SIAM J. NUMER. ANAL.
Vol. 48, No. 3, pp. 797–823

CONVERGENCE OF A FULLY CONSERVATIVE VOLUME CORRECTED CHARACTERISTIC METHOD FOR TRANSPORT PROBLEMS*

TODD ARBOGAST† AND WEN-HAO WANG‡

Abstract. We consider the convergence of a volume corrected characteristics-mixed method (VCCMM) for advection-diffusion systems. It is known that, without volume correction, the method is first order convergent, provided there is a nondegenerate diffusion term. We consider the advective part of the system and give some properties of the weak solution. With these properties we prove that the volume corrected method, with no diffusion term, gives a lower order L^1-convergence rate of $\mathcal{O}(h/\sqrt{\Delta t} + h + (\Delta t)^r)$, where r is related to the accuracy of the characteristic tracing. This result compares favorably to Godunov's method, but avoids the CFL constraint, so large time steps can be taken in practice. In fact, Godunov's method converges at $\mathcal{O}(h^{1/2})$, which is our result for $\Delta t = Ch$, where now C is not limited. However, the optimal choice, $\Delta t = Ch^{2/(2r+1)}$, gives a better rate, $\mathcal{O}(h^{2r/(2r+1)})$, than Godunov's method, e.g., $\mathcal{O}(h^{2/3})$ if $r = 1$. With a nondegenerate diffusion term, we obtain an L^2-error estimate for the problem. We also prove the existence of, and give an error estimate for, a perturbed velocity field for which the volume is locally conserved. Finally, some convergence tests are given to verify the optimal convergence rate for $r = 1$.

Key words. advection-diffusion, characteristics-mixed method, Lagrangian method, Eulerian–Lagrangian localized adjoint methods, local conservation, convergence, error estimates, total variation boundedness

AMS subject classifications. 35L65, 65M12, 65M15, 65M25, 76S05

DOI. 10.1137/09077415X

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 $c(\mathbf{x}, t)$ in an incompressible bulk fluid moving according to the velocity field $\mathbf{u}(\mathbf{x}, t)$ satisfies the advection-diffusion system

$$\nabla \cdot \mathbf{u} = q \qquad \text{in } \Omega \times J, \tag{1.1}$$

$$(\phi c)_t + \nabla \cdot (c\mathbf{u} - D\nabla c) = q_c := c_I q^+ + cq^- \qquad \text{in } \Omega \times J, \tag{1.2}$$

$$\mathbf{u} \cdot \boldsymbol{\nu} = 0 \qquad \text{on } \partial\Omega \times J, \tag{1.3}$$

$$c(\mathbf{x}, 0) = c^0(\mathbf{x}) \qquad \text{in } \Omega, \tag{1.4}$$

where $J = (0, \infty)$ is the time interval, $q = q(\mathbf{x}, t)$ represents isolated external sources $q^+ = \max\{q, 0\} \geq 0$ and sinks $q^- = q - q^+ \leq 0$, $c_I = c_I(\mathbf{x}, t)$ is the injected concentration, $\phi = \phi(\mathbf{x}) \in [\phi_*, 1]$ $(\phi_* > 0)$ is the storage factor of the medium called porosity, $D = D(\mathbf{x}, t)$ is the diffusion-dispersion tensor (assumed uniformly bounded and positive definite), $c^0 = c^0(\mathbf{x})$ is the initial concentration, subscript t is

*Received by the editors October 19, 2009; accepted for publication (in revised form) March 26, 2010; published electronically June 23, 2010. This work was supported in part by U.S. National Science Foundation grant DMS-0713815.

http://www.siam.org/journals/sinum/48-3/77415.html

†Department of Mathematics, The University of Texas at Austin, 1 University Station C1200, Austin, TX 78712, and Institute for Computational Engineering and Sciences, The University of Texas at Austin, 1 University Station C0200, Austin, TX 78712 (arbogast@ices.utexas.edu).

‡Institute for Computational Engineering and Sciences, The University of Texas at Austin, 1 University Station C0200, Austin, TX 78712 (wwang@ices.utexas.edu).

time partial differentiation, and $\boldsymbol{\nu}$ is the outward unit normal vector with respect to $\partial\Omega$. By a *dilute* tracer we mean that c does not change the overall velocity $\mathbf{u}$.

Note that we have two fluids in this problem: the tracer fluid and the ambient fluid. The mass conservation principle requires that we conserve both fluids locally over regions of space. Since the fluids are incompressible, we can more easily describe the situation as (1) *local mass conservation* of the tracer c and (2) *local volume conservation* of the combined fluid. Numerical methods should respect both these conservation principles over the computational mesh (i.e., locally). We call such methods *fully conservative*.

Moving mesh and characteristic methods have been developed to exploit this observation and thereby avoid any CFL constraint. Characteristic methods became viable in 1982 when Douglas and Russell introduced a Lagrangian formulation called the modified method of characteristics (MMOC) [12, 13, 9] (see also [20]). Because MMOC is based on points, it violates *both* local mass and volume constraints. A modification of the method (MMOC with adjusted advection) produced a global mass balance [10, 22], but not a local mass balance.

Eulerian–Lagrangian schemes have been developed to approximate the advection-diffusion equation (1.2), using Lagrangian characteristic methods for the transport and a fixed Eulerian grid for the diffusion. Included are the Eulerian–Lagrangian localized adjoint methods (ELLAM) [5, 7, 23, 25, 24] and the characteristics-mixed method (CMM) [1, 3] and its two-phase variant [11], which are ELLAM schemes but emphasize their development in terms of the local mass constraint.

Eulerian numerical methods based on fixed grids, such as Godunov's method [18], are locally mass conservative by design. They are also automatically volume conserving, since the volumes of the fixed grid elements do not change in time.

The volume corrected characteristics-mixed method (VCCMM) was introduced in [2]. It treats the advective part of the transport problem (i.e., $D = 0$) using a Lagrangian or characteristic method. It is based on the transport not of a single point or fluid particle, but rather the mass in an entire region of fluid. The mass is transported along the characteristic curves of the hyperbolic part of the transport equation. However, since the shape of a characteristic trace-back region must be approximated in numerical implementation, its volume may be incorrect. This is equivalent to mass conservation errors for the ambient fluid. The volume corrected method gives an efficient algorithm for adjustment of the trace-back points so that volume is conserved locally. This leads to a fully conservative characteristic method.

Without adjustment, in [3], it is proven that the method is first order convergent in the mesh spacing parameter h with a nondegenerate diffusion-dispersion tensor. Without diffusion-dispersion (i.e., $D = 0$), due to projection error accumulation [19], piecewise discontinuous constant approximations can be only $\mathcal{O}(h/\sqrt{\Delta t} + h + (\Delta t)^r)$, where r is related to the accuracy of the characteristic tracing (see Remark 7.2). With $D = 0$, the volume correction preserves the convergence of the VCCMM. That is, in the fully degenerate diffusion-dispersion tensor case, we preserve the accuracy $\mathcal{O}(h/\sqrt{\Delta t} + h + (\Delta t)^r)$ at the same time the adjustment recovers the volume conservation. Note that the optimal choice is $\Delta t = Ch^{2/(2r+1)}$ for a convergence rate $\mathcal{O}(h^{2r/2r+1})$, which tends to $\mathcal{O}(h)$ as r becomes large (i.e., we more accurately trace characteristics and take Δt large). This is better than using, e.g., Godunov's method, which is both $\mathcal{O}(h^{1/2})$ and CFL time step limited.

The rest of the paper is organized as follows. Section 2 gives a review of conservative characteristic methods and derives the local mass constraints for tracer and combined fluids. Section 3 gives an analytical representation of the weak solution and

introduces the entropy inequality. Section 4 lists and proves some properties of the weak solution and the numerical solution that are relevant to our purposes. Section 5 introduces an approximation of L^1-errors and proves some properties that play an important role in the proof of convergence. Section 6 gives the L^1-convergence result for the method with $D = 0$. We also note that the proof of convergence for the CMM [3] extends to the VCCMM, giving an L^2-error estimate for the problem with a nondegenerate diffusion term. Section 7 gives the existence and an error estimate of the perturbed velocity field, which presents the major difficulty of our overall proof. A few convergence tests are given in section 8, verifying our theoretical results. Finally, the summary and concluding remarks are given in the last section.

2. A review of conservative characteristic methods and local mass constraints. In the rest of the paper, we treat only the advective part of the system; i.e., we set $D = 0$. Furthermore, since $0 < \phi_* \le \phi(\mathbf{x}) \le 1$, without losing generality, we assume $\phi(\mathbf{x}) \equiv 1$ for simplicity. That is, we consider variable $\tilde{c} := \phi c$ as the new conserved quantity and introduce the *interstitial* velocity $\mathbf{v} := \mathbf{u}/\phi$, $\tilde{c}_I := \phi c_I$, and $\tilde{q} := q/\phi$. However, we continue to use the notations c, $\mathbf{u}$, c_I, and q. Therefore, the system (1.1)–(1.4) can be reduced to

$$\nabla \cdot (\phi \mathbf{u}) = \phi q \qquad \text{in } \Omega \times J, \tag{2.1}$$

$$c_t + \nabla \cdot (\mathbf{u}c) = q_c := c_I q^+ + c q^- \qquad \text{in } \Omega \times J, \tag{2.2}$$

$$\mathbf{u} \cdot \boldsymbol{\nu} = 0 \qquad \text{on } \partial\Omega \times J, \tag{2.3}$$

$$c(\mathbf{x}, 0) = c^0(\mathbf{x}) \qquad \text{in } \Omega. \tag{2.4}$$

Suppose we have a time interval $J_T := [0, T]$ and a grid $0 = t^0 < t^1 < \cdots < t^N = T$. In one time step $J^n := [t^n, t^{n+1})$, the characteristic trace-back $\check{\mathbf{x}}(t) = \check{\mathbf{x}}(\mathbf{x}, t) = \check{\mathbf{x}}_{n+1}(\mathbf{x}, t)$ passing through $(\mathbf{x}, t^{n+1})$ will solve

$$\check{\mathbf{x}}_t = \mathbf{u}(\check{\mathbf{x}}, t), \quad t \in J^n, \tag{2.5}$$

$$\check{\mathbf{x}}(t^{n+1}) = \mathbf{x}, \tag{2.6}$$

unless the particle were to trace to the boundary of the domain, which is excluded by our boundary condition (2.3). (We may omit the subscript of $\check{\mathbf{x}}_{n+1}$ if there is no confusion in the context.)

Let Ω be partitioned into elements $\mathcal{T}_h$ of maximal diameter h. Let $E \in \mathcal{T}_h$ be an element of Ω, and define the space-time trace-back region of E as

$$\mathcal{E} = \mathcal{E}_E^{n+1} := \{(\check{\mathbf{x}}, t) \in \Omega \times J^n : \check{\mathbf{x}} = \check{\mathbf{x}}_{n+1}(\mathbf{x}, t),\ \mathbf{x} \in E\},$$

and the fixed time slice

$$\check{E}(t) = \check{E}_{n+1}(t) := \{(\check{\mathbf{x}}, t) \in \Omega \times \{t\} : \check{\mathbf{x}} = \check{\mathbf{x}}_{n+1}(\mathbf{x}, t),\ \mathbf{x} \in E\}.$$

Then $E = \check{E}(t^{n+1})$, and the trace-back region of E is $\check{E} = \check{E}(t^n)$.

Let $\boldsymbol{\nu}_{t,\mathbf{x}} := (\nu_t, \boldsymbol{\nu}_{\mathbf{x}})^T$ be the unit outward normal vector to $\partial\mathcal{E}$ and let

$$\mathcal{S} = \mathcal{S}_E^{n+1} := \{(\check{\mathbf{x}}, t) \in \partial\mathcal{E}_E : \check{\mathbf{x}} = \check{\mathbf{x}}_{n+1}(\mathbf{x}; t),\ \mathbf{x} \in \partial E\}$$

be the space boundary of the space-time region $\mathcal{E}$. Since $\boldsymbol{\nu}_{t,\mathbf{x}}$ is the unit outward normal vector to $\mathcal{S}$, which is defined by curves tracing in the direction $(1, \mathbf{u})^T$, we have the orthogonality

$$\begin{pmatrix} 1 \\ \mathbf{u} \end{pmatrix} \cdot \boldsymbol{\nu}_{t,\mathbf{x}} = 0 \quad \text{on } \mathcal{S}. \tag{2.7}$$

Notice that (2.2) can be rewritten as the space-time divergence

$$\nabla_{t,\mathbf{x}} \cdot \left[c \begin{pmatrix} 1 \\ \mathbf{u} \end{pmatrix} \right] = q_c \quad \text{in } \Omega \times J^n. \tag{2.8}$$

Since $\mathcal{E}$ does not touch $\partial\Omega \times J^n$ by (2.3), applying the divergence theorem and (2.7) to (2.8) gives

$$\iint_{\mathcal{E}} q_c \, d\mathbf{x} \, dt = \iint_{\partial\mathcal{E}} c \begin{pmatrix} 1 \\ \mathbf{u} \end{pmatrix} \cdot \boldsymbol{\nu}_{t,\mathbf{x}} \, d\mathbf{x} \, dt = \int_E c^{n+1} \, d\mathbf{x} - \int_{\check{E}} c^n \, d\mathbf{x},$$

which means the *local mass constraint* is

$$\int_E c^{n+1} \, d\mathbf{x} = \int_{\check{E}} c^n \, d\mathbf{x} + \iint_{\mathcal{E}} q_c \, d\mathbf{x} \, dt, \tag{2.9}$$

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

Due to the approximations both of the characteristics (2.5) and of $\check{E}$ by a polygon, we actually trace to an approximation $\tilde{E}$ of $\check{E}$. Therefore, the numerical solution

$$c_h^{n+1} \in W_h(\Omega) := \{w \in L^2(\Omega) : w|_E \text{ is a constant for all } E \in \mathcal{T}_h\}$$

is defined on E to be

$$c_{h,E}^{n+1} |E| = \int_{\tilde{E}} c_h^n \, d\mathbf{x} + \iint_{\tilde{\mathcal{E}}} q_{c_h^n} \, d\mathbf{x} \, dt, \tag{2.10}$$

where we define $\tilde{\mathcal{E}}$ in (2.15) below as the space-time trace-back region from E to $\tilde{E}$, $|E|$ is the volume of E in the sense of Lebesgue measure in $\mathbb{R}^d$, and $q_{c_h^n} := c_I q^+ + c_h^n q^-$. The numerical solution $c_{h,E}^{n+1}$ is computable since $\tilde{E}$ and $\tilde{\mathcal{E}}$ have replaced $\check{E}$ and $\mathcal{E}$, respectively. We may refer to this method as a *conservative characteristic method*. It is a type of Lagrangian method.

