



**Michigan
Technological
University**

Michigan Technological University
Digital Commons @ Michigan Tech

Dissertations, Master's Theses and Master's Reports

2019

HIGH ORDER BOUND-PRESERVING DISCONTINUOUS GALERKIN METHODS AND THEIR APPLICATIONS IN PETROLEUM ENGINEERING

Ziyao Xu

Michigan Technological University, ziyaox@mtu.edu

Copyright 2019 Ziyao Xu

Recommended Citation

Xu, Ziyao, "HIGH ORDER BOUND-PRESERVING DISCONTINUOUS GALERKIN METHODS AND THEIR APPLICATIONS IN PETROLEUM ENGINEERING", Open Access Master's Report, Michigan Technological University, 2019.

<https://digitalcommons.mtu.edu/etdr/844>

Follow this and additional works at: <https://digitalcommons.mtu.edu/etdr>



Part of the [Numerical Analysis and Computation Commons](#)

HIGH ORDER BOUND-PRESERVING DISCONTINUOUS
GALERKIN METHODS AND THEIR APPLICATIONS IN
PETROLEUM ENGINEERING

By

Ziyao Xu

A REPORT

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

In Mathematical Sciences

MICHIGAN TECHNOLOGICAL UNIVERSITY

2019

© 2019 Ziyao Xu

This report has been approved in partial fulfillment of the requirements for the Degree of MASTER OF SCIENCE in Mathematical Sciences.

Department of Mathematical Sciences

Report Advisor: *Dr. Yang Yang*

Committee Member: *Dr. Zhengfu Xu*

Committee Member: *Dr. Fengjing Liu*

Department Chair: *Dr. Mark S. Gockenbach*

Contents

List of figures	vi
List of tables	vii
Preface	viii
Acknowledgments	ix
Abstract	x
1 Introduction	1
2 High-Order bound-preserving discontinuous Galerkin meth-	
ods for compressible miscible displacements in porous media	
on triangular meshes	5
2.1 Introduction	7
2.2 The DG scheme	12
2.3 Second-order bound-preserving scheme	16
2.4 Bound-preserving technique for high-order	
scheme	26

2.4.1	Flux limiter	26
2.4.2	Slope limiter	28
2.4.3	High-order time discretization	31
2.5	Numerical experiments	31
2.6	Concluding remarks	42
3	High-Order bound-preserving discontinuous Galerkin meth-	
	ods for wormhole propagation on triangular meshes	43
3.1	Introduction	44
3.2	Mathematical model	50
3.3	The DG scheme	51
3.4	Second-order bound-preserving schemes	54
3.5	Bound-preserving technique for high-order	
	schemes	68
3.5.1	High-order spatial discretization	69
3.5.2	High-order time discretization	71
3.6	Numerical experiments	72
3.7	Concluding remarks	79
4	Conclusion	83
	References	95
A	Copyright documentations	96
A.1	Copyright documentation of Chapter 2	96

A.2 Copyright documentation of Chapter 3 98

List of Figures

2.1	Two intersection points for the numerical flux in diffusion part on the triangular mesh.	22
2.2	Triangular mesh ($M = 10$)	32
2.3	Example 2.5.2: Numerical approximations of c_1 and c_2	36
2.4	Example 2.5.3: Concentrations of c_1, c_2 and $c_1 + c_2$	38
2.5	Example 2.5.4: Concentrations of c_1, c_2 and c_3 with limiters	40
2.6	Example 2.5.4: Concentrations of c_1, c_2 and c_3 without limiters . .	41
3.1	Distribution of quadrature points for $k = 1$ and $k = 2$	56
3.2	The points chosen to evaluate directional derivative in the diffusion part.	65
3.3	Triangle K_i and its sine	67
3.4	Triangular mesh ($N = 8$)	73
3.5	Example 3.6.2: Evolution of c_f without limiter	76
3.6	Example 3.6.2: Evolution of c_f with limiter.	77
3.7	Example 3.6.2: Evolution of extreme value of c_f and ϕ in Ω	78
3.8	Example 3.6.3: Concentration of acid with time evolution.	80
3.9	Example 3.6.3: Porosity of rock with time evolution.	81

List of Tables

2.1	Example 2.5.1: Accuracy test for c_1 and c_2 with and without bound-preserving technique.	34
3.1	Example 3.6.1: Accuracy test for c_f and ϕ with and without bound-preserving technique.	74

Preface

This report contains published and submitted works completed by the author of this report. The contributions of the author are detailed in the following paragraphs.

In the Chapter 2, the author has collaborated with Nattaporn Chuenjarern¹ and Yang Yang². The main works of author are the construction of the high-order slope limiter, flux limiter and numerical experiments. The work has been published in the Journal of Computational Physics. This work was supported by National Science Foundation DMS-1818467.

In the Chapter 3, the author has collaborated with Yang Yang and Hui Guo³. Most works in this chapter were performed by the author. The work has been completed as an article to submit to the Journal of Computational Physics. This work was supported by National Science Foundation DMS-1818467.

¹Department of Mathematical Sciences, Michigan Technological University, Houghton, MI 49931.

²Department of Mathematical Sciences, Michigan Technological University, Houghton, MI 49931.

³College of Science, China University of Petroleum, Qingdao 266580, China.

Acknowledgments

I would like to express my sincere gratitude to my advisor, Dr. Yang Yang, for his patient guidance and tremendous support in the past two years. He not only has inspired to be majoring in mathematics, but also has helped me to grow as a researcher in this field. I am also grateful to my committee members, Dr. Zhengfu Xu and Dr. Fengjing Liu for their time, efforts, and input for my study and research. I am thankful to the Department chair Dr. Mark Gockenbach for providing me this opportunity of study at Michigan Tech.

Without the love and the support from my friends and my family members, I won't be able to have such a wonderful and productive time at Michigan tech. Many thanks to them.

Abstract

This report contains researches in the theory of high-order bound-preserving (BP) discontinuous Galerkin (DG) method and their applications in petroleum engineering. It contains both theoretical analysis and numerical experiments. The compressible miscible displacements and wormhole propagation problem, arising in petroleum engineering, is used to describe the evolution of the pressure and concentrations of different components of fluid in porous media. The important physical features of concentration and porosity include their boundedness between 0 and 1, as well as the monotone increasing for porosity in wormhole propagation model. How to keep these properties in the simulation is crucial to the robustness of the numerical algorithm. In the first project, we develop high-order bound-preserving discontinuous Galerkin methods for the coupled system of compressible miscible displacements on triangular meshes. We consider the problem with multi-component fluid mixture and the (volumetric) concentration of the j th component, c_j , should be between 0 and 1. The main idea is stated as follows. First, we apply the second-order positivity-preserving techniques to all concentrations c_j 's and enforce $\sum_j c_j = 1$ simultaneously to obtain physically relevant boundedness for every components. Then, based on the second-order BP schemes, we use the second-order numerical fluxes as the lower-order one to combine with high-order numerical fluxes to achieve the high-order accuracy. Finally, since the classical slope limiter cannot be applied to polynomial upper bounds, we introduce a new limiter to our algorithm. Numerical

experiments are given to demonstrate the high-order accuracy and good performance of the numerical technique. In our second project, we propose high-order bound-preserving discontinuous Galerkin methods to keep the boundedness for the porosity and concentration of acid, as well as the monotone increasing for porosity. The main technique is to introduce a new variable r to replace the original acid concentration and use a consistent flux pair to deduce a ghost equation such that the positive-preserving technique can be applied on both original and deduced equations. A high-order slope limiter is used to keep a polynomial upper bound which changes over time for r . Moreover, the high-order accuracy is attained by the flux limiter. Numerical examples are given to demonstrate the high-order accuracy and bound-preserving property of the numerical technique.

