$L^\infty$-Stability of IMEX-BDF2 Finite Volume Scheme for Convection-Diffusion Equation
Caterina Calgaro, Meriem Ezzoug

To cite this version:
Caterina Calgaro, Meriem Ezzoug. $L^\infty$-Stability of IMEX-BDF2 Finite Volume Scheme for Convection-Diffusion Equation. FVCA 2017: Finite Volumes for Complex Applications VIII - Methods and Theoretical Aspects, Jun 2017, Lille, France. pp.245-253, 10.1007/978-3-319-57397-7_17 . hal-01574893

HAL Id: hal-01574893
https://hal.science/hal-01574893
Submitted on 16 Aug 2017
Abstract  In this paper, we propose a finite volume scheme for solving a two-dimensional convection-diffusion equation on general meshes. This work is based on an implicit-explicit (IMEX) second order method and it is issued from the seminal paper [2]. In the framework of MUSCL methods, we will prove that the local maximum property is guaranteed under an explicit Courant-Friedrichs-Levy condition and the classical hypothesis for the triangulation of the domain.

Key words: Convection-diffusion equation, finite volume scheme, IMEX-BDF2 scheme, $L^\infty$-stability

MSC (2010): 65M99, 76M12, 76E17

1 Introduction

Convection-diffusion processes appear in many areas of science, e.g. fluid dynamics or heat and mass transfer. In the study of the evolution of a mixture, the system of PDEs derives from the compressible Navier-Stokes equations. The mixture of two viscous fluids is described by the density $\rho \geq 0$, the mass velocity field $\mathbf{v}$ (which is not solenoidal) and the pressure $p$. Following Kazhikhov and Smagulov [11], we set

$$
\mathbf{u} = \mathbf{v} + \lambda \nabla \ln(\rho),
$$

(1)

---

Caterina Calgaro
Univ. Lille, CNRS, UMR 8524 - Laboratoire Paul Painlevé,
F-59000 Lille, France
e-mail: caterina.calgaro@univ-lille1.fr

Meriem Ezzoug
UR: Multifractals et Ondelettes. FSM. Université de Monastir.
5019 Monastir, Tunisie
ISSIG. Université de Gabès, 6032 Gabès, Tunisie
e-mail: meriemezzoug@yahoo.fr
for some mass diffusion coefficient $\lambda > 0$. This Fick’s law describes the diffusive fluxes of one fluid into the other. Clearly, the \textit{volume velocity} field $\mathbf{u}$ satisfies $\text{div} \mathbf{u} = 0$ and we obtain the non-standard constraint $\text{div} \mathbf{v} = -\text{div}(\lambda \nabla \ln(p))$, which is relied on the definition of the pressure $p$. Using (1), the mass conservation equation becomes

$$\frac{\partial \rho}{\partial t} + \text{div}(\rho \mathbf{u}) = \lambda \Delta \rho. \quad (2)$$

The momentum equation can also be rewritten in order to obtain the Kazhikhov-Smagulov model [11]. This model was firstly studied in [14, 1] (see also references therein). The mathematical analysis in a three-dimensional domain of Kazhikhov-Smagulov type models was carried out in recent works [5, 8], where the authors study the Kazhikhov-Smagulov models with a specific Korteweg stress tensor. The numerical study of a Kazhikhov-Smagulov model for the two-dimensional case can be found in [4], where the authors propose an hybrid finite volume-finite element method combined with the backward Euler method in time. In order to generalize the analysis given in [4] to second-order methods in time and space, the first goal is to recover the $L^\infty$-stability of the finite volume method used for the convection-diffusion equation. This is the purpose of this paper.

2 Description of the numerical scheme

This section is devoted to the design of a numerical scheme to approximate (2), using the vertex-based MUSCL finite volume methods introduced in [13] and used in [2] for a second-order accuracy in space, and an implicit-explicit (IMEX) linear multistep methods [10] for a second-order in time.

