well.

For the implementation of (5.1) we substitute \( F = Qc \), see (4.9). This leads to
\[ A^{n+1} c^{n+1} = b^n, \quad A^{n+1} = \frac{1}{\tau_n} L + \frac{1}{2} [Q_1 + Q^{n+1}_\psi]. \] (5.7)

The vector \( Q^n c^n \) is computed and a time step \( \tau_n \) is chosen for which
\[ b^n = \left[ \left( \frac{1}{\tau_n} L - \frac{1}{2} Q^n \right) c^n \right] \geq 0. \]
Next, (5.7) has to be solved. In this section, we omit the time step index \( n \) in further notation. The following iterative procedure, with \( c^{(0)} = c^n \), is introduced:
\[ J^{(k)} c^{(k+1)} = b + [J^{(k)} - A^{(k)}] c^{(k)}. \] (5.8)

It is common practice to perform only one iteration with (5.8) and the final result is a linearly implicit variant of the trapezoidal rule. Second-order time accuracy is only preserved if \( J \) is the full Jacobian of the function \( Ac \).

Several choices for \( J \) are possible. Often \( J = (1/\tau_n)L + \frac{1}{2} Q_1 \), is taken (defect-correction technique), e.g., [20,10]. It is also possible to rewrite the equation using \( \tilde{Q} \) instead of \( Q \) and to take \( J = (1/\tau_n)L + \frac{1}{2} \tilde{Q} \) [19,10]. This approach leads to positivity at all possible stages of (5.8). We take
\[ J = \frac{1}{\tau_n} L + \frac{1}{2} Q = \frac{1}{\tau_n} L + \frac{1}{2} [Q_1 + Q_\psi]. \] (5.9)

Substitution in (5.8) leads to
\[ J^{(k)} c^{(k+1)} = b. \] (5.10)

As been stated, only a single iteration with (5.10) is performed. The resulting linearly implicit scheme is nearly second-order accurate in time; (5.9) is close to the full Jacobian. Only variations via the limiting functions are not taken into account, thus, avoiding a detailed analysis of these variations. Due to the fact that the levels on the lhs of (5.10) differ, it is not possible to use the arguments of Section 4. This means that positivity is not guaranteed for a fixed number of iterations in (5.10). We find this acceptable for the moment. The resulting linearly implicit scheme is conservative as a time–space scheme. We mention this, because literature may cause some confusion at this point.

Finally, system (5.10) has been solved using ILU-preconditioned Bi-CGSTAB. The matrix \( J \) is a sparse matrix with a maximum of 10 nonzero elements in each row (10-point cell molecule).
Furthermore, $J$ reads $J = J_1 + J_\psi$, $J_1 = (1/\tau_n) L + \frac{1}{2} Q_1$, $J_\psi = \frac{1}{2} Q_\psi$, see (5.9). The matrix $J_1$ is a compact $M$-matrix with good iterative properties. The matrix $J_\psi$ corresponds with a fully decentral approximation. Furthermore, the function $J(c)c$ is known to be derived from an advective $K$-approximation, which implies a $K$-matrix property for this function in a transformed sense. All of this indicates that (5.10) presents a good starting point for iterative solution methods.

6. Hybridization

In case of hybrid meshes, we may encounter an $N$-point molecule as for example in Fig. 2. The edge under consideration is denoted by $e$. Cells 1 and 2 are the cells left and right of edge $e$ respectively. $N_i$ is the number of edges of cell $i$. We want to determine the advective $\mathbf{u}$ on edge $e$ thereby using a generalization of the discretization schemes of the triangular grid case. Two approaches are possible. We can design numerical flux functions that use, apart from the primary cells 1 and 2, all secondary cells $3, \ldots, N_i + 1$ and $N_i + 2, \ldots, N_1 + N_2$, or we can select two secondary cells on both sides of the edge and use these cells for the numerical functions of the triangular case as described in Section 3. We have chosen for the second option, because this option is simpler to implement.

We are left with the choice of selecting the secondary cells. At the moment we have worked out this only for quadrilateral meshes. Consider the situation presented in Fig. 3. First of all, our choice must be such that (4.1) is satisfied. This condition can be expressed by stating that the two secondary points to be chosen on the left side of the edge $e$ must lie on different sides of the line connecting the cell center 1 with the midpoint of the edge $e$, thus leaving two possible choices. Next, we consider for both cases the triangle used in the Green–Gauss procedure for computing the gradient. For accuracy reasons we want to avoid excessive stretching, i.e., we prefer small angles. The outcome is that the cell centers 4 and 5 are chosen to complete the molecule on the left side of the edge $e$. For the other side a similar reasoning holds.
7. A numerical example

We consider a test problem from [5,18], i.e. the quarter of five spots with strongly heterogeneous realistic permeability data. The velocity field, illustrated in Fig. 5, has been computed from the pressure equation with a mixed finite element code. Fig. 4 presents the computed tracer concentration at 0.6 PV using the previous version of our code with the JST-scheme. For obtaining the result in Fig. 6, we have used Durlofski’s limited reconstruction [4] in combination with the Shu–Osher time-stepping procedure [13]. A good resemblance can be observed. However, the JST-solution is more diffusive. Fig. 7 shows the tracer concentration such as computed with the spatial advection scheme described in this paper in combination with the trapezoidal rule. The results obtained in
Figs. 6 and 7 are quite similar. The computations so far have been done with a maximal Courant number close to one. In Fig. 8, we have extended Fig. 7 by doing the computations with a maximal Courant number of the order 30. It can be seen that the quality of the solution does not diminishes. This is somewhat surprising and asks for a thorough investigation of the high Courant number case.

Finally, for detailed investigations the above computed fronts may still be too spreaded. One way out is to refine the spatial mesh. In Fig. 9, we plot the concentration obtained on a mesh with approximately 51000 triangles and with a maximal Courant number of the order 30. An excellent resolution can now be observed.
8. Final remarks

In transport computations, a high degree of flexibility is wanted in choosing the structure of the computational grid. This can for example be dictated by the complex geometry of the physical domain or by the demand to be able to model very local processes such as contaminant releases. To increase the flexibility, we focus in our work on the application of hybrid grids. As a first step, we have developed a triangular solver with quite acceptable results [5,18]. Since then, we have encountered some new applications with stricter demands towards positivity. Therefore, we have developed the present positive advection scheme. Our basic goal was to develop a scheme suitable for time-accurate computations. In the present contribution, we have developed such a scheme and
we have given a proof of its positivity. The numerical example, presented in Section 7, indicates a good performance.

A fundamental step in our proof of positivity is the specific choice of the Green–Gauss volume for the computation of the directional gradient in the reconstruction procedure, see (4.1), (4.4). For the second directional gradient needed in the Sweby procedure, we have chosen the simplest central approximation, see (4.2). A next step will be to investigate whether the resulting approximation is linearity preserving. It might be that the chosen central approximation is too simple and needs to be adapted, for example similar to [3].

With respect to the use of hybrid grids we want to remark that we have started to study both triangular grids and quadrilateral grids for a problem concerning tracer transport in a coastal region [16]. So far, the results look promising and we intend to extend these computations towards a hybrid grid consisting of a mix of triangles and quadrilaterals.

References