STABILITY, MONOTONICITY, MAXIMUM AND MINIMUM PRINCIPLES, AND IMPLEMENTATION OF THE VOLUME CORRECTED CHARACTERISTIC METHOD*

TODD ARBOGAST† AND WEN-HAO WANG‡

Abstract. We consider the volume corrected characteristics-mixed method (VCCMM) for tracer transport problems. The volume correction adjustment maintains the local volume conservation of bulk fluids and the numerical convergence of the method. We discuss some details of implementation by considering the scheme from an algebraic point of view. We show that the volume correction adjustment is important for stability and necessary for the monotonicity and the maximum and minimum principles of the method. We also derive a relatively weaker stability property for the uncorrected characteristic-mixed method (CMM). Some numerical experiments of a quarter “five-spot” pattern of wells are given to verify our theoretical results and compare the concentration errors of VCCMM and CMM due to random perturbations set up in the computation of the algorithm. More numerical tests, including one related to long-time nuclear waste storage, are given to compare VCCMM with CMM and Godunov’s method, showing that VCCMM exhibits no overshoots or undershoots and less numerical diffusion.

Key words. transport, advection, method of characteristics, Lagrangian method, ELLAM, local conservation, fully conservative

AMS subject classifications. 35L65, 65M12, 65M25, 76S05

DOI. 10.1137/100788689

1. Introduction. We consider the problem of incompressible dilute miscible tracer transport, as might arise in a porous medium application (or similarly in a shallow water or atmospheric system). On a confined and bounded domain \( \Omega \subset \mathbb{R}^d \), a dilute miscible tracer of concentration \( c(x,t) \) in an incompressible bulk fluid moving according to the velocity field \( u(x,t) \) satisfies the advection system

\[
\begin{align*}
\nabla \cdot u &= q \quad \text{in } \Omega \times J, \\
(\phi c)_t + \nabla \cdot (cu) &= q_c := c_I q^+ + c q^- \quad \text{in } \Omega \times J, \\
u \cdot \nu &= 0 \quad \text{on } \partial \Omega \times J, \\
c(x,0) &= c_0(x) \quad \text{in } \Omega,
\end{align*}
\]

where \( J = (0,\infty) \) is the time interval, \( q = q(x,t) \) represents wells, i.e., isolated external sources (injection wells) \( q^+ = \max\{q,0\} \geq 0 \) and sinks (production wells) \( q^- = q - q^+ \leq 0 \), \( c_I = c_I(x,t) \) is the injected concentration, \( \phi = \phi(x) \in [\phi_*,1] \) \( (\phi_*>0) \) is the storage factor of the medium called porosity, \( \phi_0 = \phi_0(x) \) is the initial concentration, subscript \( t \) is time partial differentiation, and \( \nu \) is the outward unit...
normal vector with respect to $\partial \Omega$. The meaning of a *dilute* tracer is that we assume $c$ does not change the overall velocity $\mathbf{u}$.

The characteristic-mixed method (CMM) \cite{ArbogastChilakapatiWheeler1992, ArbogastChilakapatiWheeler1993} was introduced by Arbogast, Chilakapati, and Wheeler in 1992. It is based on the transport not of single points or fluid particles, but rather on the fluid mass in entire regions. This mass is transported along the characteristic curves of the transport equation. By design, it automatically preserves the local mass constraint of tracer, but violates the local volume constraint of the bulk fluid due to numerical approximation of trace-back regions. Many ELLAM schemes are based on similar principles (see, e.g., \cite{Yano1993, Yano1994, Yano1995, Yano1996, Yano1997}).

The volume corrected characteristics-mixed method (VCCMM) was introduced by Arbogast and Huang in 2006 \cite{ArbogastHuang2006}. They gave an efficient algorithm for adjustment of the trace-back points so that volume is conserved locally when the velocity is a potential field. This volume correction procedure leads to a fully conservative characteristic method; that is, a characteristic method that locally preserves the volume of both tracer and ambient fluids. The correction step was also shown by numerical examples to reduce the numerical diffusion significantly. Then the correction step was extended to the case of a solenoidal velocity field \cite{ArbogastHuang2007}, and numerical results showed that very long-time transport in a periodic, divergence-free (vortex-like) flow gave correct results only if the correction step was used, due to its ability to reduce the numerical diffusion. Recently, the authors of this paper showed that the VCCMM has a convergence rate of $O(h/\sqrt{\Delta t} + h + (\Delta t)^r)$ \cite{ArbogastHuang2008}, wherein $r$ is related to the accuracy of solving the characteristics. In this paper, we prove the monotonicity, maximum and minimum principles, and stability properties of the VCCMM as well as discuss its implementation. We also generalize the stability analysis to the CMM.

It is worth noting that Chilakapati in 1999 \cite{Chilakapati1999} presented a fully conservative, characteristic-based transport scheme using an entirely different approach. The flow equation \eqref{eq:flow_equation} is supplemented with the Darcy law

\begin{equation}
\mathbf{u} = -\frac{k}{\mu} \nabla p,
\end{equation}

where $p$ is the pressure, $k$ is the medium permeability, and $\mu$ is the fluid viscosity. Although the flow equation is linear, Chilakapati solves a nonlinear problem for the flow of grid elements along the characteristics, subject to the constraint of volume balance. This is tractable only when one approximates the streamlines by straight lines, which limits the effective time-step when $\mathbf{u}(x, t)$ corresponds to a complex flow. Moreover, one requires a global iteration process to achieve volume conservation (unlike the local iterations required by the VCCMM). The method is not as flexible as VCCMM, which can use any reasonable flow solver, and thus VCCMM is potentially applicable to more problems and higher levels of heterogeneity. Theoretical convergence, stability, monotonicity, and maximum and minimum principles have not been shown for Chilakapati’s method, although it seems likely that it would possess these properties.

Neubauer and Bastian in 2005 \cite{NeubauerBastian2005} presented finite volume and discontinuous Galerkin Eulerian–Lagrangian methods similar to the VCCMM. They use a purely algebraic correction to obtain volume conservation. Moreover, they prove monotonicity of the scheme. The correction step requires that a condition be satisfied \cite[eq. (48)]{NeubauerBastian2005}, but it is neither proven nor clear that the algorithm proposed actually satisfies this condition.

Finally, we note that other authors solve the problem using a fixed Eulerian grid, and use flux correction procedures utilizing characteristic information to obtain full
conservativity. For such methods, it is possible to prove stability, monotonicity, and maximum and minimum principles (see, e.g., [17, 25]).

The rest of this paper is organized as follows. Section 2 gives a review of the VCCMM. Section 3 gives an algebraic description of the VCCMM scheme as well as details of implementation of the algorithm. Section 4 proves the monotonicity, maximum and minimum principles, and stability properties of the VCCMM, and generalizes the analysis to the CMM. A few numerical tests are given in section 5 to verify and compare the stability results of VCCMM and CMM, where we also make an algebraic adjustment for VCCMM to obtain a better result. We give more computational examples of VCCMM in section 6, including one involving long-time nuclear waste storage, to compare the concentration profiles with CMM and Godunov’s method. Finally, a summary and concluding remarks are given in the last section.

2. A review of the VCCMM. Suppose we have a time interval \( J_T := [0, T] \) and a grid \( 0 = t^0 < t^1 < \cdots < t^N = T \). Let \( v := u/\phi \) be the interstitial velocity. In one time step \( J_n := [t^n, t^{n+1}) \), the characteristic trace-back \( \tilde{x}_n(t) = \tilde{x}_n(x, t) \) passing through \((x, t^{n+1})\) will solve the time-backward ordinary differential system

\[
\begin{align*}
\frac{dx_n}{dt} &= v(\tilde{x}_n, t), & t & \in J^n, \\
\tilde{x}_n(t^{n+1}) &= x,
\end{align*}
\]

unless the particle were to trace to the boundary of the domain, which is precluded by our boundary condition (1.3). Let \( \Omega \) be partitioned into elements \( T_h \) of maximal diameter \( h \). As depicted in Figure 2.1, let \( E \in T_h \) be an element of \( \Omega \), and define the space-time trace-back region of \( E \) as

\[ E_n^E := \{ (\tilde{x}, t) \in \Omega \times J^n : \tilde{x} = \tilde{x}_n(x, t), \ x \in E \}, \]

and the trace-back region of \( E \) is

\[ \hat{E}^n := \{ \tilde{x} \in \Omega : \tilde{x} = \tilde{x}_n(x, t^n), \ x \in E \}. \]

Then the local mass constraint derived in [5] is

\[
\int_E \phi c^{n+1} \, dx = \int_{E_n} \phi c^n \, dx + \iint_{E_n^E} q_c \, dx \, dt,
\]

where we use superscript \( n \) to denote a time dependent function evaluated at time \( t^n \).