With $c = c_I = \phi$ in (2.9), we have the transport of the single combined fluid (2.1), and, writing $|S|_\phi := \int_S \phi \, d\mathbf{x}$ for the pore volume of a set $S \subset \Omega$,

$$|E|_\phi = |\check{E}|_\phi + \iint_{\mathcal{E}} \phi q \, d\mathbf{x} \, dt. \tag{2.11}$$

We call (2.11) the *local volume constraint*, since the fluid incompressibly fills the pores. However, it is not likely that

$$|E|_\phi = |\tilde{E}|_\phi + \iint_{\tilde{\mathcal{E}}} \phi q \, d\mathbf{x} \, dt, \tag{2.12}$$

leading to a violation of an important physical principle.

The *volume corrected characteristics-mixed method* [2] is an Eulerian–Lagrangian method for approximating the solution. It includes an important procedure for further perturbing the trace-back element $\tilde{E}$ so that (2.12) holds. When $D = 0$, there is no Eulerian mixed method approximation of the diffusion/dispersion, and so we may refer to the remaining Lagrangian steps as the volume corrected, *fully conservative characteristic method*. We assume that $q = 0$ except in isolated elements of $\mathcal{T}_h$. Assuming for simplicity that the elements are rectangles, given $E \in \mathcal{T}_h$, we trace the four vertices as well as the four midpoints to obtain the unadjusted octagonal polygon $\tilde{E}$. The full algorithm is developed in [2]. A very brief description of the adjustment algorithm follows.

The volume correction algorithm.

Point adjustment in time. A trace-back point may be adjusted *in time* by a small amount [10], along the characteristics in the direction of the flow field. As we will see, the effect is to convert spatial errors into time errors. Moreover, in this way, no bias is introduced into the *direction* of the flow. This time adjustment is needed in Steps 1 and 2 below.

Step 1: Forward trace out of injection wells. Trace forward (not backward; see, e.g., (3.1)–(3.2)) the injection wells [16], and then adjust the trace-forward boundary in time according to the well volume constraint.

Step 2: Ring adjustment. Between the wells, starting adjacent to the injection well and moving towards the production wells, entire rings of elements are adjusted in time so as to have the correct volume. Assuming the trace-back ring edge closest to the injector has been adjusted, the points on the far edge are adjusted simultaneously.

Step 3: Individual element adjustment. Within an adjusted ring of elements, individual elements are adjusted to have the correct volume by traversing the ring, starting from a no-flow boundary if one intersects the ring. This is accomplished by a transverse movement of the midpoints (not a time adjustment).

For consistency of the trace-back tessellation, we tacitly assume that the time step is restricted so that the trace-back elements $\tilde{E}$ do not self intersect. Moreover, we assume that no sink traces all the way to a source within a single time step. For simplicity of exposition in this paper, we will not directly treat step 1, although the ideas presented here should extend to this case.

We use a key idea introduced by Arbogast and Wheeler [3], wherein it was noted that an analysis of inexact characteristic tracing, i.e., approximation of the solution to (2.5)–(2.6), could be made if one views the approximate tracing as arising from exact tracing through a perturbed velocity field. We will construct this perturbed velocity $\tilde{\mathbf{u}}$ such that each trace-back of element $E \in \mathcal{T}_h$ is the volume corrected $\tilde{E}$. That is, we replace $\mathbf{u}$ in (2.5)–(2.6) by $\tilde{\mathbf{u}}$ and solve for $\tilde{\mathbf{x}}(t) = \tilde{\mathbf{x}}(\mathbf{x}, t) = \tilde{\mathbf{x}}^{n+1}(\mathbf{x}, t)$ the approximate tracing

$$\tilde{\mathbf{x}}_t = \tilde{\mathbf{u}}(\tilde{\mathbf{x}}, t), \quad t \in J^n, \tag{2.13}$$

$$\tilde{\mathbf{x}}(t^{n+1}) = \mathbf{x}. \tag{2.14}$$

Then, for each $E \in \mathcal{T}_h$, we can define the numerical space-time region

$$\tilde{\mathcal{E}} = \tilde{\mathcal{E}}_E^{n+1} = \{(\tilde{\mathbf{x}}, t) \in \Omega \times J^n : \tilde{\mathbf{x}} = \tilde{\mathbf{x}}_{n+1}(\mathbf{x}, t),\ \mathbf{x} \in E\}, \tag{2.15}$$

and the numerical fixed time slice

$$\tilde{E}(t) = \tilde{E}_{n+1}(t) = \{(\tilde{\mathbf{x}}, t) \in \Omega \times \{t\} : \tilde{\mathbf{x}} = \tilde{\mathbf{x}}_{n+1}(\mathbf{x}, t),\ \mathbf{x} \in E\},$$

for which $E = \tilde{E}(t^{n+1})$ and the volume corrected trace-back region of E is $\tilde{E} = \tilde{E}(t^n)$.

However, the existence of $\tilde{\mathbf{u}}$ and the estimate of the error $(\mathbf{u} - \tilde{\mathbf{u}})$ present the major difficulty. The construction of $\tilde{\mathbf{u}}$ will be given in section 7. For now, we simply make the following assumption. We use $\|\cdot\|_{p,S}$ to denote the norm of $L^p(S)$ and we may omit S if $S = \Omega$ or $\Omega \times J_T$.

Assumption 2.1 (perturbed velocity field). The velocity field $\mathbf{u} = \mathbf{u}(\mathbf{x}, t) \in C^1(\Omega \times J_T)$ has divergence $\nabla \cdot \mathbf{u}(\cdot, t)$ uniformly Lipschitz continuous in time J_T, i.e.,

$$|\nabla \cdot \mathbf{u}(\mathbf{x}, t) - \nabla \cdot \mathbf{u}(\mathbf{y}, t)| \le L|\mathbf{x} - \mathbf{y}| \quad \text{for all } \mathbf{x}, \mathbf{y} \in \Omega,\ t \in J_T, \tag{2.16}$$

where $L > 0$ is a constant independent of $\mathbf{x}$, $\mathbf{y}$, and t. There exists a locally conservative velocity field $\tilde{\mathbf{u}} = \tilde{\mathbf{u}}(\mathbf{x}, t)$ on $\Omega \times J_T$ such that

$$\tilde{\mathbf{u}} \cdot \boldsymbol{\nu} = 0 \quad \text{on } \partial\Omega \times J_T, \tag{2.17}$$

each trace-back polygon $\tilde{E}$ satisfies the local volume constraint (2.12), and

$$\|\mathbf{u} - \tilde{\mathbf{u}}\|_\infty + \|\nabla \cdot \mathbf{u} - \nabla \cdot \tilde{\mathbf{u}}\|_\infty \le C(h + (\Delta t)^r), \tag{2.18}$$

where C and $r > 0$ are constants independent of h and Δt.

Assume c_h^0 is a given initial approximation of c^0. In each time step J^n, we now consider c_h a solution to the perturbed system

$$(c_h)_t + \nabla \cdot (c_h \tilde{\mathbf{u}}) = q_{c_h} \quad \text{in } \Omega \times J^n, \tag{2.19}$$

$$c_h(\mathbf{x}, t^n) = c^n(\mathbf{x}) \quad \text{in } \Omega, \tag{2.20}$$

and we define the update at t^{n+1} as

$$c_h^{n+1}(\mathbf{x}) := P_h c_h(\mathbf{x}, t^{n+1}-) = P_h c_h^{n+1-}(\mathbf{x}), \tag{2.21}$$

where the L^2-projection operator P_h is defined as

$$(P_h f, w) = (f, w) \quad \text{for all } w \in W_h(\Omega). \tag{2.22}$$

3. An analytical representation of the weak solution and the entropy inequality. Taking advantage of the linear structure of transport equation (2.2), as is well known, we can actually solve system (2.2)–(2.4) analytically by integration along characteristics. Let $\hat{\mathbf{x}} = \hat{\mathbf{x}}(\mathbf{x}, t)$ be the trace-forward characteristics of $\mathbf{u}$, i.e.,

$$\hat{\mathbf{x}}_t = \hat{\mathbf{u}} \quad \text{in } \Omega \times J_T, \tag{3.1}$$

$$\hat{\mathbf{x}}(\mathbf{x}, 0) = \mathbf{x} \quad \text{in } \Omega, \tag{3.2}$$

where $\hat{f}(\mathbf{x}, t) := f(\hat{\mathbf{x}}(\mathbf{x}, t), t)$ is the evaluation along trace-forward characteristics for a generic scalar or vector-valued function f.

LEMMA 3.1 (analytical representation). *Let* $\mathbf{u}$ *be a smooth velocity field on the domain* $\Omega \times J_T$ *and* $\hat{\mathbf{x}}$ *be the trace-forward characteristics of* $\mathbf{u}$ *defined in* (3.1)–(3.2). *For any* $t \in J_T$, *assume* $\hat{\mathbf{x}}(\cdot, t)$ *is a diffeomorphism in* Ω, *and denote the inverse as* $\check{\mathbf{x}}(\cdot, t)$. *Then the weak solution to system* (2.2)–(2.4) *evaluated along characteristics is given by*

$$\hat{c} = F_0 + F_1 c^0, \tag{3.3}$$

where

$$F_1(\mathbf{x}, t) := \exp\left(\int_0^t (q^- - \nabla \cdot \mathbf{u})^\wedge(\mathbf{x}, s)\, ds \right), \tag{3.4}$$

$$F_0(\mathbf{x}, t) := \int_0^t f_0(\mathbf{x}, t, s)\, ds, \tag{3.5}$$

$$f_0(\mathbf{x}, t, s) := (c_I q^+)^\wedge(\mathbf{x}, s) \frac{F_1(\mathbf{x}, t)}{F_1(\mathbf{x}, s)}. \tag{3.6}$$

Proof. Rearrange (2.2), and we have

$$c_t + \mathbf{u} \cdot \nabla c = c_I q^+ + (q^- - \nabla \cdot \mathbf{u})c.$$

Notice that

$$(\hat{c})_t = (c(\hat{\mathbf{x}}(\mathbf{x}, t), t))_t = \widehat{c_t} + \hat{\mathbf{u}} \cdot \nabla \hat{c},$$

so $\hat{c}$ solves the well-posed initial value problem of an ordinary differential equation

$$\begin{aligned} (\hat{c})_t &= (c_I q^+)^\wedge + (q^- - \nabla \cdot \mathbf{u})^\wedge \hat{c} \quad \text{in } \Omega \times J_T, \\ \hat{c}(\mathbf{x}, 0) &= c^0(\mathbf{x}) \qquad\qquad\qquad\qquad\quad \text{in } \Omega. \end{aligned}$$

Then we obtain (3.3) by solving the ordinary differential equation above. ∎

The analytical representation implies the existence and uniqueness of the weak solution.

COROLLARY 3.2 (existence and uniqueness). *If the trace-forward characteristics* $\hat{\mathbf{x}}$ *of* $\mathbf{u}$ *form a diffeomorphism, then there exists a unique weak solution* c *to system* (2.1)–(2.4) *given by* (3.3).

By the theory of conservation laws, the weak solution $c = c(\mathbf{x}, t)$ also satisfies a stability condition, which is called the *entropy inequality* or *entropy admissibility condition*, relative to a convex entropy η; that is,

$$\eta_t + \nabla \cdot \mathbf{Q} \le H$$

in the sense of distributions, i.e.,

$$\begin{aligned} (\eta^n, \varphi^n) - (\eta^{n+1-}, \varphi^{n+1}) + \int_{J^n} (\eta, \varphi_t)\, dt \\ - \int_{J^n} \langle \mathbf{Q} \cdot \boldsymbol{\nu}, \varphi \rangle_{\partial\Omega}\, dt + \int_{J^n} (\mathbf{Q}, \nabla\varphi)\, dt + \int_{J^n} (H, \varphi)\, dt \ge 0 \end{aligned} \tag{3.7}$$

for any nonnegative test function $\varphi = \varphi(\mathbf{x}, t) \in C^\infty(\Omega \times J^n)$. Any convex function $\eta = \eta(c)$ may serve as an entropy [6, p. 54], with the associated entropy flux $\mathbf{Q}$ and entropy production H computed by

$$\mathbf{Q} = \eta\, \mathbf{u} \quad \text{and} \quad H = \eta' q_c + (\eta - \eta' c)\nabla \cdot \mathbf{u}.$$

Note that the term involving $\mathbf{Q} \cdot \boldsymbol{\nu}$ in (3.7) vanishes by the boundary condition (2.3). In general, the entropy solution is the weak solution which is physically relevant. In our case, there is only one solution, and we will use (3.7) freely.

4. Properties of the weak solution. It is well known from the theory of scalar conservation laws, with a flux $\mathbf{F}$ in the canonical form

$$c_t + \nabla \cdot \mathbf{F}(c) = 0 \quad \text{in } \mathbb{R}^d \times \mathbb{R}^+, \tag{4.1}$$

$$c(\mathbf{x}, 0) = c^0(\mathbf{x}) \qquad \text{in } \mathbb{R}^d, \tag{4.2}$$

that the law has reached a state of virtual completeness, such as L^1-contraction, uniqueness, L^∞-monotonicity, uniform boundedness, and total variation diminishing (TVD) properties of the entropy solution [6, pp. 126–142].

It should be noted that our transport equation (2.2) is similar to, but not a subcase of, the canonical form (4.1), which is homogeneous, and the flux $\mathbf{F}$ does not explicitly depend on spatial and time variables, but only on the conserved quantity c. In this section, we prove some properties of the weak solution to the system (2.1)–(2.4) that are relevant to our purposes in the following analysis.

4.1. Uniform boundedness. Physically, the tracer mass comes from the initial state and the injected concentration as time proceeds. Indeed, by the analytical representation (3.3), it is easy to see the uniform boundedness of the weak solution.

LEMMA 4.1 (boundedness of the weak solution). *If $c^0 \in L^\infty(\Omega)$ and $c_I, q \in L^\infty(\Omega \times J_T)$, the weak solution c to the system* (2.1)–(2.4) *is uniformly L^∞-bounded and L^1-bounded in $\Omega \times J_T$.*

Proof. By the analytical representation (3.3), it is easy to see the uniform L^∞-boundedness of c. Then the L^1-boundedness of c follows due to the boundedness of the space-time domain $\Omega \times J_T$. ∎

LEMMA 4.2 (boundedness of the numerical solution). *If $c_h^0 \in L^\infty(\Omega)$ and $c_{I,h}$, $q_h \in L^\infty(\Omega \times J_T)$, the numerical solution c_h to the system* (2.19)–(2.21) *is uniformly L^∞-bounded and L^1-bounded in $\Omega \times J_T$.*

Proof. Notice that the L^2-projection operator defined in (2.22) increases neither the L^∞- nor L^1-norm of a function, so we can perform an argument similar to that in Lemma 4.1 for c_h defined in (2.19)–(2.21) in each time step J^n to complete the proof. ∎

4.2. Boundedness of the total variation. Variations of solutions play an important role in hyperbolic differential equations. In this subsection, we list and prove some basic properties of functions of bounded variation, prove the total variation boundedness (TVB) property of the weak solution, and make an assumption on the L^1-TVB property of the numerical solution.