Chapter 1

Introduction

The DG methods become increasingly popular due to their good stability, high-order accuracy, and flexibility on h-p adaptivity. The first DG method was introduced in 1973 by Reed and Hill [1] for neutron linear transport. Subsequently, Cockburn et al. developed Runge-Kutta discontinuous Galerkin (RKDG) methods for hyperbolic conservation laws in their papers [2, 3, 4, 5]. In [6], Cockburn and Shu introduced the local discontinuous Galerkin (LDG) method to solve the convection-diffusion equation.

In 2010, the genuinely maximum-principle-satisfying high-order DG schemes were constructed for conservation laws on rectangular meshes in [50] by Zhang and Shu. The basic idea is to take the test function to be 1 in each cell to yield an equation satisfied by the cell average of the target variable r , and prove the desired boundedness of the cell average \bar{r} . Then a slope limiter which do not affect accuracy and mass conservation can be used to modify the variable r to obtain a new one $\tilde{r} = \bar{r} + \theta(r - \bar{r})$ such that \tilde{r} has the physically relevant bounds.

In the case that the variable r only need a lower bound zero, this technique is also called positivity-preserving technique. The physically positivity-preserving and bound-preserving numerical schemes have been actively studied since then. In 2012, this technique has been successfully extended to triangular meshes in [53], where the general criteria for quadrature rule on triangular elements was proposed. After that, this technique was applied to many problems including compressible Euler equations with source terms [52], hyperbolic equations involving δ -singularities [44, 70], relativistic hydrodynamics [29], extended MHD equations [55], shallow water equations [37], etc. For convection-diffusion equations, the genuinely second-order maximum-principle-preserving technique were introduced in [54]. Subsequently, the extension to third-order or even higher order bound-preserving techniques for parabolic equations were also developed in [69, 10, 58]. Besides the above, the flux limiter [39, 38, 25] can also be used to obtain the high-order accuracy and maintain the boundedness. Recently, in [22], the authors studied miscible displacements in porous media and applied the techniques introduced in [54, 25, 38, 39] to preserve the two bounds, 0 and 1, of the volumetric fractions.

In the Chapter 2, we extend the ideas in [38, 50] and construct high-order bound-preserving DG methods for multi-component compressible miscible displacements. However, there are significant differences from previous techniques. First of all, most of the problems in [38, 50] satisfy maximum-principles while the concentration c_j in (2.1.2) does not. To solve this problem, we would like to apply the positivity-preserving technique to each c_j and enforce $\sum_j c_j = 1$. Secondly,

the high-order positivity-preserving technique in this paper is based on the flux limiter [25, 38]. The basic idea is to combine higher order and lower order fluxes to construct a new one which yield positive numerical cell averages. However, for triangular meshes, first-order fluxes are not easy to construct. Therefore, we will consider the second-order flux as the lower order one. Finally, to obtain the equation satisfied by the cell averages, we need to numerically approximate $r_j = \phi c_j$ instead of c_j . By doing so, the upper bound of r_j is not a constant and the limiter (2.1.4) may fail to work, since such a θ may not exist [22]. Moreover, the limiter applied in [22] is not straightforward extendable to multi-component problems, since we cannot simply set the upper bound of c_j to be 1 if the fluid mixture contains more than two components. Therefore, a new bound-preserving limiter will be introduced. In summary, the whole algorithm can be separated into three parts. We first treat p_t as another source in (2.1.2) to obtain the positivity of c_j by the flux limiter [25, 38]. Then we choose consistent fluxes (see Definition 2.2.1) with suitable parameter in the flux limiter in the concentration and pressure equations to obtain the positivity of $1 - \sum_{j=1}^{N-1} c_j$. More precisely, in our analysis, instead of solving p and c_j , $j = 1, \dots, N-1$, we rewrite (2.1.1) and (2.1.2) into a system of c_j , $j = 1, \dots, N$ and enforce $\sum_{i=1}^N c_i = 1$ by choosing consistent fluxes. Finally, we will introduce a new limiter to obtain physically relevant numerical approximations.

In the Chapter 3, the whole algorithm can be separated into four parts. We first apply positivity-preserving technique to obtain positive ϕ_t and use which as another source to find the velocity and pressure. Then apply the positivity-

preserving technique again to ϕ and c_f simultaneously to obtain positive numerical cell averages by the flux limiter [38, 25]. Subsequently, we choose consistent flux pair [22, 57] with suitable parameters in the flux limiter in the concentration and pressure equations to obtain the positivity of $1 - c_f$. Finally, we introduce suitable limiters to obtain physically relevant numerical approximations.

The main accomplished work will be presented in the following two chapters. In Chapter 2, we show our research on high-order bound-preserving discontinuous Galerkin methods for compressible miscible displacements in porous media on triangular meshes. Then, in Chapter 3, we present our study on high-order bound-preserving discontinuous Galerkin methods for wormhole propagation on triangular meshes. We will end in Chapter 4 with a brief conclusion.

Chapter 2

High-order bound-preserving discontinuous Galerkin methods for compressible miscible displacements in porous media on triangular meshes¹

Abstract

In this paper, we develop high-order bound-preserving (BP) discontinuous Galerkin (DG) methods for the coupled system of compressible miscible displacements on

¹This chapter has been published as an article in Journal of Computational Physics. Citation: N. Chuenjarern, Z. Xu, Y. Yang, Journal of Computational Physics 378 (2019),110-128. <https://doi.org/10.1016/j.jcp.2018.11.003>

triangular meshes. We consider the problem with multi-component fluid mixture and the (volumetric) concentration of the j th component, c_j , should be between 0 and 1. There are three main difficulties. Firstly, c_j does not satisfy a maximum-principle. Therefore, the numerical techniques introduced in (X. Zhang and C.-W. Shu, Journal of Computational Physics, 229 (2010), 3091-3120) cannot be applied directly. The main idea is to apply the positivity-preserving techniques to all c_j 's and enforce $\sum_j c_j = 1$ simultaneously to obtain physically relevant approximations. By doing so, we have to treat the time derivative of the pressure dp/dt as a source in the concentration equation and choose suitable fluxes in the pressure and concentration equations. Secondly, it is not easy to construct first-order numerical fluxes for interior penalty DG methods on triangular meshes. One of the key points in the high-order BP technique applied in this paper is the combination of high-order and lower-order numerical fluxes. We will construct second-order BP schemes and use the second-order numerical fluxes as the lower-order one. Finally, the classical slope limiter cannot be applied to c_j . To construct the BP technique, we will not approximate c_j directly. Therefore, a new limiter will be introduced. Numerical experiments will be given to demonstrate the high-order accuracy and good performance of the numerical technique.