Due to the approximation of the characteristics (2.1), approximation of \( \hat{E}^n \) by a polygon, and the volume correction adjustment, we actually trace \( E \) to an approximation \( \hat{E}^n \) of \( E^n \), as depicted in Figure 2.2. Therefore, the numerical solution

\[
c_{h+1}^{n+1} \in W_h(\Omega) := \{ w \in L^2(\Omega) : w|_E \text{ is a constant for all } E \in T_h \}
\]

analogous to the local mass constraint (2.3) is defined on \( E \) to be

\[
c_{h,E}^{n+1}|E|_\phi = \int_{E_n} \phi c_{h}^n \, dx + \iint_{E_n^E} q_c^n \, dx \, dt,
\]

where \( |S|_\phi := \int_S \phi(x) \, dx \) is the pore volume of a generic set \( S \subset \Omega \), and \( E_n^E \) is the space-time trace-back region from \( E \) to \( \hat{E}^n \) with a perturbed velocity \( \tilde{v} \approx v \) (see Figure 2.2). The existence of and an error estimate for \( \tilde{v} \) are proved in [5].
Fig. 2.1. The true trace-back region. The mesh element \( E \subset \Omega \) at time \( t^{n+1} \) is tracked back in time under the true flow field \( v \) to the true trace-back region \( \tilde{E}^n \subset \Omega \). The space-time trace-back region \( E^n_T \subset \Omega \times [t^n, t^{n+1}] \) is also traced out.

Fig. 2.2. The approximate trace-back region. The mesh element \( E \subset \Omega \) at time \( t^{n+1} \) is tracked back in time under the approximate flow field \( \tilde{v} \) to the polygon \( \tilde{E}^n \subset \Omega \). The approximate space-time trace-back region \( \tilde{E}^n_T \subset \Omega \times [t^n, t^{n+1}] \) is also traced out.

Fig. 2.3. An injection well \( W \) at time \( t^n \) is traced forward in time to \( \tilde{W}^{n+1} \) at time \( t^{n+1} \). We numerically approximate \( \tilde{W}^{n+1} \) by a polygon, and adjust it so that it satisfies the well volume constraint, forming \( \tilde{W}^{n+1} \approx \hat{W}^{n+1} \). We therefore also define the space-time region \( \tilde{E}^n_W \), which completely covers the cylinder \( I^n_W \) above \( W \).

With \( c = c_I \equiv 1 \) in (2.3), we have the local volume constraint

\[
|E|_\phi = |\hat{E}^n|_\phi + \iint_{\tilde{E}^n_E} q \, dx \, dt,
\]

since the fluid incompressibly fills the pores. In general, without adjustment, the approximation of the characteristics (2.1) and approximation of \( \hat{E}^n \) by a polygon would destroy the equality above. The volume correction adjustment defined in [3] gives the numerical local volume constraint for \( \tilde{E}^n \),

\[
|E|_\phi = |\tilde{E}^n|_\phi + \iint_{\tilde{E}^n_E} q \, dx \, dt.
\]

We assume that \( q|_E = 0 \) for “most” \( E \in \mathcal{T}_h \) and that \( q \) is of one sign on the other elements, called wells, which are isolated from each other. For grid elements near injection wells \( (q > 0) \), the characteristic trace-back \( \hat{E}^n \) is traced back completely or partially into the well, which is hard to approximate in implementation. So, instead, we use a trace-forward technique near injection wells [3, 19]. Let \( W \) be an injection
well, considered at time \( t^n \). As depicted in Figure 2.3, let \( \hat{W}^{n+1} \) be the trace-forward region corresponding to the well \( W \) traced forward in time from \( t^n \) to \( t^{n+1} \). That is, we solve (2.1) forward in time from an initial condition at \( t = t^n \) replacing (2.2), where the initial point is taken from \( W \). Actually, \( \hat{W}^{n+1} \) is adjusted to form the perturbed polygonal region \( \hat{W}^{n+1} \) satisfying the well volume constraint like (2.6) \([3]\),

\[
|\hat{W}^{n+1}|_\phi = |W|_\phi + \int_{\hat{E}^{n+1}_W} q \, dx \, dt,
\]

where \( \hat{E}^{n+1}_W \) is the space-time trace-forward region from \( W \) to \( \hat{W}^{n+1} \). However, the local volume constraint (2.6) needs to be modified to avoid negative coefficients. We give these details in the next section.

3. An algebraic view of the VCCMM and its implementation. It has long been known that, often, linear transport schemes can be written in an algebraic form; see, e.g., \([23, 20]\). In this section, we write the VCCMM (2.4) in vector form and then give details on the data structure and computation of coefficients needed in the implementation of the scheme. Let vector \( c^n_h := (c^n_{h,E})_{E \in T_h} \in \mathbb{R}^{N_h} \) for \( 0 \leq n \leq N \), where \( N_h \) is the number of elements in the mesh \( T_h \). We tacitly assume that (1) the source \( q = 0 \) except in isolated elements of \( T_h \), (2) no sink traces all the way to a source within a single time step, (3) no self-intersected trace-back elements are produced, (4) each element is sufficiently small so that none is traced back into more than one injection well, and (5) if there is a well at element \( W \), then \( q \) on \( W \) and \( u \cdot \nu \) on \( \partial W \) are of one sign. Note that (1) and (5) place a restriction on the data \( q \) appropriate for subsurface transport problems, (2) and (3) make a restriction that each time step \( \Delta t^n := t^{n+1} - t^n \) be not too large, and (4) is a restriction on the mesh.

Each restriction is reasonable in practice.

As illustrated in Figure 2.3, tacit assumptions (1) and (5) allow us to simplify the computation of (2.7) as

\[
|\hat{W}^{n+1}|_\phi = |W|_\phi + \int_{\hat{E}^{n+1}_W} q \, dx \, dt = |W|_\phi + \int_{I^n_W} q \, dx \, dt,
\]

where \( I^n_W \) is the space-time cylinder \( W \times J^n \).

3.1. The scheme in vector form. Let \( T_{h,I} \) and \( T_{h,P} \subset T_h \) be the collection of elements which represent locations of injection and production wells, respectively. Let \( T_{h,I}^* \subset T_h \) be the collection of elements “near,” i.e., influenced by an injection well; that is,

\[
T_{h,I}^* := \{ E \in T_h \setminus (T_{h,I} \cup T_{h,P}) : E \cap \hat{W}^{n+1} \neq \emptyset \text{ for some } W \in T_{h,I} \},
\]

where \( \hat{W}^{n+1} \) is defined in (3.1). Note that \( T_{h,P}, T_{h,I}, \) and \( T_{h,I}^* \) are mutually exclusive. We consider each case separately.