4.2.1. Properties of functions of bounded variation. The total variation of a function f on Ω is defined by

$$|f|_{BV(\Omega)} = \sup_{\boldsymbol{\varphi}} (f, \nabla \cdot \boldsymbol{\varphi})_{L^2(\Omega)}, \tag{4.3}$$

where the supremum is taken for all vector-valued functions $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_d)^T \in [C_c^\infty(\Omega)]^d$ with $\|\boldsymbol{\varphi}\|_\infty := \max_{1 \le i \le d} \|\varphi_i\|_\infty \le 1$. We denote $BV(\Omega) := \{f \in L^1(\Omega) : |f|_{BV(\Omega)} < \infty\}$ to be the set of L^1 functions of bounded variation on Ω. Then $|\cdot|_{BV(S)}$ is a semi-norm on $BV(S)$, and we may omit S if $S = \Omega$. If $f \in W^{1,1}(\Omega)$, integrating by parts, we have

$$|f|_{BV} = \|\nabla f\|_1 := \sum_{i=1}^{d} \|\partial_i f\|_1 < \infty, \tag{4.4}$$

and so $W^{1,1}(\Omega) \subset BV(\Omega) \subset L^1(\Omega)$.

PROPOSITION 4.3. *If the domain Ω has a partition $\mathcal{T}$, then for any $f \in BV(\Omega)$,*

$$\sum_{E \in \mathcal{T}} |f|_{BV(E)} \le |f|_{BV(\Omega)}.$$

Proposition 4.3 is trivial to prove by definition (4.3).

PROPOSITION 4.4 (lower semicontinuity). *The BV seminorm is lower semicontinuous with respect to the L^1-topology; i.e., if $f_j \to f$ in $L^1(\Omega)$, then*

$$|f|_{BV} \le \liminf_{j \to \infty} |f_j|_{BV}.$$

Proof. See [15, p. 7]. ∎

PROPOSITION 4.5 (approximation by smooth functions). *For any $f \in BV(\Omega)$, there exists a sequence $\{f_j\}$ in $C^\infty(\Omega)$ such that $f_j \to f$ in $L^1(\Omega)$ and $|f_j|_{BV} \to |f|_{BV}$.*

Proof. See [15, p. 14]. ∎

PROPOSITION 4.6 (product rule). *For any $f \in BV(\Omega)$ and $g \in W^{1,\infty}(\Omega)$, the product $fg \in BV(\Omega)$, and*

$$|fg|_{BV} \le |f|_{BV}\|g\|_\infty + \|f\|_1\|\nabla g\|_\infty. \tag{4.5}$$

Proof. First, suppose $f \in C^\infty(\Omega)$. Taking L^1-norms on both sides of the identity

$$\nabla(fg) = g\nabla f + f\nabla g,$$

we obtain (4.5). Now for general $f \in BV(\Omega)$, by Proposition 4.5, there is a sequence $\{f_j\}$ in $C^\infty(\Omega)$ such that $f_j \to f$ in $L^1(\Omega)$ and $|f_j|_{BV} \to |f|_{BV}$. Now $f_j g \to fg$ in $L^1(\Omega)$. By Proposition 4.4 and (4.5) for smooth functions, we have

$$\begin{aligned} |fg|_{BV} &\le \liminf_{j\to\infty} |f_j g|_{BV} \le \liminf_{j\to\infty} (|f_j|_{BV}\|g\|_\infty + \|f_j\|_1\|\nabla g\|_\infty) \\ &= |f|_{BV}\|g\|_\infty + \|f\|_1\|\nabla g\|_\infty. \end{aligned}$$ ∎

PROPOSITION 4.7 (composition rule). *For any $f \in BV(\Omega)$ and diffeomorphism g on Ω, the composition $f \circ g \in BV(\Omega)$, and*

$$|f \circ g|_{BV} \le \|\nabla g\|_\infty \|\det(\nabla g^{-1})\|_\infty |f|_{BV}. \tag{4.6}$$

Proof. For $f \in C^\infty(\Omega)$, by taking L^1-norms on both sides of the identity

$$\nabla(f \circ g) = \nabla g(\nabla f) \circ g$$

and changing variables, we obtain (4.6). The result for general $f \in BV(\Omega)$ follows from Propositions 4.4 and 4.5 as in the previous proof. ∎

PROPOSITION 4.8 (difference quotient). *If the domain Ω is convex, then the integral of the difference quotient is bounded by the total variation. That is, for any $f \in BV(\Omega)$,*

$$\sup_{\mathbf{y}\neq 0} \|D_{\mathbf{y}} f\|_{1,\Omega_{\mathbf{y}}} \le |f|_{BV(\Omega)}, \tag{4.7}$$

where $D_{\mathbf{y}} := |\mathbf{y}|^{-1}(T_{\mathbf{y}} - I)$ is the difference quotient operator with the translation operator $T_{\mathbf{y}}$ defined by

$$(T_{\mathbf{y}} f)(\mathbf{x}) = f(\mathbf{x} + \mathbf{y}),$$

and $\Omega_{\mathbf{y}} = \Omega \cap (\Omega - \{\mathbf{y}\})$ is the restricted domain on which the integral is well defined.

Proof. By Proposition 4.5, we need only show (4.7) for $f \in C^\infty(\Omega)$. For any $\mathbf{y} \in \mathbb{R}^d$, $\mathbf{y} \neq 0$, and $\mathbf{x} \in \Omega_{\mathbf{y}}$, if $\Omega_{\mathbf{y}}$ is not empty, we have the identity

$$f(\mathbf{x} + \mathbf{y}) - f(\mathbf{x}) = \int_0^1 \nabla f(\mathbf{x} + s\mathbf{y}) \cdot \mathbf{y}\, ds.$$

Taking norms on both sides and integrating with respect to $\mathbf{x} \in \Omega_{\mathbf{y}}$, we have

$$\begin{aligned} \|D_{\mathbf{y}} f\|_{1,\Omega_{\mathbf{y}}} &= \frac{1}{|\mathbf{y}|} \int_{\Omega_{\mathbf{y}}} \left| \int_0^1 \nabla f(\mathbf{x} + s\mathbf{y}) \cdot \mathbf{y}\, ds \right| d\mathbf{x} \\ &\le \int_0^1 \int_{\Omega_{\mathbf{y}}} |\nabla f(\mathbf{x} + s\mathbf{y})|\, d\mathbf{x}\, ds \le \|\nabla f\|_{1,\Omega} = |f|_{BV(\Omega)}. \end{aligned}$$ ∎

4.2.2. TVB property. Since we have a balance law, i.e., a conservation law in an inhomogeneous form (2.2), unfortunately, we cannot expect it to obey the TVD property in general. However, since we study the solution in a bounded time interval J_T and the transport equation (2.2) is linear, the physical behavior of the solution should continuously change as time proceeds. It is natural to expect the solution to be TVB in J_T under some regularity assumptions of the data in the system (2.1)–(2.4). Denote

$$V(\Omega) := L^\infty(\Omega) \cap W^{1,1}(\Omega),$$
$$V(J_T^k;\Omega) := L^\infty(\Omega \times J_T^k) \cap C(J_T^k; W^{1,1}(\Omega)),$$

where k is a positive integer. Note $f \in C(J_T^k; W^{m,p}(\Omega))$ means $f(\cdot, t_1, \ldots, t_k) \in W^{m,p}(\Omega)$ and $\|f(\cdot, t_1, \ldots, t_k)\|_{W^{m,p}(\Omega)}$ is continuous for each $t_j \in J_T$.

Assumption 4.1 (regularity of data). Velocity field $\mathbf{u} \in C^1(\Omega \times J_T)$ with diffeomorphic characteristics $\hat{\mathbf{x}}$, $\nabla \cdot \mathbf{u}$ satisfies the uniform Lipschitz condition (2.16), $q \in C(J_T; W^{1,\infty}(\Omega))$, $c^0 \in V(\Omega)$, and $c_I \in V(J_T;\Omega)$.

Assumption 4.2 (regularity of initial approximation). $c_h^0 \in L^\infty(\Omega) \cap BV(\Omega)$.

Furthermore, we impose the following assumptions on the time and space domain discretizations.

Assumption 4.3 (regularity of 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

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

where $\Delta t^n := t^{n+1} - t^n$ and $\Delta t := \sup_n \Delta t^n$.

Assumption 4.4 (shape regularity of domain discretization). The mesh $\mathcal{T}_h$ of bounded domain Ω is convex and regular, i.e., each element $E \in \mathcal{T}_h$ is convex, and there exists a constant $\lambda_2 > 0$ such that

$$\sup_{E \in \mathcal{T}_h} \frac{h_E}{\rho_E} \le \lambda_2,$$

where h_E and ρ_E are the outer and inner diameters of element E, respectively, and the mesh spacing parameter $h := \sup_{E \in \mathcal{T}_h} h_E < \infty$.

LEMMA 4.9 (TVB of the weak solution). *Let Assumption* 4.1 *hold. Then the weak solution c to the system* (2.2)–(2.4) *is TVB to time T. Moreover,*

$$|c(\cdot,t)|_{BV} \le C_0 t + e^{C_1 t} |c^0|_{BV}, \tag{4.8}$$

where $C_0 > 0$ and $C_1 > 0$ are constants independent of t.

Proof. By Assumption 4.1, we see from (3.4)–(3.6) that $F_1 \in W^{1,\infty}(\Omega)$, $f_0 \in V(J_T^2;\Omega)$, and $F_0 \in V(J_T;\Omega)$. By the analytical representation (3.3), we have

$$|\hat{c}(\cdot,t)|_{BV} \le |F_0(\cdot,t)|_{BV} + |(F_1 c^0)(\cdot,t)|_{BV}, \tag{4.9}$$

and

$$|F_0(\cdot,t)|_{BV} = \|\nabla F_0(\cdot,t)\|_1 = \left\| \int_0^t \nabla f_0(\cdot,t,s)\, ds \right\|_1 \le \int_0^t \|\nabla f_0(\cdot,t,s)\|_1 ds \le t \, \|f_0\|_{V(J_T^2;\Omega)}. \tag{4.10}$$

By Proposition 4.6, since $F_1(\mathbf{x},t) = \exp(\int_0^t \hat{f}_1(\mathbf{x},s)\,ds)$ with $f_1 := q^- - \nabla\cdot\mathbf{u} \in W^{1,\infty}(\Omega)$ is an exponential,

$$
\begin{aligned}
|(F_1c^0)(\cdot,t)|_{BV} &\le \|F_1(\cdot,t)\|_\infty |c^0|_{BV} + \|\nabla F_1(\cdot,t)\|_\infty \|c^0\|_1 \\
&\le \|F_1(\cdot,t)\|_\infty \left(|c^0|_{BV} + \left\| \nabla \int_0^t \hat{f}_1(\cdot,s)\,ds \right\|_\infty \|c^0\|_1 \right) \\
&\le \exp\big(t\|f_1\|_\infty\big)\big(|c^0|_{BV} + t\|\nabla f_1\|_\infty \|\nabla\hat{\mathbf{x}}\|_\infty \|c^0\|_1\big).
\end{aligned} \tag{4.11}
$$

Substituting (4.10) and (4.11) into (4.9) gives

$$
|\hat{c}(\cdot,t)|_{BV} \le C_0 t + e^{C_1 t}|c^0|_{BV}. \tag{4.12}
$$

Noticing that $c = \hat{c}\circ\check{\mathbf{x}}$, and $\nabla\check{\mathbf{x}}(\cdot,0) \equiv I$ by (3.2), we have, by Proposition 4.7,

$$
|c(\cdot,t)|_{BV} \le (1+Ct)|\hat{c}(\cdot,t)|_{BV} \tag{4.13}
$$

for some constant $C > 0$. Combining (4.12) and (4.13) gives (4.8) and completes the proof. ∎

For a general mesh $\mathcal{T}_h$ in multidimensions, the L^2-projection operator P_h might increase the variation of a function, so we cannot expect the TVB property to hold for the numerical solution. Instead, we make a weaker assumption of L^1-TVB as follows.

Assumption 4.5 (L^1-TVB of the numerical solution). The numerical solution c_h^n to the system is uniformly L^1-TVB; i.e., there exists a constant $M > 0$ such that

$$
|c_h^n|_{L^1(J_T;BV)} := \sum_{n=0}^{N-1} |c_h^n|_{BV}\Delta t^n \le M
$$

for any h, $\Delta t > 0$.

In particular, for rectangular meshes, it is well known that the L^2-projection operator P_h is TVD; i.e., for any $f \in BV(\Omega)$,

$$
|P_h f|_{BV} \le |f|_{BV}. \tag{4.14}
$$

Therefore, the numerical solution is TVB and L^1-TVB.

5. An approximation of errors in the L^1-norm. In this section, we introduce an approximation of errors in the L^1-norm that plays an important role later in the convergence proof in section 6. This approximation was first introduced by Kuznetsov [17] in the error estimates of conservation law (4.1)–(4.2) by the smoothing method and the viscosity method. It was later used by Lucier [19] in the error estimates of Glimm's method and Godunov's method.

Without losing generality, we assume $0 \in \Omega$. Let K_ε be an approximation of the Dirac distribution in Ω, i.e.,

$$
K_\varepsilon(\mathbf{x}) := \frac{1}{\varepsilon^d} K_0\left(\frac{\mathbf{x}}{\varepsilon}\right), \qquad \varepsilon > 0,
$$

where function K_0 is nonnegative, smooth, and compactly supported in Ω with an integral of one. For the weak solution c and the numerical solution c_h, we introduce

$$
\rho_{\varepsilon,h}^n := \iint_{\Omega\times\Omega} K_\varepsilon(\mathbf{x}-\mathbf{y})|c^n(\mathbf{x}) - c_h^n(\mathbf{y})|\,d\mathbf{x}\,d\mathbf{y}. \tag{5.1}
$$

Since $c_h^n = P_h c_h^{n-}$, we also have $\rho_{\varepsilon,h}^{n-}$ defined with c_h^n replaced by c_h^{n-}.

By a proof similar to that of Remark 6.13 in [14], we have the following lemma.