\textbf{Mesh definitions and notations.} Let $\Omega$ be an open bounded polygonal subset on $\mathbb{R}^2$, with sufficiently regular boundary $\partial \Omega$, and $[0, T]$ the time interval, for $T > 0$. The discretization of (2) will be carried out on an unstructured triangular mesh. We denote by $\mathcal{T}_h$ a partition of $\Omega$ composed of conforming and isotropic triangles $T_k$, $k \in [1, K]$, with $K \in \mathbb{N}^*$. The $\mathcal{T}_h$ is called the \textit{primal mesh}. We suppose the following hypotheses:

\begin{enumerate}
\item [(H1)] Let \{ $\mathcal{T}_h$ \}$_{h>0}$ be a regular family of triangulations of $\Omega$.
\item [(H2)] The triangulation $\mathcal{T}_h$ is of weakly acute type (no triangle with an angle greater than $\pi/2$).
\end{enumerate}

For each element $T \in \mathcal{T}_h$, we denote $B_T$ the barycenter of the triangle, $|T|$ the area of $T$, and $M_i, M_j, M_k$ the three vertices of $T$. We also denote respectively $M_{ij}$ and $M_{ik}$ the middles of $[M_i M_j]$ and $[M_i M_k]$.

Let us construct the dual mesh $\mathcal{C}_h = \{ \mathcal{C}_i, i \in [1, I] \}$, which defines a second partition of $\Omega$, ($I \in \mathbb{N}^*$ is the number of vertices of $\mathcal{T}_h$). The dual finite volume $\mathcal{C}_i$ associated with each vertex $M_i, i \in [1, I]$, is a closed polygon obtained in the following way: we join the barycenter $B_T$ of every triangle $T \in \mathcal{T}_h$ which share the vertex $M_i$ with
the middle point of every side of \( T \) containing \( M_i \) (see Fig. 1). If \( M_i \in \partial \Omega \), then we complete the boundary of \( C_i \) by the segments joining \( M_i \) with the middle point of boundary sides that contain \( M_i \), \( C_i \) is often called the vertex-based control volume around the node \( M_i \). Accordingly, we have \( \bigcup_{T \in \mathcal{T}_h} T = \hat{\Omega} = \bigcup_{i \in [1, l]} C_i \).

Moreover, if we denote \( |C_i| \) the area of \( C_i \in \mathcal{G}_h \), then \( |C_i| = \sum_{T,M_i \in T} \frac{|T|}{3} \).

For \( i \in [1, l] \), let \( \mathcal{V}(i) = \{ j \in [1, l], C_j \) is a neighbor of \( C_i \} \). For \( i = 1, 2 \), we denote \( \Gamma_{ij}^{(T)} \) the segment \([M_{ij}B_T], A_{ij}^{(T)} \) its middle point, \( n_{ij}^{(T)} \) the unit outward normal to \( C_i \) along \( \Gamma_{ij}^{(T)} \) and \( |\Gamma_{ij}^{(T)}| \) the length of \( \Gamma_{ij}^{(T)} \). For \( T \in \mathcal{T}_h \) and \( M_i \in T \), we have:

\[
\sum_{i=1}^{2} \frac{|\Gamma_{ij}^{(T)}|}{2} n_{ij}^{(T)} = -|T|\nabla \psi_i, \tag{3}
\]

where \( \psi_i \) is the \( P_1 \) basis function associated to the vertex \( M_i \) of \( T \). For every \( C_i \in \mathcal{G}_h \), the boundary of \( C_i \) is

\[
\partial C_i = \bigcup_{T,M_i \in T} \left( \Gamma_{ij}^{(T)} \cup \Gamma_{ij}^{(T)} \right). \tag{4}
\]

![Fig. 1 Dual mesh-Vertex based control volume \( C_i \) around the node \( M_i \).](image)

**IMEX-BDF2 finite volume scheme.** Here, we describe the finite volume scheme for solving (2). In order to obtain the density reconstruction on the interfaces \( \Gamma_{ij}^{(T)} \), we use the MUSCL technique with a multislope gradient reconstruction. Concerning the time discretization, we adapt the implicit-explicit (IMEX) linear 2-step methods using extrapolated BDF2 scheme for the convective term combined with implicit BDF2 scheme for the diffusive term. The velocity field \( u(t,x) \in \mathbb{R}^2 \) is a given function verifying the divergence free condition. For the space discretization, the usual vertex-based finite volume scheme on control volume \( C_i \), for all \( i \in [1, l] \), reads
\[
\frac{d}{dt} \int_{\mathcal{G}_i} \rho(t, x) \, dx + \int_{\partial \mathcal{G}_i} \rho(t, x) \mathbf{u}(t, x) \cdot \mathbf{n} \sigma = \lambda \int_{\partial \mathcal{G}_i} \nabla \rho(t, x) \cdot \mathbf{n} \sigma. \tag{5}
\]