Key words: compressible miscible displacements, bound-preserving, high-order, discontinuous Galerkin method, triangular meshes, multi-component fluid, flux limiter

2.1 Introduction

In this paper, we are interested in constructing high-order bound-preserving discontinuous Galerkin (DG) schemes for compressible miscible displacements in porous media on triangular meshes. We consider the fluid mixture with N components and the governing equations over the computational domain $\Omega = [0, 1] \times [0, 1]$ read

$$d(\mathbf{c})\frac{\partial p}{\partial t} + \nabla \cdot \mathbf{u} = d(\mathbf{c})\frac{\partial p}{\partial t} - \nabla \cdot \left(\frac{\kappa(x, y)}{\mu(\mathbf{c})} \nabla p \right) = q, \quad (x, y) \in \Omega, \quad 0 < t \leq T, \quad (2.1.1)$$

$$\phi \frac{\partial c_j}{\partial t} + \nabla \cdot (\mathbf{u} \cdot c_j) - \nabla \cdot (\mathbf{D} \nabla c_j) = \tilde{c}_j q - \phi c_j z_j p_t, \quad (x, y) \in \Omega, \quad 0 < t \leq T, \quad j = 1, \dots, N-1, \quad (2.1.2)$$

where the dependent variables are the pressure in fluid mixture denoted by p , the Darcy velocity of the mixture (volume flowing across a unit cross-section per unit time) denoted by \mathbf{u} and the concentration of interested species measured in amount of species per unit volume denoted by $\mathbf{c} = (c_1, \dots, c_N)^T$, with c_j being the concentration of the j th component. ϕ and κ are the porosity and permeability of the rock, respectively. μ refer to the concentration-dependent viscosity. q is the external volumetric flow rate, and \tilde{c}_j is the concentration of the fluid in the external flow. \tilde{c}_j must be specified at points where injection ($q > 0$) takes place, and is assumed to be equal to c_j at production points ($q < 0$). The diffusion coefficient \mathbf{D} is symmetric and arises from two aspects: molecular diffusion, which is rather small for field-scale problems, and dispersion, which is velocity-dependent, in the petroleum engineering literature. Its form is

$$\mathbf{D} = \phi(x, y)(d_{\text{mol}}\mathbf{I} + d_{\text{long}}|\mathbf{u}|\mathbf{E} + d_{\text{tran}}|\mathbf{u}|\mathbf{E}^\perp), \quad (2.1.3)$$

where \mathbf{E} , a 2×2 matrix, represents the orthogonal projection along the velocity vector given as

$$\mathbf{E} = (e_{ij}(\mathbf{u})) = \left(\frac{u_i u_j}{|\mathbf{u}|^2} \right), \quad \mathbf{u} = (u_1, u_2),$$

and $\mathbf{E}^\perp = \mathbf{I} - \mathbf{E}$ is the orthogonal complement. The diffusion coefficient d_{long} measures the dispersion in the direction of the flow and d_{tran} shows that transverse to the flow. To ensure the stability of the scheme, \mathbf{D} is assumed to be strictly positive definite in almost all of the previous works. In this paper, we assume \mathbf{D} to be positive semidefinite. Moreover, the pressure is uniquely determined up to a constant, thus we assume $\int_\Omega p \, dx dy = 0$ at $t = 0$. However, this assumption is not essential. Other coefficients can be stated as follows:

$$c_N = 1 - \sum_{j=1}^{N-1} c_j, \quad d(\mathbf{c}) = \phi \sum_{j=1}^N z_j c_j,$$

where z_j is the compressibility factor of the j th component of the fluid mixture. In this paper, we consider homogeneous Neumann boundary conditions

$$\mathbf{u} \cdot \mathbf{n} = 0, \quad (\mathbf{D}\nabla c - c\mathbf{u}) \cdot \mathbf{n} = 0,$$

where \mathbf{n} is the unit outer normal of the boundary $\partial\Omega$. Moreover, the initial solutions are given as

$$c_j(x, y, 0) = c_{j0}(x, y), \quad p(x, y, 0) = p_0(x, y), \quad (x, y) \in \Omega.$$

The miscible displacements in porous media were first presented in [15, 16], where mixed finite element methods were applied. Later, the compressible problem was studied in [17] and the optimal order estimates in L^2 -norm and almost

optimal order estimates in L^∞ -norm were given in [11]. Subsequently, many new numerical methods were introduced, such as the finite difference method [47, 48, 49], characteristic finite element method [27], splitting positive definite mixed element method [40] and H1-Galerkin mixed method [9]. Besides the above, in [35], an accurate and efficient simulator was developed for problems with wells. Later, the authors introduced an Eulerian-Lagrangian localized adjoint method to solve the transport partial differential equation for concentration, while a mixed finite element method to solve the pressure equation [34]. Recently, DG methods have been popular to solve compressible miscible displacements in porous media [13, 14, 41, 42, 23, 43, 46]. Some special numerical techniques were introduced to control the jumps of numerical approximations as well as the nonlinearity of the convection term. Besides the above, there were also significant works discussing the DG methods for incompressible miscible displacements, see e.g. [7, 24, 26, 28, 31, 32, 36] and for general porous media flow, see e.g. [8, 19, 18, 33] and the references therein. However, no previous works above focused on the bound-preserving techniques. In many numerical simulations, the approximations of c_j can be placed out of the interval $[0, 1]$. Especially for problems with large gradients, the value of $d(\mathbf{c})$ might be negative, leading to ill-posedness of the problem, and the numerical approximations will blow up. We will use numerical experiments to demonstrate this point in Section 2.5. In [22], we have introduced second-order bound-preserving DG methods on rectangular meshes for two-component miscible displacements in porous media. In this paper, we will extend the idea to multi-component miscible displacements and

construct high-order bound-preserving techniques on triangular meshes. Moreover, the idea can be extended to incompressible flows with some minor changes.

The DG method gained even greater popularity for good stability, high-order accuracy, and flexibility on h-p adaptivity and on complex geometry. In 2010, the genuinely maximum-principle-satisfying high-order DG and finite volume schemes were constructed in [50] by Zhang and Shu, the extension to unstructured meshes was given in [53]. After that, the idea was applied to many problems such as compressible Euler equations [51, 52], hyperbolic equations involving δ -singularities [44, 45, 55], relativistic hydrodynamics [29] and shallow water equations [37], etc. The basic idea is to take the test function to be 1 in each cell to obtain an equation of the numerical cell average of the target variable, say r , and prove the cell average, \bar{r} , is within the desired bounds. Then we can apply a slope limiter to the numerical approximation and construct a new one

$$\tilde{r} = \bar{r} + \theta(r - \bar{r}), \quad \theta \in [0, 1]. \quad (2.1.4)$$