LEMMA 5.1. *The quantity $\rho^n_{\varepsilon,h}$ is an approximation of the L^1-error with a first order convergence rate with respect to ε. That is,*

$$\left| \rho^n_{\varepsilon,h} - \|c^n - c^n_h\|_1 \right| \le C\varepsilon, \tag{5.2}$$

where $C > 0$ is a constant independent of ε, h, and n.

By changing variables, the definition (5.1) of $\rho^n_{\varepsilon,h}$ can be rewritten as

$$\rho^n_{\varepsilon,h} = \int_\Omega K_0(\mathbf{x}) \|T_{\varepsilon\mathbf{x}} c^n - c^n_h\|_{1,\Omega_{\varepsilon\mathbf{x}}} d\mathbf{x}.$$

Then we can again employ the entropy inequality (3.7) to prove the following lemma, which gives the estimate of the change of $\rho^n_{\varepsilon,h}$ in time.

LEMMA 5.2. *The change of $\rho^n_{\varepsilon,h}$ in a single time step $J^n = [t^n, t^{n+1})$ has the estimate*

$$\rho^{n+1-}_{\varepsilon,h} - \rho^n_{\varepsilon,h} \le C(\varepsilon + h + (\Delta t)^r)\Delta t^n, \tag{5.3}$$

where r is given in (2.18) *and $C > 0$ is a constant independent of ε, h, Δt^n, and n.*

Proof. Let $\varepsilon > 0$ and $\mathbf{x} \in \Omega$ be fixed. Notice that from (2.2) and (2.19) the translated difference $d_{\varepsilon\mathbf{x},h} := T_{\varepsilon\mathbf{x}} c - c_h$ solves the linear balance law

$$(d_{\varepsilon\mathbf{x},h})_t + \nabla \cdot (d_{\varepsilon\mathbf{x},h} \tilde{\mathbf{u}}) = d_{\varepsilon\mathbf{x},h} q^- + R_{\varepsilon\mathbf{x}} \quad \text{in } \Omega_{\varepsilon\mathbf{x}} \times J^n,$$

where the remainder

$$R_{\mathbf{x}} := \nabla \cdot \big((T_{\mathbf{x}} c)(\tilde{\mathbf{u}} - T_{\mathbf{x}}\mathbf{u})\big) + \big(T_{\mathbf{x}}(c_I q^+) - c_I q^+\big) + (T_{\mathbf{x}} c)(T_{\mathbf{x}} q^- - q^-) \text{ for } \mathbf{x} \in \mathbb{R}^d.$$

For entropy $\eta(d) = |d|$ and test function $\varphi(\mathbf{x}, t) \equiv 1$, the entropy inequality (3.7) is reduced to

$$\begin{aligned} &\|d^n_{\varepsilon\mathbf{x},h}\|_{1,\Omega_{\varepsilon\mathbf{x}}} - \|d^{n+1-}_{\varepsilon\mathbf{x},h}\|_{1,\Omega_{\varepsilon\mathbf{x}}} \\ &\quad - \int_{J^n} \int_{\partial\Omega_{\varepsilon\mathbf{x}}} |d_{\varepsilon\mathbf{x},h}|\, \tilde{\mathbf{u}} \cdot \boldsymbol{\nu}\, ds\, dt + \int_{J^n} \|R_{\varepsilon\mathbf{x}}\|_{1,\Omega_{\varepsilon\mathbf{x}}} dt \ge 0, \end{aligned} \tag{5.4}$$

where, with Lemmas 4.1 and 4.2, Assumption 2.1, and (2.3),

$$\begin{aligned} \int_{\partial\Omega_{\varepsilon\mathbf{x}}} |d_{\varepsilon\mathbf{x},h}|\, \tilde{\mathbf{u}} \cdot \boldsymbol{\nu}\, ds &\le |\partial\Omega|\, \|d_{\varepsilon\mathbf{x},h}\|_\infty \|\tilde{\mathbf{u}} \cdot \boldsymbol{\nu}\|_{\infty,\partial\Omega_{\varepsilon\mathbf{x}}} \\ &\le |\partial\Omega|\, \|d_{\varepsilon\mathbf{x},h}\|_\infty (\|\mathbf{u} \cdot \boldsymbol{\nu}\|_{\infty,\partial\Omega_{\varepsilon\mathbf{x}}} + \|(\tilde{\mathbf{u}} - \mathbf{u}) \cdot \boldsymbol{\nu}\|_{\infty,\partial\Omega_{\varepsilon\mathbf{x}}}) \\ &\le |\partial\Omega|\, \|d_{\varepsilon\mathbf{x},h}\|_\infty (\|(T_{\varepsilon\mathbf{x}}\mathbf{u}) \cdot \boldsymbol{\nu}\|_{\infty,\partial\Omega} + \|\tilde{\mathbf{u}} - \mathbf{u}\|_\infty) \\ &= |\partial\Omega|\, \|d_{\varepsilon\mathbf{x},h}\|_\infty (\|(T_{\varepsilon\mathbf{x}}\mathbf{u} - \mathbf{u}) \cdot \boldsymbol{\nu}\|_{\infty,\partial\Omega} + \|\tilde{\mathbf{u}} - \mathbf{u}\|_\infty) \\ &\le |\partial\Omega|(\|c\|_\infty + \|c_h\|_\infty)\, (\varepsilon h_\Omega \|\nabla\mathbf{u}\|_\infty + C'(h + (\Delta t)^r)) \\ &\le C\, (\varepsilon + h + (\Delta t)^r) \end{aligned}$$

and, with also Assumption 2.1 and Propositions 4.6 and 4.8, for any $t \in J^n$,

$$\begin{aligned}
\|R_{\varepsilon\mathbf{x}}\|_{1,\Omega_{\varepsilon\mathbf{x}}} &\le \|\nabla(T_{\varepsilon\mathbf{x}}c)\|_{1,\Omega_{\varepsilon\mathbf{x}}}\|\tilde{\mathbf{u}} - T_{\varepsilon\mathbf{x}}\mathbf{u}\|_\infty + \|T_{\varepsilon\mathbf{x}}c\|_{1,\Omega_{\varepsilon\mathbf{x}}}\|\nabla\cdot\tilde{\mathbf{u}} - T_{\varepsilon\mathbf{x}}\nabla\cdot\mathbf{u}\|_\infty \\
&\quad + \|T_{\varepsilon\mathbf{x}}(c_I q^+) - c_I q^+\|_{1,\Omega_{\varepsilon\mathbf{x}}} + \|(T_{\varepsilon\mathbf{x}}c)(T_{\varepsilon\mathbf{x}}q^- - q^-)\|_{1,\Omega_{\varepsilon\mathbf{x}}} \\
&\le \|\nabla c\|_1(\|\tilde{\mathbf{u}} - \mathbf{u}\|_\infty + \|\mathbf{u} - T_{\varepsilon\mathbf{x}}\mathbf{u}\|_\infty) \\
&\quad + \|c\|_1(\|\nabla\cdot\tilde{\mathbf{u}} - \nabla\cdot\mathbf{u}\|_\infty + \|\nabla\cdot\mathbf{u} - T_{\varepsilon\mathbf{x}}\nabla\cdot\mathbf{u}\|_\infty) \\
&\quad + \varepsilon h_\Omega(|c_I q^+|_{BV} + \|c\|_\infty |q|_{BV}) \\
&\le \|\nabla c\|_1\left[C'(h + (\Delta t)^r) + \varepsilon h_\Omega\|\nabla\mathbf{u}\|_\infty\right] + \|c\|_1\left[C'(h + (\Delta t)^r) + \varepsilon h_\Omega L\right] \\
&\quad + \varepsilon h_\Omega(|c_I|_{BV}\|q\|_\infty + \|c_I\|_1\|\nabla q\|_\infty + \|c\|_\infty |q|_{BV}) \\
&\le C\left(\varepsilon + h + (\Delta t)^r\right).
\end{aligned}$$

So (5.4) will be

$$\|d^n_{\varepsilon\mathbf{x},h}\|_{1,\Omega_{\varepsilon\mathbf{x}}} - \|d^{n+1-}_{\varepsilon\mathbf{x},h}\|_{1,\Omega_{\varepsilon\mathbf{x}}} + C(\varepsilon + h + (\Delta t)^r)\Delta t^n \ge 0$$

for some constant $C > 0$. Multiplying by $K_0(\mathbf{x})$ and integrating with respect to $\mathbf{x} \in \Omega$, we obtain (5.3) and complete the proof. ∎

The following lemma gives an estimate of the projection error measured by $\rho^n_{\varepsilon,h}$.

LEMMA 5.3. *The projection error has the estimate*

$$\rho^n_{\varepsilon,h} - \rho^{n-}_{\varepsilon,h} \le C\frac{h^2}{\varepsilon}|c^{n-}_h|_{BV(\Omega)}, \tag{5.5}$$

where $C > 0$ is a constant independent of ε, h, and n.

Proof. We compute

$$\begin{aligned}
&\rho^n_{\varepsilon,h} - \rho^{n-}_{\varepsilon,h} \\
&= \iint_{\Omega\times\Omega} K_\varepsilon(\mathbf{x}-\mathbf{y})\{|c^n(\mathbf{x}) - P_h c^{n-}_h(\mathbf{y})| - |c^n(\mathbf{x}) - c^{n-}_h(\mathbf{y})|\}\,d\mathbf{x}\,d\mathbf{y} \\
&= \int_\Omega \sum_{E\in\mathcal{T}_h}\int_E K_\varepsilon(\mathbf{x}-\mathbf{y})\left\{\left|c^n(\mathbf{x}) - \frac{1}{|E|}\int_E c^{n-}_h(\mathbf{z})\,d\mathbf{z}\right| - |c^n(\mathbf{x}) - c^{n-}_h(\mathbf{y})|\right\}d\mathbf{x}\,d\mathbf{y} \\
&\le \int_\Omega \sum_{E\in\mathcal{T}_h}\frac{1}{|E|}\iint_{E\times E} K_\varepsilon(\mathbf{x}-\mathbf{y})\{|c^n(\mathbf{x}) - c^{n-}_h(\mathbf{z})| - |c^n(\mathbf{x}) - c^{n-}_h(\mathbf{y})|\}\,d\mathbf{z}\,d\mathbf{y}\,d\mathbf{x}.
\end{aligned}$$

If we switch variables $\mathbf{y}$ and $\mathbf{z}$ in the last inequality, the value simply changes sign, so the inequality can be written as

$$\begin{aligned}
&\rho^n_{\varepsilon,h} - \rho^{n-}_{\varepsilon,h} \\
&\quad\le \frac{1}{2}\int_\Omega \sum_{E\in\mathcal{T}_h}\frac{1}{|E|}\iint_{E\times E}\{K_\varepsilon(\mathbf{x}-\mathbf{y}) - K_\varepsilon(\mathbf{x}-\mathbf{z})\} \\
&\qquad\qquad\qquad\qquad \times\{|c^n(\mathbf{x}) - c^{n-}_h(\mathbf{z})| - |c^n(\mathbf{x}) - c^{n-}_h(\mathbf{y})|\}\,d\mathbf{z}\,d\mathbf{y}\,d\mathbf{x} \\
&\quad\le \frac{1}{2}\int_\Omega \sum_{E\in\mathcal{T}_h}\frac{1}{|E|}\iint_{E\times E}|K_\varepsilon(\mathbf{x}-\mathbf{y}) - K_\varepsilon(\mathbf{x}-\mathbf{z})||c^{n-}_h(\mathbf{z}) - c^{n-}_h(\mathbf{y})|\,d\mathbf{z}\,d\mathbf{y}\,d\mathbf{x} \\
&\quad= \frac{1}{2}\sum_{E\in\mathcal{T}_h}\frac{1}{|E|}\iint_{E\times E}\int_\Omega |K_\varepsilon(\mathbf{x}-\mathbf{y}) - K_\varepsilon(\mathbf{x}-\mathbf{z})|\,d\mathbf{x}\,|c^{n-}_h(\mathbf{z}) - c^{n-}_h(\mathbf{y})|\,d\mathbf{z}\,d\mathbf{y}.
\end{aligned} \tag{5.6}$$

For any $\mathbf{y}, \mathbf{z} \in E$, we have by Proposition 4.8 that

$$\int_\Omega |K_\varepsilon(\mathbf{x}-\mathbf{y}) - K_\varepsilon(\mathbf{x}-\mathbf{z})|\,d\mathbf{x} = \int_{\varepsilon^{-1}(\Omega-\{\mathbf{y}\})} \left| K_0(\mathbf{x}) - K_0\left(\mathbf{x} + \frac{\mathbf{y}-\mathbf{z}}{\varepsilon}\right)\right| d\mathbf{x} \le \frac{|\mathbf{y}-\mathbf{z}|}{\varepsilon}|K_0|_{BV(\Omega)} \le \frac{h}{\varepsilon}|K_0|_{BV(\Omega)}, \tag{5.7}$$

and

$$\iint_{E\times E} |c_h^{n-}(\mathbf{z}) - c_h^{n-}(\mathbf{y})|d\mathbf{z}d\mathbf{y} = \int_E \left(\int_{E-\{\mathbf{z}\}} |c_h^{n-}(\mathbf{z}) - c_h^{n-}(\mathbf{y}+\mathbf{z})|d\mathbf{y}\right) d\mathbf{z} \le \int_{E-E}\left(\int_{E_\mathbf{y}} |c_h^{n-}(\mathbf{z}) - c_h^{n-}(\mathbf{y}+\mathbf{z})|d\mathbf{z}\right) d\mathbf{y} \le h|E-E|\;|c_h^{n-}|_{BV(E)}, \tag{5.8}$$

where $E - E := \{\mathbf{x}-\mathbf{y} : \mathbf{x}, \mathbf{y} \in E\}$.

Let B_r be a ball in $\mathbb{R}^d$ with radius $r > 0$; then by regularity of $\mathcal{T}_h$ in Assumption 4.4, we have

$$\frac{|E-E|}{|E|} \le \frac{|B_{h_E}|}{|B_{\rho_E/2}|} = \left(\frac{2h_E}{\rho_E}\right)^d \le (2\lambda_2)^d \quad \text{for any } E \in \mathcal{T}_h.$$

Substituting (5.7) and (5.8) into (5.6), we have by Proposition 4.3 that

$$\rho_{\varepsilon,h}^n - \rho_{\varepsilon,h}^{n-} \le \frac{1}{2}\sum_{E\in T_h} \frac{1}{|E|}\;\frac{h}{\varepsilon}|K_0|_{BV(\Omega)}\;h|E-E|\;|c_h^{n-}|_{BV(E)} \le \frac{h^2}{2\varepsilon}|K_0|_{BV(\Omega)}\;(2\lambda_2)^d\,|c_h^{n-}|_{BV(\Omega)} \le C\frac{h^2}{\varepsilon}|c_h^{n-}|_{BV(\Omega)}. \qquad \square$$

6. Convergence results. When $D = 0$, we have the following theorem.

THEOREM 6.1 (fully degenerate diffusion). *Let Assumptions* 2.1 *and* 4.1–4.5 *hold* (*or omit Assumption* 4.5 *and assume* $\mathcal{T}_h$ *is rectangular*). *Then for the problem* (2.1)–(2.4), *the following* L^1*-error estimate holds:*

$$\max_{0\le n\le N} \|c_h^n - c^n\|_1 \le \|c_h^0 - c^0\|_1 + C\left(\frac{h}{\sqrt{\Delta t}} + h + (\Delta t)^r\right), \tag{6.1}$$

where $C > 0$ *is a constant independent of* h *and* Δt.