We denote by \( \Delta t \) the time step and \( t^n = n \Delta t, n \geq 0 \), but variable time steps can also be used. Then, the approximate solution \( \rho^n_i, i \in [1, I] \), at time \( t^n \), verifies
\[
\rho^n_i \approx \frac{1}{\mathcal{V}_i} \int_{\mathcal{V}_i} \rho(t^n, x) \, dx.
\]

In particular, the numerical approximation of the density is a piecewise constant function in space on the control volume \( \mathcal{V}_i \). For the time discretization, we consider the implicit BDF2 scheme and an extrapolated BDF2 scheme following [10]. With this choice, we obtain a second-order accuracy in time. Then, the equation (5) is rewritten as follows, for each \( i \in [1, I] \) and \( n \geq 1 \),
\[
\rho^n_i - \frac{2}{3} \frac{\Delta t}{\mathcal{V}_i} \int_{\partial \mathcal{V}_i} \nabla \rho(t^{n+1}, x) \cdot \mathbf{n} \sigma = \frac{2}{3} \frac{\Delta t}{\mathcal{V}_i} \int_{\partial \mathcal{V}_i} \rho(t^n, x) \mathbf{u}(t^n, x) \cdot \mathbf{n} \sigma
\]
\[
- \frac{1}{3} \rho^n_i - \frac{2}{3} \frac{\Delta t}{\mathcal{V}_i} \int_{\partial \mathcal{V}_i} \rho(t^{n-1}, x) \mathbf{u}(t^{n-1}, x) \cdot \mathbf{n} \sigma.
\tag{6}
\]

In order to approximate \( \nabla \rho(t^{n+1}, x) \) in (6), we consider a \( \mathbb{P}_1 \)-finite element approach for the density such that
\[
\rho^n_{i|T} \approx \sum_{M_j \in T} \psi_j \rho^n_{j|T}, \quad \text{for all } T \in \mathcal{T}_h,
\]
with \( \{ \psi_j \}_{j \in [1, I]} \) the canonical basis of the usual \( \mathbb{P}_1 \) finite element space. Using (3) and (4), we find \( \rho^n_{i|T}, i \in [1, I], n \geq 1 \), verifying the following second-order IMEX-BDF2 finite volume scheme:
\[
\rho^n_{i|T} + \frac{2}{3} \frac{\Delta t}{\mathcal{V}_i} \sum_{T, M_k \in T} |T| \sum_{M_j \in T} \nabla \psi_i \cdot \nabla \psi_j \rho^n_{j|T}
\]
\[
= \frac{4}{3} \rho^n_i - \frac{4}{3} \frac{\Delta t}{\mathcal{V}_i} \sum_{T, M_k \in T} |T| \bigg| \int_{\mathcal{M}_k} G^n_{ij} \left( \rho^n_i, \rho^n_{j|T} \right) \bigg| - \frac{1}{3} \rho^n_i - \frac{2}{3} \frac{\Delta t}{\mathcal{V}_i} \sum_{T, M_k \in T} |T| \bigg| \int_{\mathcal{M}_k} G^n_{ij} \left( \rho^n_{j|T}, \rho^n_{j|T} \right) \bigg|.
\tag{7}
\]

Here we denote by \( G_{ij} \left( \rho_1, \rho_2 \right) \) a numerical flux that satisfies the consistency, conservativity and monotonicity properties. In particular, for any constant function \( \rho_1 \), we have
\[
\sum_{k \in T} |\Gamma_k^{(T)}| \, G_{ik} \left( \rho_1, \rho_1 \right) = 0.
\tag{8}
\]

In [2], \( G_{ik} \) is the upstream flux, but many other numerical fluxes can be considered, as for instance Lax-Friedrichs or Engquist-Osher fluxes. We underline that for multi-
physics coupled models, a particular attention must be paid in the approximation of the continuous velocity associated to any point of \( \partial \mathcal{C}_i \) (see [3]).