Case 1. \( E \in T_{h,P} \). First, we consider an element \( E \) at a production well, i.e., \( E \in T_{h,P} \). Then \( q^n = 0 \) on \( \hat{E}^n_E \). Since we use a piecewise constant function \( c^n_h \in W_h(\Omega) \) to approximate the solution, the VCCMM (2.4) in an integral form is reduced to

\[
c^{n+1}_{h,E}|E|_\phi = \sum_{F \in T_h} c^n_{h,F}|\hat{E}^n \cap F|_\phi + \sum_{F \in T_h} c^n_{h,F} \int_{\hat{E}^n_E \cap I^n_F} q^- \, dx \, dt
\]

\[
= \sum_{F \in T_h} c^n_{h,F} \left( |\hat{E}^n \cap F|_\phi + \int_{\hat{E}^n_E \cap I^n_F} q^- \, dx \, dt \right), \quad E \in T_{h,P},
\]
where \( I^n_p \) is the space-time cylinder \( F \times J^n \). When \( E \in T_{h,p} \), then \( E \cap \tilde{E}^n \) and \( I^n_p \subset \tilde{E}^n \), since the trace-back of a production well boundary expands. Also, we have \( q = 0 \) in \( \tilde{E}^n \cap I^n_p \) if \( \tilde{E}^n \cap I^n_p \neq \emptyset \) and \( F \neq E \). Then (3.2) is reduced to

\[
(3.3) \quad c^{n+1}_{h,E} |E|_\phi = \sum_{F \neq E} c^n_{h,F} |\tilde{E}^n \cap F|_\phi + c^n_{h,E} \left( |E|_\phi + \int_{I^n_p} q^- \, dx \, dt \right), \quad E \in T_{h,p}.
\]

Notice that the coefficient of \( c^n_{h,E} \) in (3.3) is

\[
(3.4) \quad V^n_E := |E|_\phi + \int_{I^n_p} q^- \, dx \, dt,
\]

which is the remaining volume of the combined fluids in the production well \( E \) at time \( t^n \), so \( V^n_E \) should be nonnegative in physical terms, although \( V^n_E \) could be negative numerically if we have a strong production rate. So when \( V^n_E < 0 \), we modify (3.3) by (1) setting the remaining volume of the combined fluids in \( E \) to be \((V^n_E)^+ = 0\) and (2) reducing a certain volume of fluids from each nearby element by a proportion such that the local volume constraint (2.6) still holds. That is, for each \( E \in T_{h,p} \), we consider the identity

\[
V^n_E = (V^n_E)^+ + \sum_{F \neq E} \frac{\tilde{E}^n \cap F}{|\tilde{E}^n \cap F|_\phi} (V^n_E)^-;
\]

and modify (3.3) to be

\[
(3.5) \quad c^{n+1}_{h,E} |E|_\phi = \sum_{F \neq E} c^n_{h,F} \left( |\tilde{E}^n \cap F|_\phi + \frac{|\tilde{E}^n \cap F|_\phi (V^n_E)^-}{|\tilde{E}^n \cap F|_\phi} \right)
+ c^n_{h,E} (V^n_E)^+, \quad E \in T_{h,p}.
\]

Now the coefficient of each \( c^n_{h,F} \) is nonnegative, which is proved in Lemma 4.1.

**Case 2.** \( E \notin T_{h,p} \cup T_{h,I} \cup T_{h,I^*} \). If \( E \notin T_{h,p} \cup T_{h,I} \cup T_{h,I^*} \), then \( q = 0 \) in \( \tilde{E}^n \), so (2.4) is reduced to

\[
(3.6) \quad c^{n+1}_{h,E} |E|_\phi = \sum_{F \in T_h} c^n_{h,F} |\tilde{E}^n \cap F|_\phi, \quad E \notin T_{h,I} \cup T_{h,I^*} \cup T_{h,p}.
\]

**Case 3.** \( E \in T_{h,I^*} \). Now consider an element \( E \in T_{h,I^*} \) near but not at an injection well \( W \in T_{h,I} \). For the local volume constraint of element \( E \), we adjust \( \tilde{E}^n \) to achieve

\[
(3.7) \quad |E|_\phi = |\tilde{E}^n \setminus W|_\phi + \frac{|E \cap (\tilde{W}^{n+1} \setminus W)|_\phi}{|\tilde{W}^{n+1} \setminus W|_\phi} \int_{I^n_W} q \, dx \, dt,
\]

where the second term on the right-hand side is the volume injected from the well and is transported into element \( E \) as a proportion of the whole. After obtaining the volume adjusted \( \tilde{E}^n \), the concentration in \( E \) is defined by the local mass constraint of tracer

\[
(3.8) \quad c^{n+1}_{h,E} |E|_\phi = \int_{\tilde{E}^n \setminus W} \phi c^n_{h,F} \, dx + \frac{|E \cap (\tilde{W}^{n+1} \setminus W)|_\phi}{|\tilde{W}^{n+1} \setminus W|_\phi} \int_{I^n_W} c_I q \, dx \, dt
= \sum_{F \neq W} c^n_{h,F} |\tilde{E}^n \cap F|_\phi + \frac{|E \cap (\tilde{W}^{n+1} \setminus W)|_\phi}{|\tilde{W}^{n+1} \setminus W|_\phi} \int_{I^n_W} c_I q \, dx \, dt, \quad E \in T_{h,I^*}.
\]
where the second term on the right-hand side is the tracer mass injected from the well and transported into element \( E \) by proportion.

**Case 4.** \( E \in \mathcal{T}_{h,I} \). Finally, we consider an element \( E = W \in \mathcal{T}_{h,I} \), and simply define

\[
(3.9) \quad c_{h,W}^n = c_{I,W}^n := \frac{1}{|W|} \int_W \phi c_I^n \, dx, \quad W \in \mathcal{T}_{h,I},
\]

since the injected concentration is \( c_I \) at well \( W \), and we use the average as the numerical solution.

**Remark 3.1.** In (3.8), actually, we assume that the mass of tracer injected from well \( W \) during time \( J^n \) is uniformly distributed in \( \tilde{W}^{n+1} \), which is true only if \( c_I \) is constant in \( J^n \). So (3.8) is an approximation of the local mass conservation of tracer to \( O(\Delta t^n) \), and we could better approximate the flow from well \( W \) to element \( E \) for general \( c_I = c_I(x,t) \) by microstepping the flow into \( \tilde{W}^{n+1} \).

Combining (3.6), (3.5), (3.8), and (3.9) gives the VCCMM in vector form

\[
(3.10) \quad c_{h+1}^n = A_h^nc_h^n + b_h^n,
\]

where matrix \( A_h^n = (A_{h,E,F}) \in \mathbb{R}^{N_h \times N_h} \) and vector \( b_h^n = (b_{h,E}) \in \mathbb{R}^{N_h} \) are given by

\[
(3.11) \quad A_{h,E,F}^n := \begin{cases} 
\frac{|\tilde{E} \cap F|}{|E|} \left( 1 + \frac{(V^n_E)^-}{|E \setminus F|} \right), & E \in \mathcal{T}_{h,P}, \quad F \neq E, \\
\frac{(V^n_E)^+}{|E|}, & E = F \in \mathcal{T}_{h,P}, \\
\frac{|\tilde{E} \cap F|}{|E|}, & E \notin \mathcal{T}_{h,P} \cup \mathcal{T}_{h,I}, \quad F \notin \mathcal{T}_{h,I}, \\
0, & \text{otherwise},
\end{cases}
\]

wherein \( V^n_E \) is defined by (3.4), and

\[
(3.12) \quad b_{h,E}^n := \begin{cases} 
\bar{c}_{I,E}^n, & E \in \mathcal{T}_{h,I}, \\
\frac{|E \cap (\tilde{W}^{n+1} \setminus W)|}{|E|} \int_{\tilde{W}^n} c_I q \, dx \, dt, & E \in \mathcal{T}_{h,I^*}, \\
0, & \text{otherwise},
\end{cases}
\]

wherein \( \bar{c}_{I,E}^n \) is defined by (3.9), and \( W \in \mathcal{T}_{h,I} \) is the only injection well such that \( E \cap \tilde{W}^{n+1} \neq \emptyset \) when \( E \in \mathcal{T}_{h,I^*} \).