If the problem has only one lower bound zero, the technique is also called positivity-preserving technique. Thanks to the limiter, the whole algorithm were proved to be L^1 -stable [45, 29] for some complicated systems. Moreover, the technique does not rely on the trouble cell detector and the limiter keeps the high-order accuracy in regions with smooth solutions for scalar equations [50]. In case of convection-diffusion equations, the same idea was applied to construct genuinely second-order maximum-principle-satisfying DG method on unstructured meshes [54]. Recently, the flux limiter [25, 38, 39] and third-order maximum-principle-preserving direct DG method [10] were also introduced. However, it is

not easy to apply the flux limiter to unstructured meshes since the lower order fluxes are not easy to construct, and the only work available is [12] in which the technique for hyperbolic equations was analyzed, and no previous works aimed to discuss convection-diffusion equations. In this paper, we will extend the ideas in [38, 50] and construct high-order bound-preserving DG methods for multi-component compressible miscible displacements. However, there are significant differences from previous techniques. First of all, most of the problems in [38, 50] satisfy maximum-principles while the concentration c_j in (2.1.2) does not. To solve this problem, we would like to apply the positivity-preserving technique to each c_j and enforce $\sum_j c_j = 1$. Secondly, the high-order positivity-preserving technique in this paper is based on the flux limiter [25, 38]. The basic idea is to combine higher order and lower order fluxes to construct a new one which yield positive numerical cell averages. However, for triangular meshes, first-order fluxes are not easy to construct. Therefore, we will consider the second-order flux as the lower order one. Finally, to obtain the equation satisfied by the cell averages, we need to numerically approximate $r_j = \phi c_j$ instead of c_j . By doing so, the upper bound of r_j is not a constant and the limiter (2.1.4) may fail to work, since such a θ may not exist [22]. Moreover, the limiter applied in [22] is not straightforward extendable to multi-component problems, since we cannot simply set the upper bound of c_j to be 1 if the fluid mixture contains more than two components. Therefore, a new bound-preserving limiter will be introduced. In summary, the whole algorithm can be separated into three parts. We first treat p_t as another source in (2.1.2) to obtain the positivity of c_j by the flux

limiter [25, 38]. Then we choose consistent fluxes (see Definition 2.2.1) with suitable parameter in the flux limiter in the concentration and pressure equations to obtain the positivity of $1 - \sum_{j=1}^{N-1} c_j$. More precisely, in our analysis, instead of solving p and c_j , $j = 1, \dots, N-1$, we rewrite (2.1.1) and (2.1.2) into a system of c_j , $j = 1, \dots, N$ and enforce $\sum_{i=j}^N c_j = 1$ by choosing consistent fluxes. Finally, we will introduce a new limiter to obtain physically relevant numerical approximations.

The paper is organized as follows: we first discuss the DG scheme in two dimension on triangular mesh in Section 2.2. In Section 2.3, we demonstrate the bound-preserving technique for second-order scheme. The high-order bound-preserving technique with flux limiter will be given in Section 2.4. In Section 2.5, some numerical experiments and results will be shown. We will end in Section 2.6 with concluding remarks.

2.2 The DG scheme

In this section, we will construct the DG scheme for compressible miscible displacements in porous media. We first demonstrate the notations to be used throughout the paper. We consider triangular meshes and denote Ω_h to be the set of cells. For any $K \in \Omega_h$, we denote the three edges of K to be e_K^i ($i = 1, 2, 3$), with corresponding lengths ℓ_K^i ($i = 1, 2, 3$) and unit outer normal vectors ν_i ($i = 1, 2, 3$). We also denote the neighboring triangle along e_K^i as K_i . We use Γ for all the cell interfaces, and $\Gamma_0 = \Gamma \setminus \partial\Omega$ for all the interior ones. For any $e \in \Gamma$, denote $|e|$ to be the length of e . Let u^\pm denote the numerical solution

on the edges, evaluated from K or K_i . The ' \pm ' for each edge e_K^i is determined by the inner product of $\boldsymbol{\nu}_i$ and a predetermined constant vector $\boldsymbol{\nu}_0$ which is not parallel to any edge in the mesh: for each edge e_K^i in the cell K ,

$$\mathbf{u}^- = \mathbf{u}_K, \quad \mathbf{u}^+ = \mathbf{u}_{K_i}, \quad \text{if } \boldsymbol{\nu}_0 \cdot \boldsymbol{\nu}_i > 0,$$

$$\mathbf{u}^+ = \mathbf{u}_K, \quad \mathbf{u}^- = \mathbf{u}_{K_i}, \quad \text{if } \boldsymbol{\nu}_0 \cdot \boldsymbol{\nu}_i < 0.$$

Moreover, we define \mathbf{n}_e as the unit outer normal of each edge $e \in \Gamma_0$ such that $\mathbf{n}_e \cdot \boldsymbol{\nu}_0 > 0$ and define the jump and average of any function v at the cell interface e as

$$[v]_e = v_e^+ - v_e^-, \quad \{v\}_e = \frac{1}{2}(v_e^+ + v_e^-).$$

We also denote $\partial\Omega_+ = \{e \in \partial\Omega : \mathbf{n} \cdot \boldsymbol{\nu}_0 > 0\}$, where \mathbf{n} is the unit outer normal of $\partial\Omega$ and $\partial\Omega_- = \partial\Omega \setminus \partial\Omega_+$. The finite element space is chosen as

$$W_h = \{z : z|_K \in P^k(K), \quad \forall K \in \Omega_h\},$$

where $P^k(K)$ denotes polynomials of degree at most $k \geq 1$ in K .

To construct the DG method, we first rewrite the system (2.1.1)-(2.1.2) into the following form

$$d(\mathbf{c})p_t + \nabla \cdot \mathbf{u} = q, \tag{2.2.5}$$

$$a(\mathbf{c})\mathbf{u} = -\nabla p, \tag{2.2.6}$$

$$(\phi c_j)_t + \nabla \cdot (\mathbf{u}c_j) - \nabla \cdot (\mathbf{D}(\mathbf{u})\nabla c_j) = \tilde{c}_j q - \phi c_j z_j p_t, \quad j = 1, 2, \dots, N-1, \tag{2.2.7}$$

where $a(\mathbf{c}) = \frac{\mu(\mathbf{c})}{\kappa}$.

Next, we would like to demonstrate the key points in this paper that are quite different from most of the previous works.

1. Approximate $r_j = \phi c_j$ instead of c_j . We cannot simply take the test function to be 1 to obtain the cell average of c_j .
2. Treat p_t in (2.2.7) as a source to apply the positivity-preserving techniques.
3. Apply flux limiters to the high-order scheme by combining the second- and high-order fluxes.
4. Suitably choose the parameters in the flux limiter to obtain consistent fluxes for (2.2.5) and (2.2.7) to make $\bar{r}_j < \bar{\phi}$, where \bar{r}_j and $\bar{\phi}$ are the cell averages of r_j and ϕ , respectively.
5. Take the L^2 -projection of ϕ into W_h , denoted as Φ , and use which as the new approximation of the porosity.
6. Construct a new limiter to maintain the cell average \bar{r}_j and modify the numerical approximations of r_j such that $0 < r_j < \Phi$, which further yields $c_j = P_k \left\{ \frac{r_j}{\Phi} \right\} \in [0, 1]$, where P_k is the L^2 -projection projection into W_h is $k \geq 2$ while $P_1 u|_K$ is the interpolation of u at the three vertices of cell K .