Proof. Summing (5.3) for n in Lemma 5.2, we have

$$\sum_{k=0}^{n-1}(\rho_{\varepsilon,h}^{k+1-} - \rho_{\varepsilon,h}^k) \le C(\varepsilon + h + (\Delta t)^r)t^n \le CT(\varepsilon + h + (\Delta t)^r).$$

Rearranging, we have

$$\rho_{\varepsilon,h}^n \le \rho_{\varepsilon,h}^0 + E_{\varepsilon,h}^n + CT(\varepsilon + h + (\Delta t)^r), \tag{6.2}$$

where the total projection error is

$$E_{\varepsilon,h}^n := \sum_{k=1}^n (\rho_{\varepsilon,h}^k - \rho_{\varepsilon,h}^{k-}).$$

Summing (5.5) for n in Lemma 5.3, we have

$$E^n_{\varepsilon,h} \le C\frac{h^2}{\varepsilon}\sum_{k=1}^{n}|c_h^{k-}|_{BV}. \tag{6.3}$$

By (4.8) in time step J^{k-1},

$$|c_h^{k-}|_{BV} \le C_0\Delta t^{k-1} + e^{C_1\Delta t^{k-1}}|c_h^{k-1}|_{BV},$$

so substituting into (6.3) gives

$$\begin{aligned} E^n_{\varepsilon,h} &\le C\frac{h^2}{\varepsilon}\left(C_0T + e^{C_1T}\sum_{k=1}^{n}|c_h^{k-1}|_{BV}\right) \\ &\le C\frac{h^2}{\varepsilon\Delta t}\left(C_0T^2 + \lambda_1 e^{C_1T}\sum_{k=1}^{n}|c_h^{k-1}|_{BV}\Delta t^{k-1}\right) \\ &\le C\frac{h^2}{\varepsilon\Delta t}\left(C_0T^2 + \lambda_1 e^{C_1T}|c_h^k|_{L^1(J_T;BV)}\right) \\ &\le C\frac{h^2}{\varepsilon\Delta t}\left(C_0T^2 + \lambda_1 M e^{C_1T}\right), \end{aligned} \tag{6.4}$$

where $|c_h^k|_{L^1(J_T;BV)} \le M$ by Assumption 4.5. Combining (6.2), (6.4), and (5.2) gives

$$\|c_h^n - c^n\|_1 \le \|c_h^0 - c^0\|_1 + C\left(\varepsilon + \frac{h^2}{\varepsilon\Delta t} + h + (\Delta t)^r\right),$$

where the optimal choice for ε is to take $\varepsilon = h/\sqrt{\Delta t}$, which completes the proof. ∎

For the full system (1.1)–(1.4) with a nondegenerate diffusion-dispersion tensor D, the techniques used in the proof of CMM in [3] extend to the VCCMM using Assumption 2.1. The only difference is that we have a locally conservative perturbed velocity $\tilde{\mathbf{u}}$ corresponding to the *volume corrected* characteristic trace-back regions. Within the proof, the estimate of $\|\mathbf{u}-\tilde{\mathbf{u}}\|_\infty + \|\nabla\cdot(\mathbf{u}-\tilde{\mathbf{u}})\|_\infty$ changes now by adding an extra $\mathcal{O}(h)$. However, this extra error is negligible under the assumptions in [3, Theorem 1], and so similar conclusions on the L^2-error hold.

To be precise, we need to define the full mixed approximation. The method uses operator splitting. Over each time step, we first solve the advective part as described above in (2.10) for C_h^{n+1}. We convert this back to the form of the concentration used initially in section 1 by defining $\bar{c}_h^{n+1} = C_h^{n+1}|E|/|E|_\phi$. We then approximate the diffusive part of the system. Let $\mathbf{V}_h$ be the lowest order Raviart–Thomas space [21] for the vector variable, which in our case is the diffusive flux

$$z = -D\nabla c,$$

approximated by $z_h^{n+1} \in \mathbf{V}_h$ solving, for each $E \in \mathcal{T}_h$ and $v \in \mathbf{V}_h$,

$$\left(c_h^{n+1} - \bar{c}_h^{n+1}\right)|E|_\phi = \Delta t^n \int_E \nabla\cdot z_h^{n+1}\,dx, \tag{6.5}$$

$$\int_\Omega D^{-1}z_h^{n+1}\,v\,dx = \int_\Omega c_h^{n+1}\,\nabla\cdot v\,dx. \tag{6.6}$$

We can then convert back to $C_h^{n+1} = c_h^{n+1}|E|_\phi/|E|$ for the next advective step.

In fact, one can use the postprocessing step described in [3, equation (3.2)] to improve the concentration, which we denote by $\tilde{c}_h^{n+1}$. Finally, we state the convergence theorem of VCCMM for the nondegenerate diffusion case as follows.

THEOREM 6.2 (nondegenerate diffusion). *Let the assumptions in* [3, Theorem 1] *and Assumption* 2.1 *hold. If $c \in W^{3,2}(\Omega)$ and the initial error and time steps satisfy*

$$\|\tilde{c}_h^0 - c^0\|_2 \le Ch^2 \quad \textit{and} \quad \Delta t^n \ge Ch^2$$

for some $C > 0$ independent of h and Δt, then for h and Δt sufficiently small,

$$\max_{0\le n\le N} \|\tilde{c}_h^n - c^n\|_2 \le K(h^2 + \Delta t), \tag{6.7}$$

$$\max_{0\le n\le N} \|c_h^n - c^n\|_2 + \left(\sum_{n=1}^{N} \|z_h^n - z^n\|_2^2 \, \Delta t^n \right)^{1/2} \le K(h + \Delta t), \tag{6.8}$$

where $K > 0$ is a constant independent of h and Δt.

In fact, from [3], (6.7) is $\mathcal{O}(h^{3/2} + \Delta t)$ under the same assumptions except $c \in W^{2,2}(\Omega)$ and a more restricted $\Delta t^n \ge Ch^{3/2}$, but the result was improved by Arbogast and Professor Ivan Yotov in an unpublished note from 1995. We merely record here changes to the proof. We improve the estimate [3, equation (6.44)], which is not optimal. Using the notation of that paper (so, e.g., $\|\cdot\|_2$ becomes $\|\cdot\|_0$ and $\eta = C_h - c$), we have

$$\begin{aligned} &\|(\check{\phi}\check{\tilde{\eta}})^{n-1,+} - \phi^{n-1}\tilde{\eta}^{n-1}\|_0 h \\ &\quad \le \|\check{\phi}^{n-1,+}(\check{\tilde{\eta}}^{n-1,+} - \tilde{\eta}^{n-1})\|_0 h + \|(\check{\phi}^{n-1,+} - \phi^{n-1})\tilde{\eta}^{n-1}\|_0 h \\ &\quad \le C\|\check{\tilde{\eta}}^{n-1,+} - \tilde{\eta}^{n-1}\|_0 h + \|\check{\phi}^{n-1,+} - \phi^{n-1}\|_{L^\infty}\|\tilde{\eta}^{n-1}\|_0 h \\ &\quad \le C(\|\check{\tilde{\eta}}^{n-1,+} - \tilde{\eta}^{n-1}\|_0 h + \|\tilde{\eta}^{n-1}\|_0 h\Delta t^n), \end{aligned} \tag{6.9}$$

since, by [3, Lemma 3],

$$\|\check{\phi}^{n-1,+} - \phi^{n-1}\|_{L^\infty} \le \|\check{\phi}^{n-1,+} - \phi^n\|_{L^\infty} + \|\phi^n - \phi^{n-1}\|_{L^\infty} \le C\Delta t^n.$$

For the estimate of the first term on the right-hand side above, let $\tilde{P}_h$ be a quadratic projection such that

$$\|\tilde{P}_h f - f\|_0 \le C\|f\|_{W^{3,2}(\Omega)} h^3 \quad \text{for all } f \in W^{3,2}(\Omega). \tag{6.10}$$

By an inverse inequality, [3, equation (6.43)], [3, Lemma 2], and (6.10),

$$\begin{aligned} &\|\check{\tilde{\eta}}^{n-1,+} - \tilde{\eta}^{n-1}\|_0 h \\ &\quad \le C\|(\check{\tilde{C}}_h^{n-1,+} - ((\tilde{P}_h c)^\vee)^{n-1,+}) - (\tilde{C}_h^{n-1} - (\tilde{P}_h c)^{n-1})\|_{-1} \\ &\qquad + \left\{ \|((\tilde{P}_h c)^\vee)^{n-1,+} - \check{c}^{n-1,+}\|_0 + \|(\tilde{P}_h c)^{n-1} - c^{n-1}\|_0 \right\} h \\ &\quad \le C \left\{ \|\tilde{C}_h^{n-1} - (\tilde{P}_h c)^{n-1}\|_0 \Delta t^n + \|c^{n-1}\|_{W^{3,2}(\Omega)} h^4 \right\} \\ &\quad \le C \left\{ \left(\|\tilde{C}_h^{n-1} - c^{n-1}\|_0 + \|c^{n-1} - (\tilde{P}_h c)^{n-1}\|_0 \right) \Delta t^n + h^4 \right\} \\ &\quad \le C \left\{ \|\tilde{\eta}^{n-1}\|_0 \Delta t^n + h^3 \Delta t^n + h^4 \right\}. \end{aligned} \tag{6.11}$$

Combining (6.9) and (6.11) gives the improvement of [3, equation (6.44)]

$$\|(\check{\phi}\check{\tilde{\eta}})^{n-1,+} - \phi^{n-1}\tilde{\eta}^{n-1}\|_0 h \le C \left\{ \|\tilde{\eta}^{n-1}\|_0 \Delta t^n + h^3 \Delta t^n + h^4 \right\}. \tag{6.12}$$

We then follow the same argument as in [3] to obtain (6.7).

7. The existence of the perturbed velocity. In this section, we make several assumptions that will guarantee the existence of the perturbed velocity field $\tilde{\mathbf{u}}$ satisfying the requirements of Assumption 2.1. That is, we prove Assumption 2.1 by constructing a perturbed velocity field $\tilde{\mathbf{u}} = \tilde{\mathbf{u}}(\mathbf{x}, t)$ on the domain $\Omega \times J_T$. We need to impose assumptions on the choices of rings that are adjusted in Step 2 of the volume correction algorithm in section 2. For simplicity, we concentrate on the case that the domain $\Omega \subset \mathbb{R}^2$, although the ideas can carry over to higher spatial dimensions. Below we consider the effect of three main steps of volume adjustment: characteristic time perturbation, ring adjustment, and individual element adjustment. Note that, for ease of exposition, we do not treat forward tracing around wells, though clearly the ideas of the proof extend to this step.

Remark 7.1. In the rest of this section, we tacitly assume that the velocity field $\mathbf{u}$ is given by a quarter of a "five-spot" pattern of wells, which is a rectangular domain with an injection well near a corner and a production well near the opposite corner.

7.1. Point adjustment in time and the local definition of $\tilde{\mathbf{u}}$. The following lemma constructs a perturbed velocity locally at isolated points and quantifies how a small trace-back time perturbation of size $\alpha\Delta t^n$ changes a single characteristic trace-back. The proof only involves straightforward calculation and is omitted.

LEMMA 7.1. *Suppose $\alpha \in \mathbb{R}$ is fixed and $\mathbf{x} \in \overline{\Omega}$. For $t \in J^n$, let*

$$\tilde{\mathbf{x}}(t) = \tilde{\mathbf{x}}(\mathbf{x}, t) := \check{\mathbf{x}}\big(\mathbf{x}, t + \alpha(t^{n+1} - t)\big) \tag{7.1}$$

be a time perturbation of the trace-back curve $\check{\mathbf{x}}(t)$. Then the perturbed velocity

$$\tilde{\mathbf{u}}(\mathbf{x}, t) := (1 - \alpha)\, \mathbf{u}\big(\mathbf{x}, t + \alpha(t^{n+1} - t)\big) \tag{7.2}$$

has $\tilde{\mathbf{x}}$ as its characteristic passing through point $\mathbf{x}$ at time t^{n+1}. Moreover,

$$\tilde{\mathbf{u}} \cdot \boldsymbol{\nu} = 0 \quad \text{on } \partial\Omega \times J^n, \tag{7.3}$$

$$\|\mathbf{u} - \tilde{\mathbf{u}}\|_\infty + \|\nabla \cdot \mathbf{u} - \nabla \cdot \tilde{\mathbf{u}}\|_\infty \le C|\alpha|, \tag{7.4}$$

where $C = (\|\mathbf{u}_t\|_\infty + \|\nabla \cdot \mathbf{u}_t\|_\infty)T + \|\mathbf{u}\|_\infty + \|\nabla \cdot \mathbf{u}\|_\infty$.

Remark 7.2. In practice, the ordinary differential equation (2.5) for characteristics cannot be solved exactly unless the velocity field is particularly simple. Therefore, numerical techniques are needed. For example, if the single Euler step is used, then we actually trace back from a point $\mathbf{x}_0$ with local velocity field

$$\mathbf{u}_{\mathrm{E}}(\mathbf{x}, t) := \mathbf{u}(\mathbf{x}_0, t^{n+1}),$$

where $\mathbf{x} = \mathbf{x}_0 - (t^{n+1} - t)\mathbf{u}(\mathbf{x}_0, t^{n+1})$ for $t \in J^n$. This leads to an error

$$\|\mathbf{u}_{\mathrm{E}} - \mathbf{u}\|_\infty + \|\nabla \cdot \mathbf{u}_{\mathrm{E}} - \nabla \cdot \mathbf{u}\|_\infty \le C(\Delta t)^r,$$

where $r = 1$. Since $\mathbf{u} = \mathbf{u}_{\mathrm{E}} + (\mathbf{u} - \mathbf{u}_{\mathrm{E}})$, we simply replace $\mathbf{u}$ by $\mathbf{u}_{\mathrm{E}}$, and the rest of the analysis remains unchanged except that there is an extra error due to approximately solving for characteristics. In general, we may use an approximation of order $r > 1$. For ease of exposition, we tacitly omit this extra error term in this section.