**3.2. Computation of the coefficients in the scheme.** In this section, we consider only a convex polygonal mesh \( \mathcal{T}_h \). To implement (3.10), we need to compute the entries in matrix \( A_h^n \) and vector \( b_h^n \); that is, we need to compute pore volumes of polygons clipped by convex polygonal elements and integrate sources over space-time cylinders at wells. We use a polyline structure to represent polygons, which is a doubly linked list of vertices with the last vertex equal to the first to form a closed polyline. This structure gives us the connectivity of the vertices, which is convenient for computing volumes of polygons.
To compute a trace-back polygon $\tilde{E}^n$ clipped by a rectangular element $F$, we use the Sutherland–Hodgman clipping algorithm [16]. This clipping algorithm works by extending each edge of the clipping window $F$ in turn and selecting only intersection points and vertices from the subject polygon $\tilde{E}^n$ that are on the “visible” side, which is the side where $F$ lies with respect to its extended edge. After $\tilde{E}^n$ is clipped by each side of $F$, the algorithm generates a sequence of vertices of the clipped polygon $\tilde{E}^n \cap F$ given in a clockwise or counterclockwise direction if the vertices of $\tilde{E}^n$ are so given. It is convenient to implement this clipping algorithm with our class of polyline structure, since it allows us to advance through the edges of a polygon in turns, which is consistent with the feature of this clipping algorithm. It also provides the flexibility in applications to handle more general and complicated meshes.

4. Stability and monotonicity analysis. Scheme (3.10) defines the method in an algebraic sense. In practice, all entries in $A^n_h$ and $b^n_h$ are computable, and iteration of (3.10) generates the time sequence of the numerical solution $c^n_{h}$. Therefore, the nature of the method is a linear two-time-level explicit scheme, and the properties of $A^n_h$ and $b^n_h$ completely characterize the method.

4.1. Analysis for the VCCMM. We begin the analysis with two lemmas expressing important characteristics of $A^n_h$ and $b^n_h$.

**Lemma 4.1.** Each entry of the matrix $A^n_h$ defined in (3.11) is nonnegative, i.e.,

$$A^n_{h,E,F} \geq 0 \quad \forall E, F \in \mathcal{T}_h.$$  

**Proof.** By (3.11), we need only show $A^n_{h,E,F} \geq 0$ for $E \in \mathcal{T}_h,P, F \neq E, \tilde{E}^n \cap F \neq \emptyset$. Actually,

$$\frac{|E|_\phi |\tilde{E}^n \setminus E|_\phi}{|\tilde{E}^n \cap F|_\phi} A^n_{h,E,F} = |\tilde{E}^n \setminus E|_\phi + \int_{\mathcal{I}_E^R} q^- dx dt \geq |\tilde{E}^n \setminus E|_\phi + \int_{\tilde{E}^n \setminus E} q dx dt = 0,$$

where we obtain the last equality by the local volume constraint (2.6). \(\square\)

**Lemma 4.2.** Let $b^n_{h,1} := (b^n_{h,1,E}) \in \mathbb{R}^{N_h}$ be vector $b^n_h$ with $c_I \equiv 1$, and let $e_h \in \mathbb{R}^{N_h}$ be with each component $e_{h,E} = 1$. Then

$$A^n_{h} e_h + b^n_{h,1} = e_h.$$  

**Proof.** By (3.11) and (3.12), we compute each component as

$$(A^n_{h} e_h + b^n_{h,1})_E = \sum_{F \in \mathcal{T}_h} A^n_{h,E,F} + b^n_{h,1,E} \quad \forall E \in \mathcal{T}_h.$$  

We treat each element $E \in \mathcal{T}_h$ according to its type.

When $E \in \mathcal{T}_h,I$,

$$(A^n_{h} e_h + b^n_{h,1})_E = \sum_{F \in \mathcal{T}_h} 0 + 1 = 1.$$
When $E \in \mathcal{T}_{h, I^*}$, there is some $W \in \mathcal{T}_{h, I}$ such that $E \cap W^{n+1} \neq \emptyset$, and so

$$(A^n_h e_h + b^n_{h,1})_E = \frac{|\hat{E}^n \setminus W|_\phi}{|E|_\phi} + \frac{|E \cap (\hat{W}^{n+1} \setminus W)|_\phi}{|E|_\phi} \int_{I^n_E} q \, dx \, dt = 1,$$

where the last equality is obtained by (3.7).

When $E \in \mathcal{T}_{h, P}$,

$$(A^n_h e_h + b^n_{h,1})_E = \frac{|\hat{E}^n \setminus E|_\phi}{|E|_\phi} \left(1 + \frac{(V^n_n)^-}{|E_n|_\phi}\right) + \frac{(V^n_n)^+}{|E|_\phi} = \frac{1}{|E|_\phi} \left(|\hat{E}^n|_\phi + \int_{I^n_E} q^- \, dx \, dt\right) = \frac{1}{|E|_\phi} \left(|\hat{E}^n|_\phi + \int_{\hat{E}^n_E} q \, dx \, dt\right) = 1,$$

where the last equality is obtained by (2.6).

When $E \notin \mathcal{T}_{h, I} \cup \mathcal{T}_{h, I^*} \cup \mathcal{T}_{h, P}$,

$$(A^n_h e_h + b^n_{h,1})_E = \sum_{F \in \mathcal{T}_h} \frac{|\hat{E}^n \cap F|_\phi}{|E|_\phi} = \frac{|\hat{E}^n|_\phi}{|E|_\phi} = 1,$$

where the last equality is obtained by (2.6). This completes the proof.

Lemma 4.1 directly implies the monotonicity of the method.

**Theorem 4.3 (monotonicity of VCCMM).** The VCCMM given by (3.10) is monotone. That is, if the initial approximation $c^0_h \leq \tilde{c}^0_h$, then $c^n_h \leq \tilde{c}^n_h$ for all $n \geq 0$.

**Proof.** Let $c^n_h$ and $\tilde{c}^n_h \in \mathbb{R}^{N_h}$ be the vectors for numerical solutions $c^n_h$ and $\tilde{c}^n_h$, respectively. Applying scheme (3.10) to $c^n_h$ and $\tilde{c}^n_h$ and taking differences gives

$$c^{n+1}_h - c^n_h = A^n_h (\tilde{c}^n_h - c^n_h).$$

So, by Lemma 4.1, $c^n_h \leq \tilde{c}^n_h$ implies $c^{n+1}_h \leq \tilde{c}^{n+1}_h$. By induction, the proof is complete.

Lemmas 4.1 and 4.2 imply maximum and minimum principles for the method, which we state in the following theorem. This result can also be shown from theories on methods that have the general algebraic form (3.10) [23, 20].

**Theorem 4.4 (maximum and minimum principles of VCCMM).** If there exist constants $c_*$ and $c^*$ such that $c_* \leq c^0_h \leq c^*$ and $c_* \leq c_I \leq c^*$, then $c_* \leq c^n_h \leq c^*$ for all $n \geq 0$.

**Proof.** Let $e^n_h$ be the vector for numerical solution $e^n_h$. Applying scheme (3.10), by Lemmas 4.1 and 4.2, we have for each $E \in \mathcal{T}_h$

$$c^{n+1}_h, E = \sum_{F \in \mathcal{T}_h} A^n_{h,E,F} c^n_{h,F} + b^n_{h,E}$$

$$\leq \left(\sum_{F \in \mathcal{T}_h} A^n_{h,E,F}\right) \max c^n_h + b^n_{h,1,E} c^*$$

$$\leq \left(\sum_{F \in \mathcal{T}_h} A^n_{h,E,F} + b^n_{h,1,E}\right) \max\{\max c^n_h, c^*\} = \max\{\max c^n_h, c^*\},$$
which means \( c_h^n \leq c^* \) implies \( c_h^{n+1} \leq c^* \), so by induction the maximum principle holds. Similarly, we can show the minimum principle.

**Remark 4.1.** As shown in the proof of Lemma 4.2, the volume correction adjustment in the method is necessary and essential to the maximum and minimum principles of VCCMM, which implies that the method produces neither overshoots nor undershoots.

We recall that for any matrix \( A \in \mathbb{R}^{N_h \times N_h} \), the matrix norm

\[
\|A\|_\infty = \max_{E \in T_h} \sum_{F \in T_h} |A_{E,F}|
\]

is induced from the \( l^\infty \)-norm of vectors, where \( |x|_\infty := \max_{E \in T_h} |x|_E \) for any \( x = (x_E) \in \mathbb{R}^{N_h} \). Since \( b_{n,1}^n \) has nonnegative components by (3.12), combining Lemmas 4.1 and 4.2 gives that \( \|A_h^n\|_\infty \leq 1 \). Since the boundary conditions require \( \mathcal{T}_{h,P} \neq \emptyset \), then \( b_{n,1,E}^n = 0 \) for some \( E \), so we have the following corollary.