For simplicity, if not otherwise stated, we use $p, \mathbf{u}, c_j, r_j, j = 1, 2, \dots, N$ as the numerical approximations from now on. Then the DG scheme for (2.2.5) - (2.2.7) is to find $p, r_j \in W_h$ and $\mathbf{u} \in \mathbf{W}_h = W_h \times W_h$ such that for any $\xi, \zeta \in W_h$

and $\boldsymbol{\eta} \in \mathbf{W}_h$,

$$(\tilde{d}(\mathbf{r})p_t, \xi) = (\mathbf{u}, \nabla \xi) + \sum_{e \in \Gamma_0} \int_e \hat{\mathbf{u}} \cdot \mathbf{n}_e [\xi] ds + (q, \xi), \quad (2.2.8)$$

$$(a(\mathbf{c})\mathbf{u}, \boldsymbol{\eta}) = (p, \nabla \cdot \boldsymbol{\eta}) + \sum_{e \in \Gamma} \int_e \hat{p} [\boldsymbol{\eta} \cdot \mathbf{n}_e] ds, \quad (2.2.9)$$

$$\begin{aligned} (r_{jt}, \zeta) &= (\mathbf{u}c_j - \mathbf{D}(\mathbf{u})\nabla c_j, \nabla \zeta) + (\check{c}_j q - r_j z_j p_t, \zeta) + \sum_{e \in \Gamma_0} \int_e \widehat{\mathbf{u}c_j} \cdot \mathbf{n}_e [\zeta] ds \\ &\quad - \sum_{e \in \Gamma_0} \int_e \left(\{ \mathbf{D}(\mathbf{u})\nabla c_j \cdot \mathbf{n}_e \} [\zeta] + \{ \mathbf{D}(\mathbf{u})\nabla \zeta \cdot \mathbf{n}_e \} [c_j] + \frac{\tilde{\alpha}}{|e|} [c_j] [\zeta] \right) ds, \end{aligned} \quad (2.2.10)$$

where

$$c_j = P_k \left\{ \frac{r_j}{\Phi} \right\}, \quad \tilde{d}(\mathbf{r}) = \sum_{j=1}^N z_j r_j, \quad (u, v) = \int_K uv dx, \quad \check{c}_j = \begin{cases} \tilde{c}_j, & q > 0, \\ \frac{r_j}{\Phi}, & q < 0. \end{cases}$$

In (2.2.8)-(2.2.10), \hat{p} , $\hat{\mathbf{u}}$ and $\widehat{\mathbf{u}c_j}$ are the numerical fluxes. We use alternating fluxes for the diffusion term and for any $e \in \Gamma_0$

$$\hat{\mathbf{u}}|_e = \mathbf{u}^+|_e, \quad \hat{p}|_e = p^-|_e, \quad (2.2.11)$$

and on $\partial\Omega$ we take

$$\hat{p}|_e = p^-|_e, \quad \forall e \in \partial\Omega^+, \quad \hat{p}|_e = p^+|_e, \quad \forall e \in \partial\Omega^-.$$

For the convection term, for any $e \in \Gamma_0$ we take

$$\widehat{\mathbf{u}c_j} = \mathbf{u}^+ c_j^+ - \alpha [c_j] \mathbf{n}_e. \quad (2.2.12)$$

In (2.2.10) and (2.2.12), α and $\tilde{\alpha}$ are two positive constants to be chosen by the bound-preserving technique. Before we complete this subsection, we would like

to introduce the following definition that will be used in the bound-preserving technique.

Definition 2.2.1. *We say the flux $\widehat{\mathbf{u}}c_j$ is consistent with $\hat{\mathbf{u}}$ if $\widehat{\mathbf{u}}c_j = \hat{\mathbf{u}}$ by taking $c_j = 1$ in Ω .*

The numerical flux $\widehat{\mathbf{u}}c_j$ in (2.2.12) is consistent with the flux $\hat{\mathbf{u}}$ in (2.2.11), and this is required by the bound-preserving technique.

Remark 2.2.1. *There are plenty of fluxes can be used following the procedures introduced in the next section. The proofs are basically the same with some minor changes, so we only list some of them below without more details.*

- $\hat{\mathbf{u}} = \mathbf{u}^-$, $\hat{p} = p^+$, $\widehat{\mathbf{u}}c_j = \mathbf{u}^-c_j^- - \alpha[c_j]\mathbf{n}_e$.
- $\hat{\mathbf{u}} = \frac{1}{2}(\mathbf{u}^+ + \mathbf{u}^-)$, $\hat{p} = \frac{1}{2}(p^+ + p^-)$, $\widehat{\mathbf{u}}c_j = \frac{1}{2}(\mathbf{u}^+c_j^+ + \mathbf{u}^-c_j^-) - \alpha[c_j]\mathbf{n}_e$.

2.3 Second-order bound-preserving scheme

In this section, we will construct second-order bound-preserving DG scheme with Euler forward time discretization on triangular meshes. For simplicity, we only discuss the technique for cells away from $\partial\Omega$, while the boundary cells can be analyzed following the same lines with some minor changes. A similar analysis for the boundary cells can be found in [22]. We use o_K for the numerical approximation of o in K with cell average \bar{o}_K . Moreover, we use o^n as the solution o at time level n . Now, we will demonstrate the bound-preserving technique in

detail. For simplicity, we will drop the subindex j in (2.2.10) and use r, c, \check{c}, z for $r_j, c_j, \check{c}_j, z_j$, respectively.

In (2.2.10), we take $\zeta = 1$ in K to obtain the equation satisfied by the cell average of r

$$\bar{r}_K^{n+1} = H_K^c(r, \mathbf{u}, c) + H_K^d(r, \mathbf{u}, c) + H_K^s(r, \check{c}, q, z, p) \quad (2.3.13)$$

where

$$H_K^c(r, \mathbf{u}, c) = \frac{1}{3} \bar{r}_K^n - \lambda \sum_{i=1}^3 \int_{e_K^i} \widehat{\mathbf{u}} \mathbf{c} \cdot \boldsymbol{\nu}_i ds, \quad (2.3.14)$$

$$H_K^d(r, \mathbf{u}, c) = \frac{1}{3} \bar{r}_K^n + \lambda \sum_{i=1}^3 \int_{e_K^i} \left(\{ \mathbf{D}(\mathbf{u}) \nabla c \cdot \boldsymbol{\nu}_i \} + \frac{\tilde{\alpha}}{\ell_K^i} [c] \mathbf{n}_e \cdot \boldsymbol{\nu}_i \right) ds, \quad (2.3.15)$$

$$H_K^s(r, \check{c}, q, z, p) = \frac{1}{3} \bar{r}_K^n + \Delta t \overline{\check{c}q - rzp}_t, \quad (2.3.16)$$