In (7), \( \rho_{ijl} \) and \( \rho_{ij} \) denote the density reconstructions on the segments \( I_{ijl}^{(T)} \), for \( l = 1, 2 \). In order to reach a second-order accuracy in space, we use the MUSCL technique [13] with a multislope gradient reconstruction. Introducing \( M = [M_{jl} M_{jl}] \cap (M_i A_{ijl}^{(T)}) \) and \( \hat{N} \in [M_k M_{k2}] \cap (M_i A_{ijl}^{(T)}) \), we define

\[
\begin{align*}
\rho_{ij}^{\text{up}} &= \frac{\rho_i - \rho_{\hat{N}}}{\| M_i N \|}, & \text{and} & \quad \rho_{ij}^{\text{down}} = \frac{\rho_{\hat{M}} - \rho_i}{\| M_i M \|},
\end{align*}
\]

Then, \( \rho_{ijl} \) is the density evaluated at node \( A_{ijl}^{(T)} \), defined as:

\[
\rho_{ijl} = \rho_i + \rho_{ijl} \| M_i A_{ijl}^{(T)} \|, \quad \text{with} \quad \rho_{ijl} = \rho_{ij}^{\text{up}} \text{Lim}(\frac{\rho_{ij}^{\text{down}}}{\rho_{ij}^{\text{up}}}),
\]

where \( \text{Lim} \) is a so-called ”\( \tau \)-limiter” (for details see [2]). In particular, they have the following result:

**Lemma 1.** There exists some coefficients \( \omega_{ijk} \geq 0, k \in \mathcal{T}(i) \), such that

\[
\rho_{ijl} - \rho_i = \sum_{k \in \mathcal{T}(i)} \omega_{ijk} (\rho_i - \rho_k)
\]

holds, and furthermore, they verify \( \sum_{k \in \mathcal{T}(i)} \omega_{ijk} \leq \frac{\tau}{2} C_{\mathcal{B}_h} \), where the constant \( C_{\mathcal{B}_h} \) characterizes the mesh regularity (but it is more general than the classical Ciarlet ratio) and \( \tau > 0 \) is used in the definition of the \( \tau \)-limiter.

### 3 \( L^\infty \)-stability of the numerical scheme

The IMEX-BDF2 finite volume scheme (7) is rewritten as linear system:

\[
A \rho^{n+1} = F^n,
\]

where the matrix \( A \) and the right hand side \( F^n \) are defined as follows:

\[
A_{,i} = 1 + 2 \lambda \Delta t \sum_{T, M \in T} \| \nabla \psi_i \|^2, \quad A_{,j} = 2 \lambda \Delta t \sum_{T, M \neq M_j \in T} \nabla \psi_i \cdot \nabla \psi_j, \quad \forall i, j \in [1, I],
\]

\[
F^n_i = \frac{4}{3} \rho^n_i - \frac{4}{3 \| \mathcal{C}_i \|} \sum_{T, M \in T} \sum_{l=1}^{2} |I_{ijl}^{(T)}| G_{ijl}^n (\rho^n_{ijl}, \rho^n_{jil})
- \frac{2}{3} \rho^{n-1}_i + \frac{3}{3 \| \mathcal{C}_i \|} \sum_{T, M \in T} \sum_{l=1}^{2} |I_{ijl}^{(T)}| G_{ijl}^{n-1} (\rho^{n-1}_{ijl}, \rho^{n-1}_{jil}), \quad \forall i \in [1, I].
\]

Under the hypotheses (H1) and (H2) on the mesh \( \mathcal{B}_h \), the matrix \( A \) is an M-matrix.
Remark 1. The hypothesis (H2), necessary to establish error estimates, is classical for the vertex-centered finite volume scheme [12] or the combined finite volume–finite element scheme [9]. Obviously, the M-matrix property still holds for Delaunay triangulations (see [6], Sect. 3.4).