**Corollary 4.5.** The matrix \( A_h^n \) defined in (3.11) has the induced norm

\[
\|A_h^n\|_\infty = 1.
\]

**Theorem 4.6 (stability of VCCMM).** The VCCMM given by (3.10) is stable. That is, if \( c_h^n (0 \leq n \leq N) \) satisfies scheme (3.10) in time \( J_T \) with an initial approximation \( c_h^0 \), and if \( \tilde{c}_h^n (0 \leq n \leq N) \) satisfies the perturbed scheme

\[
\tilde{c}_h^{n+1} = A_h^n \tilde{c}_h^n + b_h^n + \Delta t^n \delta_h^n
\]

in time \( J_T \) with an initial approximation \( \tilde{c}_h^0 \), where \( \Delta t^n := t^{n+1} - t^n \), and if \( \delta_h^n \in \mathbb{R}^{N_h} \) is a perturbation at time step \( J^n \), then the following error estimate holds:

\[
\max_{0 \leq n \leq N} |\tilde{c}_h^n - c_h^n|_\infty \leq |\tilde{c}_h^0 - c_h^0|_\infty + T \max_{0 \leq n \leq N} |\delta_h^n|_\infty.
\]

**Proof.** Subtracting (3.10) from (4.3), we have

\[
\tilde{c}_h^{n+1} - c_h^{n+1} = A_h^n (\tilde{c}_h^n - c_h^n) + \Delta t^n \delta_h^n.
\]

Taking the \( l^\infty \)-norm on both sides of (4.5) and using Corollary 4.5 gives

\[
|\tilde{c}_h^{n+1} - c_h^{n+1}|_\infty \leq |A_h^n (\tilde{c}_h^n - c_h^n)|_\infty + \Delta t^n |\delta_h^n|_\infty \leq |\tilde{c}_h^n - c_h^n|_\infty + \Delta t^n \max_{0 \leq n \leq N} |\delta_h^n|_\infty.
\]

Iterating (4.6) for \( n \) gives

\[
|\tilde{c}_h^n - c_h^n|_\infty \leq |\tilde{c}_h^0 - c_h^0|_\infty + t^n \max_{0 \leq n \leq N} |\delta_h^n|_\infty,
\]

and we obtain (4.4). \( \square \)

**4.2. Analysis of the CMM.** Now we consider the method without the volume correction, which means we trace back only finitely many points on the boundary of each grid element \( E \) to form a polygonal approximation \( \tilde{E}_h^n \) without trace-back point adjustment. By the same argument, we still have the CMM in the vector form (3.10)–(3.12) as

\[
c_h^{n+1} = \hat{A}_h^n c_h^n + \hat{b}_h^n,
\]
where matrix $\tilde{A}_h^n$ and vector $\tilde{b}_h^n$ take the previous form but for unadjusted trace regions. Due to the violation of the local volume constraint, we lose the properties of $\tilde{A}_h^n$ and $\tilde{b}_h^n$ in Lemmas 4.1 and 4.2 and Corollary 4.5. Therefore, we no longer have the monotonicity and the maximum and minimum principles for the CMM. This has been observed in numerical examples (see [3] and sect. 5.1 and 6.1). Instead, we have an estimate of the norm of $\tilde{A}_h^n$ as shown in Lemma 4.7 below. For the purpose of the proof, besides the assumptions made in section 4, we further impose the following two assumptions on time and space discretizations.

**Assumption 4.1** (regularity of the time discretization). The time grid $0 = t^0 < t^1 < \cdots < t^N = T$ of $J_T$ is regular, i.e., there exists a constant $\lambda_1 > 0$ such that

$$\Delta t \leq \lambda_1 \inf_n \Delta t^n,$$

where $\Delta t := \sup_n \Delta t^n$.

**Assumption 4.2** (regularity of the space discretization). The mesh $\mathcal{T}_h$ of bounded domain $\Omega \subset \mathbb{R}^d$ is regular, i.e., there exists a constant $\lambda_2 > 0$ such that the pore volume of each element $E \in \mathcal{T}_h$ satisfies

$$|E|_\phi \geq \lambda_2 h^d.$$

**Lemma 4.7.** Let Assumption 4.2 hold. Then the matrix $\tilde{A}_h^n$ for CMM satisfies

$$(4.8) \quad \|\tilde{A}_h^n\|_\infty \leq 1 + C \Delta t^n,$$

where $C > 0$ is a constant independent of $n$, $h$, and $\Delta t^n$.

**Proof.** By (4.2), we need only estimate each absolute row sum of $\tilde{A}_h^n$. For $E \notin \mathcal{T}_h$, the exact trace-back element has pore volume $|\tilde{E}^n|_\phi \leq |E|_\phi$. So, by (3.11), we have

$$\sum_{F \in \mathcal{T}_h} |\tilde{A}_{h,F}^n| \leq \frac{|\tilde{E}^n|_\phi}{|E|_\phi} = \frac{|\tilde{E}^n|_\phi}{|E|_\phi} + \frac{|\tilde{E}^n|_\phi - |\tilde{E}^n|_\phi}{|E|_\phi} \leq 1 + C \Delta t^n,$$

where the last inequality is obtained by Assumption 4.2 and the estimate of the discrepancy pore volume

$$(4.9) \quad \|\tilde{E}^n|_\phi - |\tilde{E}^n|_\phi| \leq C h^d \Delta t^n,$$

where $C > 0$ is a constant independent of $n$, $h$, and $\Delta t^n$. The estimate (4.9) for $d = 2$ is proved in [5], and the general result in $\mathbb{R}^d$ can be derived by a similar technique.

For $E \in \mathcal{T}_h$, we have

$$\sum_{F \in \mathcal{T}_h} |\tilde{A}_{h,F}^n| = \frac{|\tilde{E}^n \setminus E|_\phi}{|E|_\phi} + \frac{(V_E^*)^-}{|E|_\phi} + \frac{(V_E^*)^+}{|E|_\phi} \leq \frac{1}{|E|_\phi} |\tilde{E}^n \setminus E|_\phi + \frac{|V_E^*|}{|E|_\phi} + \int_{I_E} q^+ dx \, dt \leq 1 + \frac{|\tilde{E}^n|_\phi + |E|_\phi - |\tilde{E}^n|_\phi|}{|E|_\phi} \leq 1 + C \Delta t^n,$$

where we use (2.5) and, again, the last inequality is given by (4.9). \qed
THEOREM 4.8 (stability of CMM). Let Assumptions 4.1 and 4.2 hold. The CMM given by (4.7) is stable. That is, if \( c^0_h \ (0 \leq n \leq N) \) satisfies (4.7) in time \( J_T \) with an initial approximation \( c^0_h \), and \( c^0_h \ (0 \leq n \leq N) \) satisfies the perturbed scheme

\[
(4.10) \quad \bar{c}^{n+1}_h = A^n_h \bar{c}_h^n + b^n_h + \Delta t^n \delta^n_h
\]

in time \( J_T \) with an initial approximation \( \bar{c}^0_h \), where \( \delta^n_h \in \mathbb{R}^{N_h} \) is a perturbation at time step \( J^n \), then the following error estimate holds:

\[
(4.11) \quad \max_{0 \leq n \leq N} |\bar{c}_h^n - c_h^n|_{\infty} \leq e^{C T \lambda_1} |\bar{c}^0_h - c^0_h|_{\infty} + C^{-1}(e^{C T \lambda_1} - 1) \max_{0 \leq n \leq N} |\delta^n_h|_{\infty},
\]

where \( \lambda_1 > 0 \) is the constant in Assumption 4.1, and \( C > 0 \) is a constant independent of \( h \) and \( \Delta t \).