with $\lambda = \frac{\Delta t}{|K|}$ being the ratio of the time step and the area of triangle K , and $\overline{\check{c}q - rzp}_t$ being the cell average of $\check{c}q - rzp_t$. We denote $V_i, i = 1, 2, 3$ as the three vertices of cell K . In this section, we will construct the bound-preserving technique in K , hence for any $w \in W_h$, we define $w(V_i)$ to be the limit evaluated in K . We use the $(k+1)$ -point Gaussian quadrature to approximate the integrals along the cell interfaces in (2.3.14)-(2.3.16), and denote $x_{i,\beta}, \beta = 1, 2, \dots, k+1$ as the quadrature points on e_K^i with w_β as the corresponding weights on the reference interval $[-\frac{1}{2}, \frac{1}{2}]$. Moreover, we use quadratures discussed in [53] to compute the cell average \bar{r}_K^n . The quadrature contains $L = 3(N_G - 2)(k + 1)$ quadrature points, denoted as x_γ , lying in the interior of K with $2N_G - 3 \geq k$, and the quadratures points on the cell interfaces are exactly the $k + 1$ Gaussian

quadratures points. We denote the quadrature weights corresponding to the interior quadrature points as \tilde{w}_γ and those on the cell interfaces as \hat{w}_β . In [53], it was shown that $\hat{w}_\beta = \frac{2}{3}w_\beta\hat{w}$, where \hat{w} is the quadrature weight corresponding to the first quadrature point in the N_G -point Gauss-Lobatto quadrature on the interval $[-\frac{1}{2}, \frac{1}{2}]$. Based on the above notations, we define the values of o ($o = r, c, p, q, \Phi$) at the quadrature points as $o_K^{i,\beta} = o(x_{i,\beta})$ along the boundary of K and $o_K^\gamma = o(x_\gamma)$ in cell K . Now, we can demonstrate the bound-preserving techniques. We will consider the source term H_K^s first, and discuss the high-order bound-preserving technique.

Lemma 2.3.1. *Suppose $r^n > 0$ ($c^n > 0$), then $H_K^s(r, \check{c}, q, z, p) > 0$ under the conditions*

$$\Delta t \leq \frac{1}{6z p_M}, \quad \Delta t \leq \frac{\Phi_m}{6q_M}, \quad (2.3.17)$$

where

$$p_M = \max_{i,\beta,\gamma}((p_t)_K^{i,\beta}, (p_t)_K^\gamma, 0) \quad \Phi_m = \min_x \Phi(x), \quad q_M = \max_{i,\beta,\gamma} \left\{ -q_K^{i,\beta}, -q_K^\gamma, 0 \right\}. \quad (2.3.18)$$

Proof. We can write H_K^s as

$$H_K^s(r, \check{c}, q, z, p) = \left(\frac{1}{6} \bar{r}_K^n - \Delta t \bar{r} z p_t \right) + \left(\frac{1}{6} \bar{r}_K^n + \Delta t \check{c} q \right) := L_1 + L_2.$$

Applying the quadrature in [53], we have

$$\begin{aligned}
L_1 &= \frac{1}{6} \bar{r}_K^n - \Delta t \overline{r z p_t} \\
&= \frac{1}{6} \left(\sum_{i=1}^3 \sum_{\beta=1}^{k+1} \hat{w}_\beta r_K^{i,\beta} + \sum_{\gamma=1}^L \tilde{w}_\gamma r_K^\gamma \right) \\
&\quad - \Delta t z \left(\sum_{i=1}^3 \sum_{\beta=1}^{k+1} \hat{w}_\beta r_K^{i,\beta}(p_t)_{i,\beta} + \sum_{\gamma=1}^L \tilde{w}_\gamma r_K^\gamma(p_t)_\gamma \right) \\
&= \sum_{i=1}^3 \sum_{\beta=1}^{k+1} \hat{w}_\beta \left(\frac{1}{6} - \Delta t z (p_t)_{i,\beta} \right) r_K^{i,\beta} + \sum_{\gamma=1}^L \tilde{w}_\gamma \left(\frac{1}{6} - \Delta t z (p_t)_\gamma \right) r_K^\gamma.
\end{aligned}$$

Then $L_1 > 0$ under the condition (2.3.17). We apply the same quadrature for L_2 to obtain

$$\begin{aligned}
L_2 &= \frac{1}{6} \left(\sum_{i=1}^3 \sum_{\beta=1}^{k+1} \hat{w}_\beta r_K^{i,\beta} + \sum_{\gamma=1}^L \tilde{w}_\gamma r_K^\gamma \right) + \Delta t \left(\sum_{i=1}^3 \sum_{\beta=1}^{k+1} \hat{w}_\beta \check{c}_K^{i,\beta} q_K^{i,\beta} + \sum_{\gamma=1}^L \tilde{w}_\gamma \check{c}_K^\gamma q_K^\gamma \right) \\
&= \sum_{i=1}^3 \sum_{\beta=1}^{k+1} \hat{w}_\beta \left(\frac{1}{6} r_K^{i,\beta} + \Delta t \check{c}_K^{i,\beta} q_K^{i,\beta} \right) + \sum_{\gamma=1}^L \tilde{w}_\gamma \left(\frac{1}{6} r_K^\gamma + \Delta t \check{c}_K^\gamma q_K^\gamma \right).
\end{aligned}$$

Notice that $\check{c} = r/\Phi$ if $q < 0$ while $\check{c} > 0$ if $q > 0$. Therefore, under the condition (2.3.17), each term in the summation above is positive. \square

In the rest part of this section, we will consider second-order scheme only, i.e. $k = 1$, $N = 2$, $L = 0$, then $\hat{w} = \frac{1}{2}$ and $w_\beta = 3\hat{w}_\beta$. Now we can analyze the convection term H_K^c and the result is given below.

Lemma 2.3.2. *Suppose $r^n > 0$ ($c^n > 0$), if α satisfies*

$$\alpha > \max_{i,\beta} \{ |\mathbf{u}_{K_i}^{i,\beta}|, 0 \}, \tag{2.3.19}$$

and the time step satisfies

$$\Delta t \leq \min_{i,\beta} \left\{ \frac{1}{9\ell_K^i \alpha}, \frac{1}{9\ell_K^i (|\mathbf{u}_K^{i,\beta}| + \alpha)} \right\} \Phi_m |K|. \tag{2.3.20}$$

we have $H_K^c(r, \mathbf{u}, c) > 0$.

Proof. Following the same analysis for the source term, we write

$$H_K^c = \sum_{i=1}^3 \sum_{\beta=1}^2 w_\beta H_{i,\beta}^c, \quad H_{i,\beta}^c = \frac{1}{9} r_K^{i,\beta} - \lambda \ell_K^i \widehat{\mathbf{u}} \mathbf{c}^{i,\beta} \cdot \boldsymbol{\nu}_i.$$

We only need to show $H_{i,\beta}^c > 0$.