Remark 7.3. If $\mathbf{u}$ is unknown, then we may need to approximate $\mathbf{u}$ with $\mathbf{u}_h$ by numerical techniques, which leads to some error $\mathcal{O}(h^{r_1} + \Delta t^{r_2})$, where r_1 and $r_2 > 0$. If so, this error would enter the estimates as well.

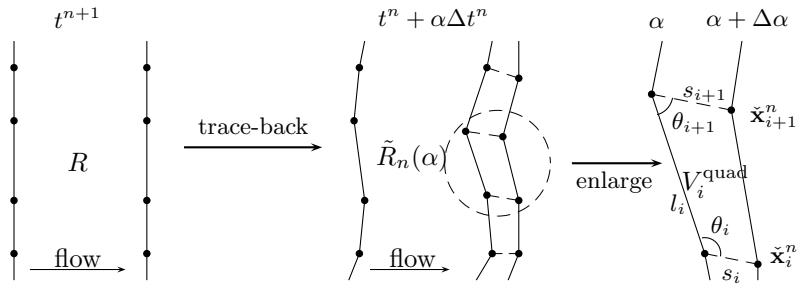

FIG. 7.1. *Ring R at time t^{n+1} is traced back to time t^n and approximated by $\tilde{R}_n$. The solid dots represent the points which are traced back. The exterior boundary of $\tilde{R}_n$ is perturbed in location by a time change of $\alpha\Delta t^n$ and $(\alpha + \Delta\alpha)\Delta t^n$ along the direction of characteristics.*

7.2. Global $\tilde{\mathbf{u}}$ and the ring adjustment. Now consider the ring adjustment phase of the volume correction algorithm. We have defined a local perturbed velocity field $\tilde{\mathbf{u}}$ for a single characteristic in Lemma 7.1. Here we further perturb $\tilde{\mathbf{u}}$ to obtain volume conservation over rings of elements.

At time t^{n+1}, let $R \subset \Omega$ be a ring (Figure 7.1, left) and $\check{R}_n$ be the exact trace-back region with velocity field $\mathbf{u}$ for time Δt^n. Vertices and midpoints $\mathbf{x}_i$ on ∂R are traced back to $\check{\mathbf{x}}^n_i$, where $1 \le i \le N_R$. Without losing generality, assume points $\check{\mathbf{x}}^n_i$, where $1 \le i \le N_{\text{ext}}$ for some $N_{\text{ext}} < N_R$, are on the "exterior" boundary (i.e., away from injection sites) of $\check{R}_n$ which need to be adjusted. We perturb these points in time of size $\alpha\Delta t^n$ as defined in Lemma 7.1, i.e., $\tilde{\mathbf{x}}^n_i = \check{\mathbf{x}}(\mathbf{x}_i, t^n + \alpha\Delta t^n)$, $1 \le i \le N_{\text{ext}}$. Denote this perturbed trace-back polygon as $\tilde{R}_n(\alpha)$ (Figure 7.1, middle) with the "exterior" boundary $\Gamma_n(\alpha)$. We will choose the ring such that the shape of the ring is approximately "perpendicular" to the direction of the flow; that is, the volume change of the ring should be sensitive to the adjustment in the direction of the flow.

Assumption 7.1. There exists a constant $C' > 0$ independent of h such that the number of vertices and midpoints on ∂R satisfies $N_R \le C' h^{-1}$.

Assumption 7.2 (monotonicity and differentiability). When $\alpha_1 \le \alpha_2$, $\tilde{R}_n(\alpha_1) \subseteq \tilde{R}_n(\alpha_2)$, and the pore volume $V_n(\alpha) := |\tilde{R}_n(\alpha)|_\phi$ is differentiable with respect to α.

Assumption 7.3 (nondegeneracy). There exist constants $\phi_* > 0$, $u_* > 0$, and $\Gamma_* > 0$ such that $1 \ge \phi(\mathbf{x}) \ge \phi_*$ in Ω, $|\mathbf{u}| \ge u_*$ in a sufficiently large neighborhood of $\Gamma_n(\alpha)$, and $|\Gamma_n(\alpha)| \ge \Gamma_*$.

Assumption 7.4 (nonparallelism). There exists a constant $\nu_* > 0$ such that $\mathbf{u} \cdot \boldsymbol{\nu}_\alpha \ge \nu_* |\mathbf{u}|$ in a neighborhood of $\Gamma_n(\alpha)$, where $\boldsymbol{\nu}_\alpha$ is the unit outward normal vector with respect to $\Gamma_n(\alpha)$.

Remark 7.4. The condition $|\Gamma_n(\alpha)| \ge \Gamma_*$ in Assumption 7.3 implies that the trace-back procedure should only be performed away from injection wells, where points do not trace into the well-bore and become arbitrarily close. Therefore, a trace-forward technique is used near injection wells in the volume correction algorithm. We note also that $|\mathbf{u}| > u_*$ in Assumption 7.3 does not cover the case of velocity fields with stagnation points when $\Gamma_n(\alpha)$ is near the point. The "sufficiently large" condition is defined in the proof of Lemma 7.3 (see (7.10)).

The following lemma shows the existence of the perturbed velocity field $\tilde{\mathbf{u}}$ such that the trace-back region of a ring R satisfies the *local volume constraint* (2.12) in the absence of source q.

LEMMA 7.2. *Let $R \subset \Omega$ be a ring to be adjusted. If Assumptions* 7.1–7.4 *hold, then there exists some α^* such that*

$$V_n(\alpha^*) = |\check{R}_n|_\phi, \tag{7.5}$$

where $|\alpha^*| \le Ch$ for some constant $C > 0$ independent of n, h, and Δt.

To show Lemma 7.2, we need another lemma which simply says that the change rate of the pore volume $V_n(\alpha)$ is bounded away from zero during the ring adjustment.

LEMMA 7.3. *If Assumptions* 7.1–7.4 *hold, then*

$$V_n'(\alpha) \ge \beta_* \Delta t^n, \tag{7.6}$$

where $\beta_* > 0$ is a constant independent of n, h, and Δt^n.

Proof. For a small $\Delta\alpha > 0$, by Assumptions 7.2 and 7.3, we have

$$V_n(\alpha + \Delta\alpha) - V_n(\alpha) = |\tilde{R}_n(\alpha + \Delta\alpha) \setminus \tilde{R}_n(\alpha)|_\phi \ge \phi_* |\tilde{R}_n(\alpha + \Delta\alpha) \setminus \tilde{R}_n(\alpha)|, \tag{7.7}$$

where the set $\tilde{R}_n(\alpha + \Delta\alpha) \setminus \tilde{R}_n(\alpha)$ can be decomposed as a union of quadrilaterals (Figure 7.1, middle).

As shown in Figure 7.1, right, the volume of each quadrilateral V_i^{quad} ($1 \le i \le N'_{\text{ext}}$, $N'_{\text{ext}} = N_{\text{ext}}$ if the ring R does not intersect $\partial\Omega$ and $N'_{\text{ext}} = N_{\text{ext}} - 1$ otherwise) is

$$\begin{aligned} V_i^{\text{quad}} &= \frac{1}{2}(s_i \sin\theta_i + s_{i+1}\sin\theta_{i+1})(l_i - s_i\cos\theta_i - s_{i+1}\cos\theta_{i+1}) \\ &\quad + \frac{1}{2}s_i^2 \sin\theta_i\cos\theta_i + \frac{1}{2}s_{i+1}^2 \sin\theta_{i+1}\cos\theta_{i+1} \\ &= \frac{1}{2}[\, l_i s_i \sin\theta_i + l_i s_{i+1}\sin\theta_2 - s_i s_{i+1}\sin(\theta_i + \theta_{i+1})\,] \\ &\ge \frac{1}{2}(\, l_i s_i \sin\theta_i + l_i s_{i+1}\sin\theta_{i+1} - s_i s_{i+1}\,). \end{aligned} \tag{7.8}$$

Each displacement s_i is

$$\begin{aligned} s_i &= \left|\check{\mathbf{x}}_i\big(t^n + (\alpha + \Delta\alpha)\Delta t^n\big) - \check{\mathbf{x}}_i(t^n + \alpha\Delta t^n)\right| \\ &= \left|\int_{t^n + \alpha\Delta t^n}^{t^n + (\alpha+\Delta\alpha)\Delta t^n} \mathbf{u}(\check{\mathbf{x}}_i(t), t)\, dt\right| \le \|\mathbf{u}\|_\infty \Delta\alpha\Delta t^n, \end{aligned} \tag{7.9}$$

and by Assumptions 7.3 and 7.4,

$$\begin{aligned} s_i &= \left|\int_{t^n + \alpha\Delta t^n}^{t^n + (\alpha+\Delta\alpha)\Delta t^n} \mathbf{u}(\check{\mathbf{x}}_i(t), t)\, dt\right| \\ &\ge \left|\int_{t^n + \alpha\Delta t^n}^{t^n + (\alpha+\Delta\alpha)\Delta t^n} \mathbf{u}(\check{\mathbf{x}}_i(t), t)\cdot \boldsymbol{\nu}_\alpha\, dt\right| \ge \nu_* u_* \Delta\alpha\Delta t^n. \end{aligned} \tag{7.10}$$

Substituting (7.9), (7.10), and each $\sin\theta_i \ge \nu_*$ by Assumption 7.4 into (7.8) gives

$$V_i^{\text{quad}} \ge \left(\nu_*^2 u_* l_i - \frac{1}{2}\|\mathbf{u}\|_\infty^2 \Delta\alpha\Delta t^n\right)\Delta\alpha\Delta t^n. \tag{7.11}$$

To obtain a lower bound of the difference $V_n(\alpha+\Delta\alpha)-V_n(\alpha)$ in (7.7), summing (7.11) for all $V_i^{\rm quad}$ in the ring $\tilde{R}_n(\alpha)$, by Assumptions 7.1 and 7.3, we have

$$
\begin{aligned}
(7.12)\qquad V_n(\alpha+\Delta\alpha)-V_n(\alpha) &\ge \phi_* \sum_i V_i^{\rm quad} \\
&\ge \phi_* \left(\nu_*^2 u_* |\Gamma_n(\alpha)| - \frac{N_{\rm ext}}{2} \|\mathbf{u}\|_\infty^2 \Delta\alpha \Delta t^n \right) \Delta\alpha \Delta t^n \\
&\ge \phi_* \left(\nu_*^2 u_* \Gamma_* - \frac{C\Delta t^n}{2h} \|\mathbf{u}\|_\infty^2 \Delta\alpha \right) \Delta\alpha \Delta t^n .
\end{aligned}
$$

Divide (7.12) by $\Delta\alpha$ and let $\Delta\alpha \to 0$. We obtain (7.6) with $\beta_* = \phi_* \nu_*^2 u_* \Gamma_*$. ∎

Now we are ready to prove Lemma 7.2.

Proof of Lemma 7.2. For any α in a neighborhood of zero, consider the difference

$$
\begin{aligned}
(7.13)\qquad V_n(\alpha) - |\check{R}_n|_\phi &= (V_n(\alpha) - V_n(0)) + (V_n(0) - |\check{R}_n|_\phi) \\
&= V_n'(\xi)\alpha + (|\tilde{R}_n(0)|_\phi - |\check{R}_n|_\phi),
\end{aligned}
$$

where $\xi = \xi(\alpha)$ comes from the mean value theorem. For the second term on the right-hand side, since $0 \le \phi \le 1$

$$
(7.14)\qquad \big| |\tilde{R}_n(0)|_\phi - |\check{R}_n|_\phi \big| \le \big| (\tilde{R}_n(0) \setminus \check{R}_n) \cup (\check{R}_n \setminus \tilde{R}_n(0)) \big|,
$$

which is the discrepancy of volumes between $\tilde{R}_n(0)$ and $\check{R}_n$. This discrepancy is the sum of the discrepancies associated to each edge.

As illustrated in Figure 7.2, at time t^{n+1}, let e_i $(1 \le i \le N_R)$ be an edge of R with ends $\mathbf{x}_i$ and $\mathbf{x}_{i+1}$ $(\mathbf{x}_{N_R+1} = \mathbf{x}_1)$, which is traced back with velocity $\mathbf{u}$ to a curve $\check{e}_i(t)$ at time $t \in J^n$ with ends $\check{\mathbf{x}}_i(t) = \check{\mathbf{x}}(\mathbf{x}_i, t)$ and $\check{\mathbf{x}}_{i+1}(t) = \check{\mathbf{x}}(\mathbf{x}_{i+1}, t)$. Curve $\check{e}_i(t)$ is approximated by a line segment $\tilde{e}_i(t)$ by connecting $\check{\mathbf{x}}_i(t)$ and $\check{\mathbf{x}}_{i+1}(t)$. Let $\tilde{\mathbf{e}}_i(t) := \check{\mathbf{x}}_{i+1}(t) - \check{\mathbf{x}}_i(t)$. The local discrepancy $V_i^{\rm dis}$ at time t^n associated to edge e_i is the net difference in area using the correct curve $\check{e}_i^n$ versus the segment $\tilde{e}_i^n$.