Proof. Subtracting (3.10) from (4.10), we have

\[
\bar{c}^{n+1}_h - c_h^{n+1} = A^n_h (\bar{c}^n_h - c^n_h) + \Delta t^n \delta^n_h.
\]

Taking the \( t \)-/\( \infty \)-norm on both sides and using Lemma 4.7 give

\[
|\bar{c}^{n+1}_h - c_h^{n+1}|_{\infty} \leq |A^n_h (\bar{c}^n_h - c^n_h)|_{\infty} + \Delta t^n |\delta^n_h|_{\infty} \\
\leq (1 + C \Delta t) |\bar{c}^n_h - c^n_h|_{\infty} + \Delta t \max_{0 \leq n \leq N} |\delta^n_h|_{\infty}.
\]

Iterating above to solve for sequence \( \bar{c}^n_h - c^n_h \) gives

\[
|\bar{c}^n_h - c^n_h|_{\infty} \leq (1 + C \Delta t)^n |\bar{c}^0_h - c^0_h|_{\infty} + C^{-1}((1 + C \Delta t)^n - 1) \max_{0 \leq n \leq N} |\delta^n_h|_{\infty}.
\]

Note that, by Assumption 4.1,

\[
(1 + C \Delta t)^n \leq (1 + C \Delta t)^N \leq e^{CN \Delta t} \leq e^{C T \lambda_1},
\]

so we obtain (4.11). \( \square \)

Remark 4.2. Without the volume correction adjustment, CMM is still stable, but with a much larger error estimate for the numerical solution than VCCMM.

5. Stability tests. Without the volume correction adjustment, from Theorem 4.8, the error estimate of CMM in (4.11) has a much larger constant than VCCMM, although it maintains the stability property. We show here through numerical experiments that, in fact, the CMM is relatively less stable to perturbations, and is more likely to produce larger errors than VCCMM.

To verify the stability properties in Theorems 4.6 and 4.8, we consider a quarter of a “five-spot” pattern of wells. We take a rectangular domain \( \Omega = (0, 150) \times (0, 200) \) meters with a tracer injection well near the corner \((0, 0)\), a production well near the corner \((150, 200)\), and the no-flow boundary condition (1.3). We impose a uniform 50 \times 50 rectangular grid over \( \Omega \) and a uniform time step \( \Delta t = 50 \) minutes. It is initially clean: \( c^0(x) = 0 \). The injector covers one cell near the corner \((0, 0)\) and has a constant rate of \( q = 10 \) m\(^2\)/hour, injecting an inert tracer with concentration \( c_I = 1 \). The cell comprising the producer near the opposite corner \((150, 200)\) has the rate opposite that of the injector. For simplicity, we solve (1.1)–(1.4) with a constant porosity \( \phi(x) \equiv 1 \) and an isotropic but heterogeneous permeability tensor \( k(x) \) depicted in Figure 5.1, which is geostatically generated with mean \( m_k = 10 \) md and dimensionless coefficient of variation \( C_v = 2 \), where

\[
m_k = \frac{1}{|\Omega|} \int_{\Omega} k(x) \, dx, \quad C_v = \frac{1}{m_k} \left( \frac{1}{|\Omega|} \int_{\Omega} (k(x) - m_k)^2 \, dx \right)^{1/2}.
\]
5.1. Concentration profiles with perturbations. Figure 5.2 shows the concentration profiles of CMM and VCCMM at time $t = 10^4$ minutes (about 6.94 days) and the results with perturbations. Without the volume correction adjustment, CMM (Figure 5.2, upper left) produces relative errors of trace-back element volumes as large as 22% near wells, and a maximum overshoot of concentration about 40%, so it shows many nonphysical local extrema. With the volume correction adjustment, the VCCMM (Figure 5.2, upper right) corrects relative errors of trace-back element volumes to the $10^{-5}$–$10^{-6}$ range under a certain error tolerance, and shows monotone contours of concentrations in the domain without overshoots or undershoots.

With each component given a random initial perturbation and a random perturbation in each time step $\delta_n^h$ uniformly distributed in $(-10^{-3}, 10^{-3})$, the perturbed concentration profiles of CMM and VCCMM show contours similar to their results without perturbations, respectively (see Figure 5.2). Therefore, both methods are qualitatively stable to perturbations in the schemes given by (4.3) and (4.10).

5.2. The long time behavior and the algebraic adjustment. To test the long time behavior of the stability of VCCMM and CMM, we use a much smaller
injection rate $q = 0.1 \text{ m}^2/\text{hour}$, so that the errors in concentration due to the perturbations will propagate along the flow streamlines for a much longer time before being extracted by the production well. Also, we use only positive perturbations so that perturbation errors do not cancel each other but rather accumulate as time proceeds. In addition, to emphasize the necessity of the volume correction, we use a more heterogeneous permeability field with $m_k = 10 \text{ md}$ and $C_v = 4$.

Figure 5.3 plots the maximum concentration errors of CMM and VCCMM up to time $t = 10^6$ minutes (about 694.44 days) with random perturbations uniformly distributed in $(0, 10^{-7})$. As time proceeds, the maximum concentration errors of both methods accumulate. The curve of CMM seems more like the graph of an exponential function as expected from Theorem 4.8. The curve of VCCMM shows much smaller errors than CMM. The error of VCCMM at time $t = 10^6$ minutes is only about 54% of CMM. However, due to the error of the volume correction adjustment, the computed matrix $A^n_h$ does not strictly satisfy Lemma 4.2, so the curve does not seem to be perfectly linear. This issue occurs because the volume correction procedure can be accomplished in a discrete fashion only within finitely many steps, and both an error tolerance and maximum number of steps for the adjustment are imposed for a certain accuracy and efficiency. Therefore, to guarantee that the computed matrix $A^n_h$ satisfies Lemma 4.2, we perform the following algebraic adjustment. That is, we set

$$\tilde{A}^n_h,E,F := \frac{A^n_h,E,F}{\sum_{G \in T_h} A^n_h,E,G}$$

if the row sum $\sum_{G \in T_h} A^n_h,E,G > 1$. Then the adjusted matrix $\tilde{A}^n_h := (\tilde{A}^n_h,E,F)$ will satisfy Lemma 4.2. As a result, Figure 5.3 shows that the curve of VCCMM with algebraic adjustment (VCCMM-AA) seems to be linear, and has smaller errors than VCCMM. The error of VCCMM-AA at time $t = 10^6$ minutes is only about 20% of CMM.

Table 5.1 shows the maximum concentration errors of CMM, VCCMM, and VCCMM-AA at time $t = 10^6$ minutes with respect to random perturbations uniformly distributed in $(0, \varepsilon)$ for different $\varepsilon$. The results show that, for the fixed simulation time, the maximum concentration errors of all three methods are approximately linearly dependent on the maximum perturbation size $\varepsilon$, as expected from Theorems 4.6 and 4.8. Again, due to the volume correction adjustment, VCCMM and VCCMM-AA
Table 5.1
Maximum errors of concentration with respect to perturbations in different sizes.

<table>
<thead>
<tr>
<th>ε</th>
<th>CMM</th>
<th>VCCMM</th>
<th>VCCMM-AA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1e-03</td>
<td>2.49133 e+03</td>
<td>1.32531 e+03</td>
<td>5.02741 e+02</td>
</tr>
<tr>
<td>1e-04</td>
<td>2.49061 e+02</td>
<td>1.32627 e+02</td>
<td>5.02678 e+01</td>
</tr>
<tr>
<td>1e-05</td>
<td>2.48911 e+01</td>
<td>1.32496 e+01</td>
<td>5.01977 e+00</td>
</tr>
<tr>
<td>1e-06</td>
<td>2.48786 e+00</td>
<td>1.32430 e+00</td>
<td>5.01934 e+00</td>
</tr>
<tr>
<td>1e-07</td>
<td>2.48758 e+00</td>
<td>1.32569 e+00</td>
<td>5.01863 e+00</td>
</tr>
<tr>
<td>1e-08</td>
<td>2.48758 e+00</td>
<td>1.32569 e+00</td>
<td>5.01863 e+00</td>
</tr>
<tr>
<td>1e-09</td>
<td>2.48938 e-03</td>
<td>1.32314 e-03</td>
<td>5.02050 e-03</td>
</tr>
<tr>
<td>1e-10</td>
<td>2.48841 e-04</td>
<td>1.32414 e-04</td>
<td>5.00915 e-04</td>
</tr>
<tr>
<td>1e-11</td>
<td>2.49221 e-05</td>
<td>1.32613 e-05</td>
<td>5.03443 e-06</td>
</tr>
<tr>
<td>1e-12</td>
<td>2.49138 e-06</td>
<td>1.32135 e-06</td>
<td>5.02401 e-07</td>
</tr>
</tbody>
</table>

show much smaller errors than CMM.