Case 1: $\boldsymbol{\nu}_i = \mathbf{n}_e$, i.e. $\mathbf{u}^- = \mathbf{u}_K$, $\mathbf{u}^+ = \mathbf{u}_{K_i}$, $c^- = c_K$ and $c^+ = c_{K_i}$. Then

$$H_{i,\beta}^c = \frac{1}{9} r_K^{i,\beta} - \lambda \ell_K^i (\mathbf{u}_{K_i}^{i,\beta} \mathbf{c}_{K_i}^{i,\beta} \cdot \boldsymbol{\nu}_i - \alpha c_{K_i}^{i,\beta} + \alpha c_K^{i,\beta}).$$

Since r and c are both linear functions, we can write the function values of r and c as the interpolation of the values at vertices $\{V_1, V_2, V_3\}$ of K , i.e. for any point x_ρ in K ,

$$r_K^\rho = \mu_1^\rho r_K(V_1) + \mu_2^\rho r_K(V_2) + \mu_3^\rho r_K(V_3), \quad c_K^\rho = \mu_1^\rho c_K(V_1) + \mu_2^\rho c_K(V_2) + \mu_3^\rho c_K(V_3), \quad (2.3.21)$$

with $\mu_m^\rho \geq 0$, $m = 1, 2, 3$, and $\sum_{m=1}^3 \mu_m^\rho = 1$. Then

$$\begin{aligned} H_{i,\beta}^c &= \sum_{m=1}^3 \mu_m^{i,\beta} \left(\frac{1}{9} r_K(V_m) - \lambda \ell_K^i \alpha c_K(V_m) \right) + \lambda \ell_K^i (\alpha - \mathbf{u}_{K_i}^{i,\beta} \cdot \boldsymbol{\nu}_i) c_{K_i}^{i,\beta} \\ &= \sum_{m=1}^3 \mu_m^{i,\beta} \left(\frac{1}{9} \Phi_K(V_m) - \lambda \ell_K^i \alpha \right) c_K(V_m) + \lambda \ell_K^i (\alpha - \mathbf{u}_{K_i}^{i,\beta} \cdot \boldsymbol{\nu}_i) c_{K_i}^{i,\beta}. \end{aligned}$$

Then we have $H_{i,\beta}^c > 0$, if α and Δt satisfy (2.3.19) and (2.3.20), respectively.

Case 2: $\boldsymbol{\nu}_i = -\mathbf{n}_e$, i.e. $\mathbf{u}^+ = \mathbf{u}_K$, $\mathbf{u}^- = \mathbf{u}_{K_i}$, $c^+ = c_K$ and $c^- = c_{K_i}$. Then

$$H_{i,\beta}^c = \frac{1}{9} r_K^{i,\beta} - \lambda \ell_K^i (\mathbf{u}_K^{i,\beta} \mathbf{c}_K^{i,\beta} \cdot \boldsymbol{\nu}_i - \alpha c_{K_i}^{i,\beta} + \alpha c_K^{i,\beta}).$$

Applying (2.3.21) again, we have

$$H_{i,\beta}^c = \sum_{m=1}^3 \mu_m \left(\frac{1}{9} \Phi_K(V_m) - \lambda \ell_K^i \mathbf{u}_K^{i,\beta} \cdot \boldsymbol{\nu}_i - \lambda \ell_K^i w_\beta \alpha \right) c_K(V_m) + \lambda \ell_K^i \alpha c_{K_i}^{i,\beta}.$$

Then we have $H_{i,\beta}^c > 0$ under the condition (2.3.20). \square

Finally, we discuss the diffusion part. We also take $k = 1$, $N = 2$, $L = 0$ and the result is given in the following lemma.

Lemma 2.3.3. *Assume the minimum angle of each triangle K is uniformly bounded away from zero. Suppose $r^n > 0$ ($c^n > 0$), then $H_K^d(r, \mathbf{u}, c) > 0$ under the conditions*

$$\tilde{\alpha} \geq \frac{(3 + \sqrt{3})\Lambda}{2 \min_{K,i,j} (\sin(\theta_K^{i,j}))}, \quad (2.3.22)$$

and

$$\Delta t \leq \frac{\Phi_m |K|}{18\tilde{\alpha}}, \quad \frac{\Delta t}{|K|} \frac{(3 + \sqrt{3})\Lambda}{\min_{K,i,j} (\sin(\theta_K^{i,j}))} \leq \frac{1}{54} \Phi_m, \quad (2.3.23)$$

where $\theta_K^{i,j}$, $i, j = 1, 2, 3, i \neq j$ denotes the angle between the edge e_K^i and e_K^j , and Λ is the largest absolute value of the eigenvalue of \mathbf{D} .

Proof. First, we will consider the term

$$\int_{e_K^i} \left(\{ \mathbf{D}(\mathbf{u}) \nabla c \cdot \boldsymbol{\nu}_i \} + \frac{\tilde{\alpha}}{\ell_K^i} [c] \mathbf{n}_e \cdot \boldsymbol{\nu}_i \right) ds.$$

Following [54], we write

$$\mathbf{D}(\mathbf{u}) \nabla c \cdot \boldsymbol{\nu}_i = \nabla c \cdot \mathbf{D}(\mathbf{u}) \boldsymbol{\nu}_i = \frac{\partial c}{\partial \boldsymbol{\eta}_i} \|\tilde{\boldsymbol{\eta}}_i\|,$$

where

$$\tilde{\boldsymbol{\eta}}_i = \mathbf{D}(\mathbf{u}) \boldsymbol{\nu}_i, \quad \boldsymbol{\eta}_i = \frac{\tilde{\boldsymbol{\eta}}_i}{\|\tilde{\boldsymbol{\eta}}_i\|}.$$

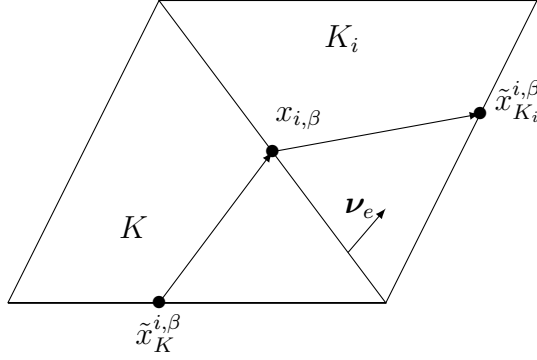


Figure 2.1: Two intersection points for the numerical flux in diffusion part on the triangular mesh.

Define $\boldsymbol{\eta}_K = \boldsymbol{\eta}_i|_K$ and $\boldsymbol{\eta}_{K_i} = \boldsymbol{\eta}_i|_{K_i}$. Likewise for $\tilde{\boldsymbol{\eta}}_K$ and $\tilde{\boldsymbol{\eta}}_{K_i}$. For each quadrature point $x_{i,\beta}$ on the edge e_K^i , we can draw a straight line from $x_{i,\beta}$ with direction $\boldsymbol{\eta}_{K_i}$ intersects ∂K_i at $\tilde{x}_{K_i}^{i,\beta}$. Similarly, we can draw another straight line from $x_{i,\beta}$ with direction $-\boldsymbol{\eta}_K$ intersects ∂K at $\tilde{x}_K^{i,\beta}$. See Figure 2.1 for an illustration. It is easy to verify that at $x = x_{i,\beta}$