Now, we prove the following result:

**Proposition 1.** If for any \( n \geq 1 \), we have \( \rho^{n-1} \geq 0 \) and \( \rho^n \geq 0 \), then the right hand side of linear system (9) satisfy \( F^n \geq 0 \) under the CFL condition:

\[
\Delta t \leq \min_{1 \leq i \leq l} \frac{2}{3} \left( \frac{7\tau}{12} C_{\mathcal{G}_h} + 2 \right) \left\| u_{i} \right\| \left( \int_{T} |G^i_{T,1}| + |G^i_{T,2}| \right),
\]

with \( \left\| u_{i} \right\| = \max_{T \in T_F} \left\| u_{T} \right\|_{L^2(\mathbb{R}^2)} \) where \( u_{T} \) is the cell average velocity.

**Proof.** Let \( i \in [1, l] \) and \( n \geq 1 \). Thanks to (8), the \( i \)-th row of (9) is given by:

\[
\left( A \rho^{n+1} \right)_i = \frac{4}{3} \rho^n_i - \frac{4}{3} \Delta t \sum_{T,M \in T} \left( G^i_{ij} \left( \rho^n_{j_1}, \rho^n_{j_2} \right) - G^i_{ij} \left( \rho^n_{j_1}, \rho^n_{j_2} \right) \right) - \frac{1}{3} \rho^n_i + \frac{2}{3} \Delta t \sum_{T,M \in T} \left( G^i_{ij} \left( \rho^{n-1}_{j_1}, \rho^{n-1}_{j_2} \right) - G^i_{ij} \left( \rho^{n-1}_{j_1}, \rho^{n-1}_{j_2} \right) \right),
\]

Let us introduce some definitions and notations dropping the time indices, such that

\[
\Delta \rho_{ij} = \rho_{ij} - \rho_i, \quad \tilde{\Delta} \rho_{ij} = \rho_{ij} - \rho_i, \quad \text{for } l = 1, 2.
\]

Thanks to Lemma 1, there exists for \( l = 1, 2 \), some coefficients \( \omega_{ijk} \geq 0, k \in \mathcal{Y}(i) \), such that

\[
\Delta \rho_{ij} = \sum_{k \in \mathcal{Y}(i)} \omega_{ijk} \left( \rho_i - \rho_k \right), \quad \text{with} \quad \sum_{k \in \mathcal{Y}(i)} \omega_{ijk} \leq \frac{7\tau}{12} C_{\mathcal{G}_h}.
\]

Also, there exists for \( l = 1, 2 \), some coefficients \( \tilde{\omega}_{ijk} \geq 0, k \in \mathcal{Y}(i) \), such that

\[
\tilde{\Delta} \rho_{ij} = \sum_{k \in \mathcal{Y}(i)} \tilde{\omega}_{ijk} \left( \rho_k - \rho_i \right), \quad \text{with} \quad \sum_{k \in \mathcal{Y}(i)} \tilde{\omega}_{ijk} \leq 2.
\]

Next, for \( 0 < \delta_{ij} < 1, l = 1, 2 \), we consider the following quantities:

\[
E_{ij} = \frac{|G^i_{ij}|}{|G^i_{ij}|} \left( \frac{\partial G^i_{ij}}{\partial \rho_1} \left( \rho_i + \delta_{ij} \Delta \rho_{ij}, \rho_i + \delta_{ij} \tilde{\Delta} \rho_{ij} \right) \right), \quad l = 1, 2,
\]

\[
F_{ij} = -\frac{|G^i_{ij}|}{|G^i_{ij}|} \left( \frac{\partial G^i_{ij}}{\partial \rho_2} \left( \rho_i + \delta_{ij} \Delta \rho_{ij}, \rho_i + \delta_{ij} \tilde{\Delta} \rho_{ij} \right) \right), \quad l = 1, 2.
\]