6. Computational examples. In this section, we first demonstrate a numerical implementation of the VCCMM on a quarter of a “five-spot” problem to compare the concentration profile with the CMM and Godunov’s method [21, 9, 15], where we also discuss two different flow approximations. Then we give an application to a nuclear waste disposal problem to show that the VCCMM produces less numerical diffusion, shows similar shapes of concentration contours, and is more efficient than the Godunov method.

6.1. A quarter “five-spot” problem. We consider a quarter “five-spot” problem with a rectangular domain (0, 15) × (0, 20) meters, a rate \( q = 1.2 \text{ m}^2/\text{min} \), and a heterogeneous permeability \( k(x) \) depicted in Figure 6.1. The permeability has \( m_k = 10 \text{ milidarcies (md)} \) and \( C_v = 2.5 \), which varies by about four orders of magnitude (10\(^{-2}\) to 10\(^{2}\) md).

First, we approximate the flow equation (1.1), supplemented with Darcy’s law

(6.1)

\[
\mathbf{u} = -\frac{k}{\mu} \nabla p,
\]

by the mixed finite element method for \( p \) and \( \mathbf{u} \), where \( p \) is the fluid pressure, and \( \mu \) is the fluid viscosity. Figure 6.2 shows the divergences of velocity using \( \text{RT}_0 \) [24] and \( \text{AW}_0 \) [7] mixed finite element spaces on a uniform 50 × 50 grid. The reason for this choice is that both methods are locally conservative. One advantage of \( \text{RT}_0 \) is that the simple basis functions allow analytic tracing of streamlines [18], although the velocities themselves have discontinuities on element boundaries in the tangential direction. The advantage of \( \text{AW}_0 \) is that its velocities are continuous, so a simple numerical solver can be used to obtain the characteristic traces.

At wells, both \( \text{RT}_0 \) and \( \text{AW}_0 \) compute \( \nabla \cdot \mathbf{u} = 0.1667 \text{ sec}^{-1} \) for the injector and \( \nabla \cdot \mathbf{u} = -0.1667 \text{ sec}^{-1} \) for the producer. This is consistent, since the well rate per unit pore volume is

\[
\frac{\text{well rate}}{\text{well volume}} = \frac{1.2 \text{ m}^2/\text{min}}{0.12 \text{ m}^2} = 10 \text{ min}^{-1} \approx 0.1667 \text{ sec}^{-1}.
\]

Due to the local heterogeneity of the permeability, the divergence on the rest of the field varies by about four orders of magnitude (10\(^{-16}\) to 10\(^{-13}\) sec\(^{-1}\)). The \( \text{RT}_0 \) approximation satisfies the conservation of bulk fluid (1.1) pointwise, but the \( \text{AW}_0 \) approximation satisfies only (1.1) on the average in each grid element. Thus, the
Fig. 6.1. A heterogeneous permeability field in millidarcies (md).

Fig. 6.2. Divergence of velocity in sec$^{-1}$ (grid: 50 × 50).

Divergence for RT$_0$ is essentially accurate to rounding errors and solver tolerances, and it shows a better accuracy than that for AW$_0$. However, these errors are extremely small, so that they do not affect the quality of the transport approximations. Actually, AW$_0$ shows a better concentration profile as shown in Figures 6.3 and 6.4.

Figure 6.3 shows the tracer concentration profiles at time $t = 100$ minutes for six cases. The reference solution is the first order Godunov method (FOG) with the RT$_0$ flow approximation, shown in the top-right corner. It uses the CFL restricted $\Delta t_{\text{CFL}} = 6$ seconds. The other two cases shown on the top row use the volume uncorrected CMM, also with the RT$_0$ flow approximation, but with $\Delta t = 80$ seconds ($\Delta t \approx 13.3 \Delta t_{\text{CFL}}$). The top-left solution has considerable overshoot (up to about 30%), and introduces many nonphysical local minima and maxima into the solution. When the algebraic adjustment (5.1) is applied naively to CMM (shown top-middle), no overshoots develop. However, the solution is no longer mass conservative, indicating that more care must be used in removing the volume error.

The volume corrected method uses more care in removing the volume error. Three cases of VCCMM are shown in the bottom row of Figure 6.3. In these results, we used the algebraic adjustment to remove the numerical rounding error that affects the volume error (see subsection 5.2), so the volume error identically vanishes. In all three cases, VCCMM gives physically reasonable solutions that compare well with FOG (top-right). Moreover, VCCMM exhibits no overshoots nor undershoots and has a monotone contour. The bottom-left case uses $\Delta t = 80$ seconds and the RT$_0$ flow approximation. The bottom-middle case again uses $\Delta t = 80$ seconds but with the AW$_0$ flow approximation. Finally, the bottom-right case uses the CFL restricted $\Delta t_{\text{CFL}} = 6$ seconds and the RT$_0$ flow approximation, which is strictly comparable.
Fig. 6.3. Tracer concentration at time $t = 100$ min (grid: $50 \times 50$). The right-most cases use $\Delta t = \Delta t_{\text{CFL}} = 6$ seconds, while the left and middle cases use $\Delta t = 80$ seconds.

Fig. 6.4. Tracer concentration at time $t = 100$ min (grid: $100 \times 100$). The right-most cases use $\Delta t = \Delta t_{\text{CFL}} = 1.5$ seconds, while the left and middle cases use $\Delta t = 30$ seconds.
to the FOG case. The FOG and long-time VCCMM results show similar levels of numerical diffusion. Of the three VCCMM cases, we note that VCCMM-AW shows the least numerical diffusion, and VCCMM-RT0-ΔtCFL shows the most, due to taking more time steps.

Figure 6.4 shows completely analogous results using a refined 100 × 100 grid. In this example, the restricted CFL time step is ΔtCFL = 1.5 seconds, used for FOG and VCCMM-RT0-ΔtCFL (right-most cases), and Δt = 30 seconds for CMM and the other VCCMM cases (left and middle). Note that Δt = 20ΔtCFL. Completely similar conclusions can be seen on the refined grid.

Remark 6.1. Instead of using piecewise constants to approximate solutions, one could use piecewise linears or multilinears, postprocessed from the constant values and appropriately slope limited. Such a spatial postprocessing procedure, appropriate for VCCMM, was defined in [6], and was proven to improve the spatial accuracy to second order for smooth solutions [5]. In that case, we should compare it with the higher order Godunov method (HOG) [9, 15]. Our comparison here of VCCMM to FOG is appropriate, as neither is postprocessed.

6.2. A nuclear waste disposal problem. We now consider a problem defined originally by ANDRA [10] in the early 2000’s for safety assessment in nuclear waste management. It leads to a classical advection-diffusion-reaction problem. Since we are demonstrating VCCMM, we make some modifications to the problem to better match the limitations of our demonstration code.

6.2.1. The problem and governing equations. The problem is restricted to a two-dimensional rectangular disposal site Ω = (0, 25000) × (0, 695) in meters. We use a constant layer hydraulic conductivity kh [8, p. 133] in meter/year, as depicted in Figure 6.5. The layers are dogger (kh = 25.2288 m/year), clay (kh = 3.1536 × 10⁻⁶ m/year), limestone (kh = 6.3072 m/year), and marl (kh = 3.1536 × 10⁻⁵ m/year). Note that kh varies by about seven orders of magnitude. A deep geological repository, denoted by R, is modeled by a rectangular region in the clay layer with dimensions R = (18440, 21680) × (244, 250) meters. The computation is carried over time JT with T = 10⁶ years.