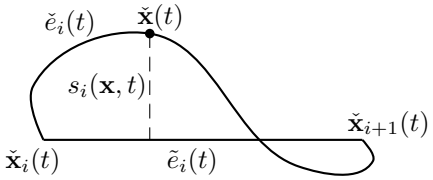

FIG. 7.2. *An edge e of ring R is traced back to a curve $\check{e}_i(t)$ with two ends $\check{\mathbf{x}}_i^n(t)$ and $\check{\mathbf{x}}_{i+1}^n(t)$, which is approximated by a line segment $\tilde{e}_i(t)$.*

For any $\mathbf{x} \in e_i$ which is traced back to $\check{\mathbf{x}}^n(t) = \check{\mathbf{x}}(\mathbf{x}, t^n) \in \check{e}_i(t)$, let

$$
(7.15)\qquad s_i(\mathbf{x},t) := \frac{\det\big(\check{\mathbf{x}}(t) - \check{\mathbf{x}}_i(t), \tilde{\mathbf{e}}_i(t)\big)}{|\tilde{\mathbf{e}}_i(t)|}
$$

be the algebraic distance from point $\check{\mathbf{x}}(t)$ to segment $\tilde{e}_i(t)$, where $\det(\mathbf{x},\mathbf{y})$ is the determinant of a 2×2 matrix formed by column vectors $\mathbf{x}$ and $\mathbf{y}$. Since $\check{\mathbf{x}}(\cdot,t)$ is a diffeomorphism by Assumption 4.1, $|\tilde{\mathbf{e}}_i(t)| \ne 0$ and (7.15) is well defined. Then

$$
(7.16)\qquad V_i^{\rm dis} \le 2\|s_i^n\|_{\infty,e_i} \sup_{\mathbf{x},\mathbf{y}\in e_i} |\check{\mathbf{x}}^n - \check{\mathbf{y}}^n| \le 2\|s_i^n\|_{\infty,e_i} \|\nabla\check{\mathbf{x}}\|_\infty h,
$$

where $\|\nabla\check{\mathbf{x}}\|_\infty$ is bounded since, by taking gradients of (2.5) and (2.6), $\nabla\check{\mathbf{x}}$ solves the *linear* ordinary differential equation in time

$$\begin{aligned}(\nabla\check{\mathbf{x}})_t &= \nabla\mathbf{u}(\check{\mathbf{x}},t)\nabla\check{\mathbf{x}} \quad \text{in } \Omega\times J^n,\\ \nabla\check{\mathbf{x}}^{n+1} &= I \qquad\qquad\qquad \text{in } \Omega.\end{aligned}$$

At time t^{n+1}, $\check{\mathbf{x}}^{n+1}=\mathbf{x}\in e_i$, so by (7.15),

$$s_i^{n+1}(\mathbf{x}) = \frac{\det\left(\mathbf{x}-\mathbf{x}_i,\tilde{\mathbf{e}}_i^{n+1}\right)}{|\tilde{\mathbf{e}}_i^{n+1}|}=0.$$

By the mean value theorem, there exists some $\tau\in J^n$ such that

$$\begin{aligned}\left|s_i^n(\mathbf{x})\right| &= \left|\frac{\partial s_i}{\partial t}(\mathbf{x},\tau)\right|\Delta t^n\\ &= \left|\frac{\det\left(\check{\mathbf{x}}'(\tau)-\check{\mathbf{x}}_i'(\tau),\tilde{\mathbf{e}}_i(\tau)\right)}{|\tilde{\mathbf{e}}_i(\tau)|}+\frac{\det\left(\check{\mathbf{x}}(\tau)-\check{\mathbf{x}}_i(\tau),\tilde{\mathbf{e}}_i'(\tau)\right)}{|\tilde{\mathbf{e}}_i(\tau)|}\right.\\ &\qquad\left.-\det\left(\check{\mathbf{x}}(\tau)-\check{\mathbf{x}}_i(\tau),\tilde{\mathbf{e}}_i(\tau)\right)\frac{\tilde{\mathbf{e}}_i(\tau)\cdot\tilde{\mathbf{e}}_i'(\tau)}{|\tilde{\mathbf{e}}_i(\tau)|^3}\right|\Delta t^n.\end{aligned}$$

Applying inequalities $|\det(\mathbf{x},\mathbf{y})|\le|\mathbf{x}|\,|\mathbf{y}|$, for any $\mathbf{x}$, $\mathbf{y}\in\mathbb{R}^2$, and

$$|\tilde{\mathbf{e}}_i'(\tau)| = |\check{\mathbf{x}}_{i+1}'(\tau)-\check{\mathbf{x}}_i'(\tau)| = |\mathbf{u}(\check{\mathbf{x}}_{i+1}(\tau),\tau)-\mathbf{u}(\check{\mathbf{x}}_i(\tau),\tau)|\le\|\nabla\mathbf{u}\|_\infty|\tilde{\mathbf{e}}_i(\tau)|,$$

we have

$$\begin{aligned}\left|s_i^n(\mathbf{x})\right| &\le \left(|\check{\mathbf{x}}'(\tau)-\check{\mathbf{x}}_i'(\tau)|+2\|\nabla\mathbf{u}\|_\infty|\check{\mathbf{x}}(\tau)-\check{\mathbf{x}}_i(\tau)|\right)\Delta t^n\\ &= \left(|\mathbf{u}(\check{\mathbf{x}}(\tau),\tau)-\mathbf{u}(\check{\mathbf{x}}_i(\tau),\tau)|+2\|\nabla\mathbf{u}\|_\infty|\check{\mathbf{x}}(\tau)-\check{\mathbf{x}}_i(\tau)|\right)\Delta t^n\\ &\le 3\|\nabla\mathbf{u}\|_\infty|\check{\mathbf{x}}(\tau)-\check{\mathbf{x}}_i(\tau)|\Delta t^n\\ &\le 3\|\nabla\mathbf{u}\|_\infty\|\nabla\check{\mathbf{x}}\|_\infty h\Delta t^n.\end{aligned}\tag{7.17}$$

Combining (7.16) and (7.17) gives

$$V_i^{\mathrm{dis}}\le C''h^2\Delta t^n,\tag{7.18}$$

and summing over all edges e_i of ring $\tilde{R}_n(0)$, by Assumption 7.1, we have

$$\left|(\tilde{R}_n(0)\setminus\check{R}_n)\cup(\check{R}_n\setminus\tilde{R}_n(0))\right| = \sum_{i=1}^{N_R}V_i^{\mathrm{dis}}\le N_RC''h^2\Delta t^n\le C'C''h\Delta t^n.\tag{7.19}$$

Combining (7.13), (7.14), (7.19), and (7.6) gives

$$V_n(\alpha)-|\check{R}_n|_\phi\le C'C''h\Delta t^n+\beta_*\Delta t^n\alpha<0 \text{ when } \alpha<-\frac{C'C''}{\beta_*}h,\tag{7.20}$$

$$V_n(\alpha)-|\check{R}_n|_\phi\ge -C'C''h\Delta t^n+\beta_*\Delta t^n\alpha>0 \text{ when } \alpha>\frac{C'C''}{\beta_*}h.\tag{7.21}$$

By the continuity of $V_n(\alpha)-|\check{R}_n|_\phi$, inequalities (7.20) and (7.21) imply that there exists some α^*, where $|\alpha^*|\le Ch$, such that (7.5) holds. ∎

7.3. Individual element adjustment. Finally, we consider the individual element adjustment of the volume correction algorithm. Let E be a grid element in a ring R, and let $\mathbf{x}_{\rm m}$ be the midpoint of an edge $e = \mathbf{x}_{\rm l}\mathbf{x}_{\rm r}$ of E between the inner and outer ring boundaries which requires adjustment. Vertices and midpoints of edges of E are traced back for time Δt^n and are adjusted to a polygon $\tilde{E}_n(\alpha^*)$ (Figure 7.3, left) in the ring adjustment, where α^* is determined by (7.5). The following lemma gives the local construction of the perturbed velocity field near midpoint $\mathbf{x}_{\rm m}$.

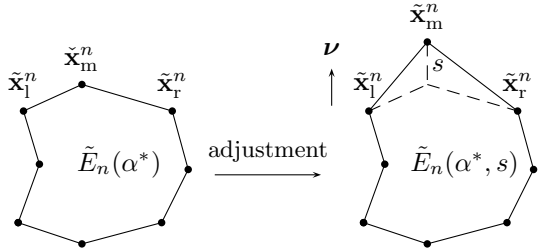

FIG. 7.3. *The trace-back midpoint $\check{\mathbf{x}}^n_{\rm m}$ of element $\tilde{E}_n(\alpha^*)$ is adjusted to $\tilde{\mathbf{x}}^n_{\rm m}$ in the direction of $\boldsymbol{\nu}$.*

LEMMA 7.4. *For $t \in J^n$, let*

$$\tilde{\mathbf{x}}(\mathbf{x}, t) := \check{\mathbf{x}}(\mathbf{x}, t) + s\left(\frac{t^{n+1} - t}{\Delta t^n}\right)\boldsymbol{\nu} \tag{7.22}$$

be a perturbation of the trace-back characteristic $\check{\mathbf{x}}(t)$, so that, in particular, $\tilde{\mathbf{x}}^n_{\rm m} = \check{\mathbf{x}}^n_{\rm m} + s\boldsymbol{\nu}$ is a perturbation of the trace-back midpoint $\check{\mathbf{x}}^n_{\rm m} = \check{\mathbf{x}}(\mathbf{x}_{\rm m}, t^n)$, where $\boldsymbol{\nu}$ is the unit normal vector with respect to the trace-back segment $\tilde{\mathbf{x}}^n_{\rm l}\tilde{\mathbf{x}}^n_{\rm r}$, and $s \in \mathbb{R}$ is the adjustment distance (Figure 7.3, right). Then the perturbed velocity

$$\tilde{\mathbf{u}}(\mathbf{x}, t) := \mathbf{u}\left(\mathbf{x} - s\left(\frac{t^{n+1} - t}{\Delta t^n}\right)\boldsymbol{\nu}, t\right) - \frac{s}{\Delta t^n}\boldsymbol{\nu} \tag{7.23}$$

has $\tilde{\mathbf{x}}$ as its characteristic passing through point $\mathbf{x}$ at time t^{n+1} and

$$\|\mathbf{u} - \tilde{\mathbf{u}}\|_\infty + \|\nabla \cdot \mathbf{u} - \nabla \cdot \tilde{\mathbf{u}}\|_\infty \le C\frac{|s|}{\Delta t^n}, \tag{7.24}$$

where $C > 0$ is a constant independent of h and Δt^n.

Proof. We compute

$$\begin{aligned}\tilde{\mathbf{x}}'(t) &= \check{\mathbf{x}}'(t) - \frac{s}{\Delta t^n}\boldsymbol{\nu} = \mathbf{u}(\check{\mathbf{x}}(t), t) - \frac{s}{\Delta t^n}\boldsymbol{\nu} \\ &= \mathbf{u}\left(\tilde{\mathbf{x}}(t) - s\left(\frac{t^{n+1} - t}{\Delta t^n}\right)\boldsymbol{\nu}, t\right) - \frac{s}{\Delta t^n}\boldsymbol{\nu} = \tilde{\mathbf{u}}(\tilde{\mathbf{x}}(t), t).\end{aligned}$$

Since clearly $\tilde{\mathbf{x}}(t^{n+1}) = \check{\mathbf{x}}(t^{n+1}) = \mathbf{x}$, we have the claimed characteristic curve. Now

$$\begin{aligned}|\mathbf{u}(\mathbf{x}, t) - \tilde{\mathbf{u}}(\mathbf{x}, t)| &= \left|\mathbf{u}(\mathbf{x}, t) - \mathbf{u}\left(\mathbf{x} - s\left(\frac{t^{n+1} - t}{\Delta t^n}\right)\boldsymbol{\nu}, t\right) + \frac{s}{\Delta t^n}\boldsymbol{\nu}\right| \\ &\le \|\nabla \mathbf{u}\|_\infty |s| + \frac{|s|}{\Delta t^n} \le C\frac{|s|}{\Delta t^n},\end{aligned} \tag{7.25}$$

and by the uniform Lipschitz continuity of $\nabla \cdot \mathbf{u}$ in (2.16),

$$|\nabla \cdot \mathbf{u}(\mathbf{x},t) - \nabla \cdot \tilde{\mathbf{u}}(\mathbf{x},t)| = \left|\nabla \cdot \mathbf{u}(\mathbf{x},t) - \nabla \cdot \mathbf{u}\left(\mathbf{x} - s\left(\frac{t^{n+1}-t}{\Delta t^n}\right)\boldsymbol{\nu}, t\right)\right| \le L|s|\left|\frac{t^{n+1}-t}{\Delta t^n}\right| \le L|s|. \tag{7.26}$$

Combining (7.25) and (7.26) gives (7.24). ∎

LEMMA 7.5. *Let $\tilde{E}_n(\alpha^*, s)$ be the trace-back polygonal approximation of $\check{E}_n$ with velocity field $\tilde{\mathbf{u}}$ defined in Lemma 7.4 (Figure 7.3, right), and let $V_{E_n}(\alpha^*, s) := |\tilde{E}_n(\alpha^*, s)|_\phi$ be its pore volume. Assume that no self-intersected polygons are created during the adjustment. If*

$$|\tilde{\mathbf{x}}_{\mathrm{l}}^n - \tilde{\mathbf{x}}_{\mathrm{r}}^n| \ge \lambda_* h \tag{7.27}$$

for some constant $\lambda_* > 0$, then there exists some s^* such that

$$V_{E_n}(\alpha^*, s^*) = |\check{E}_n|_\phi, \tag{7.28}$$

where $|s^*| \le Ch\Delta t^n$ for some constant $C > 0$ independent of n, h, and Δt^n.

Proof. For any s in a neighborhood of zero, consider the difference

$$V_{E_n}(\alpha^*, s) - |\check{E}_n|_\phi = (V_{E_n}(\alpha^*, s) - V_{E_n}(\alpha^*, 0)) + (V_{E_n}(\alpha^*, 0) - V_{E_n}(0,0)) + (V_{E_n}(0,0) - |\check{E}_n|_\phi). \tag{7.29}$$

For the first term on the right-hand side of (7.29), since no self-intersected polygons are created during the adjustment, $E_n(\alpha^*, s)$ is monotone in s, so by (7.27),

$$|V_{E_n}(\alpha^*, s) - V_{E_n}(\alpha^*, 0)| \ge \frac{1}{2}\phi_* |\tilde{\mathbf{x}}_{\mathrm{l}}^n - \tilde{\mathbf{x}}_{\mathrm{r}}^n|\,|s| \ge \frac{1}{2}\phi_* \lambda_* h|s|. \tag{7.30}$$

For the second term on the right-hand side of (7.29), notice that $\tilde{E}_n(\alpha, 0) = \tilde{E}_n(\alpha) \subset \tilde{R}_n(\alpha)$ and the diameter of $\tilde{E}_n(\alpha)$ is $h_{\tilde{E}_n(\alpha)} \le \|\nabla \check{\mathbf{x}}\|_\infty h$, so by (7.9) and Lemma 7.2, we have

$$|V_{E_n}(\alpha^*, 0) - V_{E_n}(0,0)| \le |\tilde{E}_n(\alpha^*) \setminus \tilde{E}_n(0)| + |\tilde{E}_n(0) \setminus \tilde{E}_n(\alpha^*)| \le 2\big(h_{\tilde{E}_n(\alpha^*)} + h_{\tilde{E}_n(0)}\big)\|\mathbf{u}\|_\infty |\alpha^*| \Delta t^n \le C' h^2 \Delta t^n. \tag{7.31}$$

For the third term on the right-hand side of (7.29), since $\tilde{E}(0)$ is an octagon, by (7.18), we have

$$\big|V_{E_n}(0,0) - |\check{E}_n|_\phi\big| = \big||\tilde{E}(0)|_\phi - |\check{E}_n|_\phi\big| \le C' h^2 \Delta t^n. \tag{7.32}$$

Combining (7.29), (7.30), (7.31), and (7.32) gives

$$V_{E_n}(\alpha^*, s) - |\check{E}_n|_\phi \le \frac{1}{2}\phi_* \lambda_* h s + 2C' h^2 \Delta t^n < 0 \text{ when } s < -\frac{4C'}{\phi_* \lambda_*} h \Delta t^n, \tag{7.33}$$

$$V_{E_n}(\alpha^*, s) - |\check{E}_n|_\phi \ge \frac{1}{2}\phi_* \lambda_* h s - 2C' h^2 \Delta t^n > 0 \text{ when } s > \frac{4C'}{\phi_* \lambda_*} h \Delta t^n. \tag{7.34}$$

By the continuity of $V_{E_n}(\alpha^*, s) - |\check{E}_n|_\phi$, inequalities (7.33) and (7.34) imply that there exists some s^*, where $|s^*| \le Ch\Delta t^n$, such that (7.28) holds. ∎

Remark 7.5. If self-intersected polygons are created during the adjustment, one should reduce the distance $|s^*|$ in (7.28) by tracing and adjusting more points on an edge of a grid element. The assumption (7.27) implies that, again, the trace-back procedure should only be performed away from injection wells so that the length of segment $\tilde{\mathbf{x}}_{\rm l}^n \tilde{\mathbf{x}}_{\rm r}^n$ is nondegenerate.