Of course, by monotonicity of the numerical flux, we have \( E_{ij} \geq 0 \) and \( F_{ij} \geq 0 \). Hence, using the mean value theorem, the numerical scheme (11) is rewritten as
follows:

\[
(A\rho^{n+1})_i = \frac{4}{3}\rho^n_i - \frac{4}{3}\Delta t \sum_{T,M\in T} \sum_{k\in Y(i)} \sum_{l=1}^{2} \left( \alpha_{ij}kE^n_{ij} (\rho^n_j - \rho^n_k) - \bar{\alpha}_{ij}kF^n_{ij} (\rho^n_k - \rho^n_l) \right)
- \frac{1}{3}\rho^n_{i-1} + \frac{2}{3}\Delta t \sum_{T,M\in T} \sum_{k\in Y(i)} \sum_{l=1}^{2} \left( \alpha_{ij}kE^{n-1}_{ij} (\rho^{n-1}_j - \rho^{n-1}_k) - \bar{\alpha}_{ij}kF^{n-1}_{ij} (\rho^{n-1}_k - \rho^{n-1}_l) \right).
\]

(12)

Finally, we obtain the following equations for each \( i \in [1, I] \) and for all \( n \geq 1 \):

\[
(A\rho^{n+1})_i = a_{ii} \rho^n_i + b_{ii} \rho^{n-1}_i + \sum_{k \in Y(i)} (a_{ik} \rho^n_k + b_{ik} \rho^{n-1}_k),
\]

(13)

where \( a_{ii}, b_{ii}, a_{ik} \) and \( b_{ik} \) are easily determined from (12). Clearly, we have

\[
a_{ii} + b_{ii} + \sum_{k \in Y(i)} (a_{ik} + b_{ik}) = 1.
\]

(14)

Moreover, by choosing the time step \( \Delta t \) such that for all \( i \in [1, I] \),

\[
\Delta t \leq \left( \sum_{T,M\in T} \sum_{k\in Y(i)} \sum_{l=1}^{2} \left( \frac{4}{3}(\alpha_{ij}kE^n_{ij} + \bar{\alpha}_{ij}kF^n_{ij}) - \frac{2}{3}(\alpha_{ij}kE^{n-1}_{ij} + \bar{\alpha}_{ij}kF^{n-1}_{ij}) \right)^2 \right)^{-1},
\]

(15)

we have

\[
0 \leq a_{ii} + b_{ii} \leq 1 \quad \text{and} \quad 0 \leq a_{ik} + b_{ik} \leq 1.
\]

(16)

Hence, (13), (14) and (16) allow us to conclude that for each \( i \in [1, I] \), \((A\rho^{n+1})_i\) is written as convex combination of \( \rho^n_i, \rho^{n-1}_i, \rho^n_k \) and \( \rho^{n-1}_k, k \in Y(i) \).

Finally, as a consequence of Proposition 1, and recalling that an M-matrix is invertible with positive inverse, we obtain:

**Theorem 1.** Let the velocity field \( u \) divergence free and the initial density \( \rho_0 \) such that \( \rho_0(x) \geq 0 \). Then, under the CFL condition (10) and the hypotheses (H1) and (H2) on the mesh, the linear system (9) is invertible, and

\[
\rho^{n+1} \geq 0, \quad \forall \ n \geq 1.
\]

(17)

**Numerical results.** Here we consider structured meshes on \( \Omega = [-1, 1]^2 \), a stationary rotating velocity field \( u = (x_2, -x_1) \) and a small diffusion coefficient \( \lambda = 10^{-6} \). Setting \( r = \sqrt{(x_1 + 0.5)^2 + x_2^2} \), the discontinuous initial condition is \( \rho_0 = 1000 \) if \( r \leq 0.25 \) and \( \rho_0 = 1 \) if \( r > 0.25 \). The computations are performed for different values of \( h \geq 0.004 \), until \( T = 0.3 \). In Fig. 2 we show the evolution of the density contours (left) and the solution profiles for some horizontal sections (right). We can remark that the maximum principle is well verified using the IMEX-BDF2 scheme, unlike other classical order two schemes, such as Crank-Nicolson Adams-Bashforth or Crank-Nicolson Runge-Kutta. Some other numerical results can be found in [7].
Fig. 2 The evolution of the density contours (left) and the solution profiles (right) for $\lambda = 10^{-6}$.

References