It is assumed that all rock layers are saturated with water and that boundary conditions are stationary, so that the flow and pressure are independent of time. Darcy’s law gives the velocity

\[ \mathbf{u} = -k_h \nabla \Psi \]

in terms of the hydrodynamic load \( \Psi := p/(\rho g) + y \), where we assume the density \( \rho \)

![Fig. 6.5. The computational domain of the disposal site showing four layers and the repository.](image_url)
is a constant and $u$ satisfies the mass conservation (1.1) with $q = 0$. Also, we impose the following boundary conditions (in meters):

$$
\begin{align*}
\Psi &= 286 \quad \text{on } \{0\} \times (0,200), \\
\Psi &= 200 \quad \text{on } \{0\} \times (295,595), \\
\Psi &= 289 \quad \text{on } \{25000\} \times (0,200), \\
\Psi &= 310 \quad \text{on } \{25000\} \times (350,595), \\
\nabla \Psi \cdot \nu &= 0 \quad \text{elsewhere}.
\end{align*}
$$

At the initial time, the repository has a leak, and we consider the long-lived radioactive element iodine-129 that escapes from the repository cave into the water. The leak maintains a repository concentration $c^0 = 0.133 \text{ mol/m}^3$. The concentration $c$ is given by the advection-diffusion-reaction equation

$$(6.3) \quad \phi (c_t + \lambda c) + \nabla \cdot (cu - D \nabla c) = 0 \quad \Omega \times J_T,$$

where the effective porosity $\phi = 0.001$ in the clay layer and 0.1 elsewhere, the radioactive decay constant $\lambda = \log(2)/T_{\text{half}}$ with the half life time $T_{\text{half}} = 1.57 \times 10^7 \text{ years}$, and the effective diffusion/dispersion tensor $D$ depends on the Darcy velocity $u$ as

$$(6.4) \quad D(u) = \phi d_{\text{mol}} I + |u| [d_{\text{long}} E(u) + d_{\text{trans}} (I - E(u))],$$

where $E(u) = uu^T/|u|^2$ and molecular diffusion, longitudinal and transverse dispersion coefficients assumed constant in each layer, are given in Table 6.1. Finally, we impose the boundary conditions for transport as

$$
\begin{align*}
\nabla c \cdot \nu &= 0 \quad \text{on } \{0\} \times \{(0,200) \cup (295,595)\}, \\
c &= 0 \quad \text{elsewhere}.
\end{align*}
$$

### Table 6.1

**Diffusion/dispersion coefficients in the four layers.**

<table>
<thead>
<tr>
<th>Layer</th>
<th>$d_{\text{mol}}$ (m$^2$/year)</th>
<th>$d_{\text{long}}$ (m)</th>
<th>$d_{\text{trans}}$ (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dogger</td>
<td>$5 \times 10^{-4}$</td>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td>Clay</td>
<td>$9.48 \times 10^{-7}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Limestone</td>
<td>$5 \times 10^{-4}$</td>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td>Marl</td>
<td>$5 \times 10^{-4}$</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

#### 6.2.2. The numerical results.

We use a nonuniform $108 \times 70$ rectangular mesh with local refinement near the repository as shown in Figure 6.6. We use an operator splitting technique to solve (6.3) by approximating the advection with FOG or VCCMM, approximating the diffusion with the expanded mixed finite element method [12], and solving the reaction analytically by solving an ordinary differential equation.

Figure 6.7 shows the hydrodynamic load (top) and speed (bottom) with the RT$0$ mixed finite element approximation. There is approximately a linear hydrodynamic load drop in the limestone layer since the conductivity is constant. The clay layer shows a nearly constant hydrodynamic load since the hydrodynamic load drop in this layer is small. For the speed, as we should expect, there is a relatively large speed in the limestone layer, since not only is this layer more permeable, but also there is a
larger hydrodynamic load drop on the boundaries. Clay and marl layers have little speed due to low conductivities and no-flow boundary conditions.

We take $\Delta t = 100$ years for Godunov's method because the CFL restricted
Δt_{CFL} ≈ 102.52 years. Due to the simple structure of the conductivity distribution in the domain, we are able to use a large time step Δt = 2500 years for VC-CMM, and even then, there is little work needed for the trace-back adjustment. Figure 6.8 shows the characteristic trace-back mesh, where we treat the inflow boundary \{25000\} \times (0, 695) as an injection well using the trace-forward technique. Due to the large time step, there are some self-intersected trace-back polygons created near the “sharp corner” interface between the clay and limestone layers. However, this degeneracy results only in a locally minor inaccuracy of the transport approximation as shown in Figure 6.11 (bottom), so we chose to use this large time step to reduce the computational cost. For better accuracy, we could locally have refined the mesh or traced back more points near the “sharp corner.”

Up to time 3 \times 10^4 years (Figure 6.9), almost all of the iodine-129 is still in the clay layer, which has a low conductivity and a no-flow boundary condition, so there
is little advection and the diffusion effect is dominant. Figure 6.10 shows the concentration profiles at $2.5 \times 10^5$ years. Due to the relatively high speed in the dogger and limestone layers, the flow front is moving much faster after it escapes from the clay layer. In addition, due to the restricted time step, Godunov’s method has more numerical diffusion and shows a much wider distribution of concentration.

Figure 6.11 shows the concentration profiles at $3 \times 10^5$ years. The flow front is moving much faster in the limestone layer than in the dogger layer due to a much higher conductivity. Each profile shows a sharp concentration jump cross the interface between the limestone and clay layers since the speeds in the two layers differ greatly. There is some inaccuracy of VCCMM near the “sharp corner” of the interface due to the creation of self-intersected trace-back polygons. As noted above, this inaccuracy could be removed by a local refinement or by tracing back more points. But it has...
very little effect on the overall flow.

Figure 6.12 shows the results at $10^6$ years, where Godunov’s method and the VCCMM predict a similar shape of the plume.

6.2.3. **Comparison of the advection computational time.** A summary of the actual time of computation on the same machine is shown in Table 6.2. Indeed, due to the simple structure of the conductivity distribution, little time is spent on characteristic tracing and volume adjustment, and VCCMM takes only about 59.56% of the advection time of Godunov’s method simulated by Parssim1 [1].

<table>
<thead>
<tr>
<th>Flow</th>
<th>Tracing and adjustment</th>
<th>Transport (advection)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Godunov</td>
<td>0.05</td>
<td>N/A</td>
</tr>
<tr>
<td>VCCMM</td>
<td>0.05</td>
<td>2.97</td>
</tr>
</tbody>
</table>

7. **Summary and conclusions.** We formulated the VCCMM algebraically, and gave special consideration to the implementation of wells. The local mass conservation (2.4) is modified when a mesh element is either a production well or near an injection well. All coefficients in the scheme (3.10) are computable, for example, with the implementation of the polyline structure and the Sutherland–Hodgman clipping algorithm.

The volume correction adjustment of the method not only preserves a basic physical principle, but also guarantees the monotonicity and maximum and minimum principles of the VCCMM. Moreover, we also proved a strong stability property of the VCCMM; that is, the maximum error due to perturbations is proportional to the initial error and the size of the perturbations, and it grows at most linearly with respect to the simulation time. With a similar technique, we generalized the stability analysis to CMM, where the error could grow at an exponential rate with respect to the simulation time. A few tests were given to verify the stability results of VCCMM and CMM. From these numerical results, the maximum concentration errors are approximately linearly dependent on the size of the perturbations for both methods. The VCCMM shows much smaller errors than CMM over time. However, both methods showed exponential growth in the error over time. For VCCMM, due to the error tolerance in the volume correction adjustment, we lost the strict $L^\infty$-norm bound of the matrix $A_n$ in Corollary 4.5. After imposing this strict bound in the VCCMM through algebraic adjustment (5.1), we saw only linear growth of the error over time, as expected. Thus, our results are consistent with Theorems 4.6 and 4.8.

Computational examples showed that the VCCMM produces no overshoots or undershoots and has less numerical diffusion than CMM and Godunov’s method. Moreover, in certain long-time problems, such might arise in a nuclear waste storage simulation, VCCMM can both better approximate the concentrations and also be more computationally efficient.

**REFERENCES**