Finally, combining Lemmas 7.1, 7.2, 7.4, and 7.5, we construct a perturbed velocity field $\tilde{\mathbf{u}}$ locally for all trace-back points, and they all have the L^∞-error $\mathcal{O}(h)$ for $\mathbf{u}$ and $\nabla \cdot \mathbf{u}$. Then we can extend $\tilde{\mathbf{u}}$ to the entire domain $\Omega \times J_T$ by interpolating the local definitions of $\tilde{\mathbf{u}}$, and we keep the same bound for the error. In addition, due to the error $(\Delta t)^r$ of the approximately characteristic tracing in Remark 7.2, we obtain (2.18) and Assumption 2.1 holds.

8. Some convergence tests. We consider a quarter of a "five-spot" pattern of wells, which is a rectangular domain $\Omega = (0, 15) \times (0, 20)$ meters with tracer injection and production wells near opposite corners and boundary condition (2.3). We impose a uniform $n \times n$ rectangular grid over Ω and a uniform time step Δt. It is initially clean: $c^0(\mathbf{x}) = 0$. The injector covers one cell near the corner $(0, 0)$ and has a constant rate of $q = 1.2$ m^2/minute, injecting an inert tracer with concentration $c_I = 1$. The cell comprising the producer near the opposite corner $(15, 20)$ has a rate opposite that of the injector. The velocity $\mathbf{u}$ satisfies Darcy's law

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

where k is the permeability, μ is the fluid viscosity, and p is the pressure. We assume that $\mu = 0.01$ poise is constant (i.e., the concentration of the tracer is too small to affect the viscosity of the fluid, which is water). For simplicity, we solve (2.1) with a constant porosity $\phi(\mathbf{x}) \equiv 1$ and a uniform permeability $k(\mathbf{x}) \equiv 10$ millidarcies.

To test the optimal convergence rate with Euler's method for solving characteristics (i.e., $r = 1$ in Theorem 6.1), let $\Delta t = Ch^{2/3}$ and compute the normalized discrete $L^\infty(J_T; L^1(\Omega))$-error

$$E_h := \frac{1}{|\Omega|} \max_{0 \le k \le N} \|c_h^k - c^k\|_1 \tag{8.1}$$

in Theorem 6.1. We approximate (2.2)–(2.4) using VCCMM for the simulation time $T = 1$ hour, and consider the "exact" solution c computed by the higher order Godunov's method [4, 8] on a fine 256×256 grid using the restricted CFL time step $\Delta t_{\mathrm{CFL},256} \approx 0.23$ second.

Table 8.1 shows the error E_{h_n} and the ratio $C_{h_n} := E_{h_n}/h_n^{2/3}$ on grids for six different sizes n. From the results, the sequence of the ratio C_{h_n} shows an upper bound C^* as h_n decreases to zero, so indeed

$$E_{h_n} \le C^* h_n^{2/3},$$

which is consistent with Theorem 6.1, and indicates that VCCMM is convergent and has the optimal convergence rate of at least $\mathcal{O}(h^{2/3})$.

The next test indicates that $\mathcal{O}(h^{2/3})$ is exactly the optimal convergence rate of VCCMM when $\Delta t = Ch^{2/3}$. Consider a constant velocity field $\mathbf{u} \equiv (0.03, 0.04)$ m/second and no source or sink (i.e., $q = 0$). Then the in-flow boundary $\Gamma_{\rm in}$ is the union of the left and bottom edges of Ω. We impose the boundary and initial conditions

$$c(\mathbf{x}, t) = 1 \text{ on } \Gamma_{\rm in} \times J_T \quad \text{and} \quad c^0(\mathbf{x}) = 0 \text{ in } \Omega,$$

TABLE 8.1
Convergence test 1 *for* $\Delta t = Ch^{2/3}$. *The sequence of* $C_{h_n} \le C^*$, *so* $E_{h_n} \le C^* h^{2/3}$.

n	h_n (m)	Δt_n (sec)	E_{h_n}	C_{h_n}
8	3.1250	115.35	0.46870	0.2193
16	1.5625	72.66	0.19137	0.1421
32	0.7813	45.78	0.07837	0.0924
64	0.3906	28.84	0.03507	0.0656
128	0.1953	18.17	0.01767	0.0525
256	0.0977	11.44	0.00882	0.0416

where $T = 500$ seconds. Then the exact solution c is

$$c(\mathbf{x}, t) = \begin{cases} 1 & \text{if } x_1 \le 0.03\,t \text{ or } x_2 \le 0.04\,t, \\ 0 & \text{otherwise,} \end{cases}$$

and at time T, the entire domain Ω is flooded.

Due to the simplicity of $\mathbf{u}$, there is no need for the polygonal approximation and volume adjustment procedures of VCCMM. Table 8.2 shows the error E_{h_n} defined in (8.1) and the ratio $C_{h_n} := E_{h_n}/h_n^{2/3}$ with grids of 7 different sizes n. From the results, the sequence of the ratio C_{h_n} is stable around 0.03 as h_n decreases to zero, so the optimal convergence rate is apparently exactly $\mathcal{O}(h^{2/3})$ as expected from Theorem 6.1.

TABLE 8.2
Convergence test 2 *for* $\Delta t = Ch^{2/3}$. *The sequence of* $C_{h_n} \approx C^* = 0.03$, *so* $E_h \approx C^* h^{2/3}$.

n	h_n (m)	Δt_n (sec)	E_{h_n}	C_{h_n}
8	3.1250	166.67	0.07252	0.0339
16	1.5625	105.00	0.04508	0.0335
32	0.7813	66.14	0.02506	0.0295
64	0.3906	41.67	0.01639	0.0307
128	0.1953	26.25	0.00972	0.0289
256	0.0977	16.54	0.00670	0.0316
512	0.0488	10.42	0.00396	0.0297
1024	0.0244	6.56	0.00262	0.0311

Extensive numerical examples comparing VCCMM with CMM are included in [2]. They show that the CMM exhibits overshoots and introduces many nonphysical local minima and maxima into the solution. In contrast, the VCCMM corrects these problems and gives monotone values of concentration contours.

9. Summary. The main result of this paper is the proof of convergence of the fully conservative, volume corrected characteristics-mixed method for advection problem (2.1)–(2.4) without diffusion. Usually, we take the initial approximation $c_h^0 = P_h c^0$, which leads to an initial error $\|c_h^0 - c^0\|_1 = \mathcal{O}(h)$. The overall error is $\mathcal{O}(h/\sqrt{\Delta t} + h + (\Delta t)^r)$, where r is related to the accuracy of the characteristic tracing itself (see Remark 7.2). In practice, we usually take the ratio $\Delta t/h$ to be a constant so the trace-back elements do not degenerate and self-intersect. Then the convergence rate of the method given by Theorem 6.1 is $\mathcal{O}(\sqrt{h})$. This rate is the same as Godunov's method, but we avoid the CFL constraint which puts an upper bound on the ratio $\Delta t/h$. Therefore, large time steps Δt can be taken. However, as long as we do not introduce self-intersected trace-back regions, we can use much larger time steps. The optimal choice is $\Delta t = Ch^{2/(2r+1)}$, i.e., $\Delta t = Ch^{2/3}$ if $r = 1$, for a convergence rate $\mathcal{O}(h^{2/3})$. This is a better convergence rate than Godunov's method achieves.

In the case of nondegenerate diffusion, in [3] it was shown that the CMM approximates the full advection-diffusion system (1.1)–(1.4) in the L^2-norm to $\mathcal{O}(h + \Delta t)$. Since the volume adjustment produces only a small extra error in the perturbed velocity, the proof extends to VCCMM to obtain a similar L^2-error estimate. Moreover, postprocessing the concentration can improve the convergence to $\mathcal{O}(h^2 + \Delta t)$.

The major difficulty of the proofs is to verify the existence and estimate the error of the locally conservative perturbed velocity field $\tilde{\mathbf{u}}$ in Assumption 2.1. Under some additional assumptions, our results guarantee that the volume correction step only produces a sufficiently small perturbation and therefore maintains the convergence of the method. Actually, in practice, we do not calculate $\tilde{\mathbf{u}}$ or verify Assumptions 7.1–7.4. We just need to verify in the code that there exist α^* and s^*, which satisfy Lemmas 7.2 and 7.5, respectively; i.e., α^* and s^* are not too large ($|\alpha^*| \le Ch$ and $|s^*| \le Ch\Delta t$).

REFERENCES

[1] T. Arbogast, A. Chilakapati, and M.F. Wheeler, *A characteristic-mixed method for contaminant transport and miscible displacement*, in Computational Methods in Water Resources IX, Vol. 1: Numerical Methods in Water Resources, T. F. Russell et al., eds., Computational Mechanics Publications, Southampton, UK, 1992, pp. 77–84.

[2] T. Arbogast and C.-S. Huang, *A fully mass and volume conserving implementation of a characteristic method for transport problems*, SIAM J. Sci. Comput., 28 (2006), pp. 2001–2022.

[3] T. Arbogast and M.F. Wheeler, *A characteristics-mixed finite element method for advection-dominated transport problems*, SIAM J. Numer. Anal., 32 (1995), pp. 404–424.

[4] J.B. Bell, C.N. Dawson, and G.R. Shubin, *An unsplit higher-order Godunov method for scalar conservation laws in two dimensions*, J. Comput. Phys., 74 (1988), pp. 1–24.

[5] M.A. Celia, T.F. Russell, I. Herrera, and R.E. Ewing, *An Eulerian-Lagrangian localized adjoint method for the advection-diffusion equation*, Adv. Water Res., 13 (1990), pp. 187–206.

[6] C.M. Dafermos, *Hyperbolic Conservation Laws in Continuum Physics*, Springer-Verlag, Berlin, 2005.

[7] H.K. Dahle, R.E. Ewing, and T.F. Russell, *Eulerian-Lagrangian localized adjoint methods for a nonlinear advection-diffusion equation*, Comput. Methods Appl. Mech. Engrg., 122 (1995), pp. 223–250.

[8] C.N. Dawson, *Godunov-mixed methods for advection-diffusion equations in multidimensions*, SIAM. J. Numer. Anal., 30 (1993), pp. 1315–1332.

[9] C.N. Dawson, T.F. Russell, and M.F. Wheeler, *Some improved error estimates for the modified method of characteristics*, SIAM J. Numer. Anal., 26 (1989), pp. 1487–1512.

[10] J. Douglas, Jr., C.-S. Huang, and F. Pereira, *The modified method of characteristics with adjusted advection*, Numer. Math., 83 (1999), pp. 353–369.

[11] J. Douglas, Jr., F. Pereira, and L.-M. Yeh, *A locally conservative Eulerian-Lagrangian numerical method and its application to nonlinear transport in porous media*, Comput. Geosci., 4 (2000), pp. 1–40.

[12] J. Douglas, Jr., and T.F. Russell, *Numerical methods for convection-dominated diffusion problems based on combining the method of characteristics with finite element or finite difference procedures*, SIAM J. Numer. Anal., 19 (1982), pp. 871–885.

[13] R.E. Ewing, T.F. Russell, and M.F. Wheeler, *Convergence analysis of an approximation of miscible displacement in porous media by mixed finite elements and a modified method of characteristics*, Comput. Methods Appl. Mech. Engrg., 47 (1984), pp. 73–92.

[14] R. Eymard, T. Gallouët, and R. Herbin, *Finite volume methods*, in Handbook of Numerical Analysis, Vol. 7, P. G. Ciarlet et al., eds., North–Holland, Amsterdam, 2000, pp. 713–1020.

[15] E. Giusti, *Minimal Surfaces and Functions of Bounded Variation*, Birkhäuser Verlag, Basel, 1984.

[16] R.W. Healy and T.F. Russell, *Treatment of internal sources in the finite-volume ELLAM*, in Computational Methods in Water Resources XIII, Vol. 2, L. Bentley et al., eds., A. A. Balkema, Rotterdam, 2000, pp. 619–622.

[17] N.N. Kuznetsov, *Accuracy of some approximate methods for computing the weak solutions of*

a first-order quasilinear equation, USSR Comput. Math. Math. Phys., 16 (1976), pp. 105–119.

[18] R.J. LeVeque, *Numerical Methods for Conservation Laws*, 2nd ed., Birkhäuser Verlag, Basel, 1992.

[19] B.J. Lucier, *Error bounds for the methods of Glimm, Godunov and Leveque*, SIAM J. Numer. Anal., 22 (1985), pp. 1074–1081.

[20] O. Pironneau, *On the transport-diffusion algorithm and its applications to the Navier-Stokes equations*, Numer. Math., 38 (1981/82), pp. 309–332.

[21] P.-A. Raviart and J.M. Thomas, *A mixed finite element method for 2nd order elliptic problems*, in Mathematical Aspects of Finite Element Methods, Lecture Notes in Math. 606, Springer, Berlin, 1977, pp. 292–315.

[22] H. Wang, *An optimal-order error estimate for MMOC and MMOCAA schemes for multidimensional advection-reaction equations*, Numer. Methods Partial Differential Equations, 18 (2002), pp. 69–84.

[23] H. Wang and M. Al-Lawatia, *A locally conservative Eulerian-Lagrangian control-volume method for transient advection-diffusion equations*, Numer. Methods Partial Differential Equations, 22 (2006), pp. 577–599.

[24] H. Wang, D. Liang, R.E. Ewing, S.L. Lyons, and G. Qin, *An ELLAM approximation for highly compressible multicomponent flows in porous media*, Comput. Geosci., 6 (2002), pp. 227–251.

[25] H. Wang, W. Zhao, and R.E. Ewing, *A numerical modeling of multicomponent compressible flows in porous media with multiple wells by an Eulerian-Lagrangian method*, Comput. Vis. Sci., 8 (2005), pp. 69–81.