$$\begin{aligned}
& \{\mathbf{D}(\mathbf{u})\nabla c \cdot \boldsymbol{\nu}_i\} + \frac{\tilde{\alpha}}{\ell_K^i} [c] \mathbf{n}_e \cdot \boldsymbol{\nu}_i \\
&= \frac{1}{2} \mathbf{D}(\mathbf{u}_K) \nabla c_K \cdot \boldsymbol{\nu}_i + \frac{1}{2} \mathbf{D}(\mathbf{u}_{K_i}) \nabla c_{K_i} \cdot \boldsymbol{\nu}_i + \tilde{\alpha} \frac{(c_{K_i} - c_K)}{\ell_K^i} \\
&= \frac{1}{2} \frac{c_K^{i,\beta} - c(\tilde{x}_K^{i,\beta})}{\|x_K^{i,\beta} - \tilde{x}_K^{i,\beta}\|} \|\tilde{\boldsymbol{\eta}}_K\| + \frac{1}{2} \frac{c(\tilde{x}_{K_i}^{i,\beta}) - c_{K_i}^{i,\beta}}{\|\tilde{x}_{K_i}^{i,\beta} - x_{K_i}^{i,\beta}\|} \|\tilde{\boldsymbol{\eta}}_{K_i}\| + \frac{\tilde{\alpha}}{\ell_K^i} (c_{K_i}^{i,\beta} - c_K^{i,\beta}) \\
&= \left(\frac{\|\tilde{\boldsymbol{\eta}}_K\|}{2\|x_K^{i,\beta} - \tilde{x}_K^{i,\beta}\|} - \frac{\tilde{\alpha}}{\ell_K^i} \right) c_K^{i,\beta} + \left(\frac{\tilde{\alpha}}{\ell_K^i} - \frac{\|\tilde{\boldsymbol{\eta}}_{K_i}\|}{2\|\tilde{x}_{K_i}^{i,\beta} - x_{K_i}^{i,\beta}\|} \right) c_{K_i}^{i,\beta} \\
&\quad - \frac{\|\tilde{\boldsymbol{\eta}}_K\|}{2\|x_K^{i,\beta} - \tilde{x}_K^{i,\beta}\|} c(\tilde{x}_K^{i,\beta}) + \frac{\|\tilde{\boldsymbol{\eta}}_{K_i}\|}{2\|\tilde{x}_{K_i}^{i,\beta} - x_{K_i}^{i,\beta}\|} c(\tilde{x}_{K_i}^{i,\beta}).
\end{aligned}$$

We write the cell average \bar{r}_K^n as

$$\bar{r}_K^n = \sum_{i=1}^3 \sum_{\beta=1}^2 \hat{w}_\beta r_K^{i,\beta} = \sum_{i=1}^3 \sum_{\beta=1}^2 \sum_{m=1}^3 \hat{w}_\beta \mu_m^{i,\beta} \Phi_K(V_m) c_K(V_m).$$

we can rewrite $H_K^d(r, \mathbf{u}, c)$ as

$$\begin{aligned} H_K^d &= \frac{1}{3} \sum_{i=1}^3 \sum_{\beta=1}^2 \sum_{m=1}^3 \hat{w}_\beta \mu_m^{i,\beta} \Phi_K(V_m) c_K(V_m) \\ &\quad + \lambda \sum_{i=1}^3 \ell_K^i \sum_{\beta=1}^2 w_\beta \left[\{\mathbf{D}(\mathbf{u}) \nabla c \cdot \boldsymbol{\nu}_i\} + \frac{\tilde{\alpha}}{\ell_K^i} [c] \mathbf{n}_e \cdot \boldsymbol{\nu}_i \right]_{x=x_{i,\beta}} \\ &= \sum_{i=1}^3 \sum_{\beta=1}^2 w_\beta \left(\frac{1}{9} \sum_{m=1}^3 \mu_m^{i,\beta} \Phi_K(V_m) c_K(V_m) \right. \\ &\quad \left. + \lambda \ell_K^i \left[\{\mathbf{D}(\mathbf{u}) \nabla c \cdot \boldsymbol{\nu}_i\} + \frac{\tilde{\alpha}}{\ell_K^i} [c] \mathbf{n}_e \cdot \boldsymbol{\nu}_i \right]_{x=x_{i,\beta}} \right) \\ &:= \sum_{i=1}^3 \sum_{\beta=1}^2 w_\beta L_{i,\beta} + L, \end{aligned}$$

where

$$\begin{aligned} L_{i,\beta} &= \frac{1}{18} \sum_{m=1}^3 \mu_m^{i,\beta} \Phi_K(V_m) c_K(V_m) + \lambda \ell_K^i \left[\left(\frac{\|\tilde{\boldsymbol{\eta}}_K\|}{2\|x_K^{i,\beta} - \tilde{x}_K^{i,\beta}\|} - \frac{\tilde{\alpha}}{\ell_K^i} \right) c_K^{i,\beta} \right. \\ &\quad \left. + \left(\frac{\tilde{\alpha}}{\ell_K^i} - \frac{\|\tilde{\boldsymbol{\eta}}_{K_i}\|}{2\|\tilde{x}_{K_i}^{i,\beta} - x_{K_i}^{i,\beta}\|} \right) c_{K_i}^{i,\beta} + \frac{\|\tilde{\boldsymbol{\eta}}_{K_i}\|}{2\|\tilde{x}_{K_i}^{i,\beta} - x_{K_i}^{i,\beta}\|} c(\tilde{x}_{K_i}^{i,\beta}) \right], \\ L &= \frac{1}{6} \bar{r}_K^n - \lambda \sum_{i=1}^3 \sum_{\beta=1}^2 \frac{\ell_K^i \|\tilde{\boldsymbol{\eta}}_K\|}{2\|x_K^{i,\beta} - \tilde{x}_K^{i,\beta}\|} c(\tilde{x}_K^{i,\beta}). \end{aligned}$$

We need to make $L_{i,\beta} > 0$. In fact

$$\begin{aligned}
L_{i,\beta} &= \frac{1}{18} \sum_{m=1}^3 \mu_m^{i,\beta} \Phi_K(V_m) c_K(V_m) + \lambda \ell_K^i \left(\frac{\|\tilde{\boldsymbol{\eta}}_K\|}{2\|x_K^{i,\beta} - \tilde{x}_K^{i,\beta}\|} - \frac{\tilde{\alpha}}{\ell_K^i} \right) c_K^{i,\beta} \\
&\quad + \lambda \ell_K^i \left(\frac{\tilde{\alpha}}{\ell_K^i} - \frac{\|\tilde{\boldsymbol{\eta}}_{K_i}\|}{2\|\tilde{x}_{K_i}^{i,\beta} - x_{K_i}^{i,\beta}\|} \right) c_{K_i}^{i,\beta} + \lambda \ell_K^i \frac{\|\tilde{\boldsymbol{\eta}}_{K_i}\|}{2\|\tilde{x}_{K_i}^{i,\beta} - x_{K_i}^{i,\beta}\|} c(\tilde{x}_{K_i}^{i,\beta}) \\
&= \sum_{m=1}^3 \mu_m^{i,\beta} \left(\frac{1}{18} \Phi_K(V_m) + \lambda \ell_K^i \left(\frac{\|\tilde{\boldsymbol{\eta}}_K\|}{2\|x_K^{i,\beta} - \tilde{x}_K^{i,\beta}\|} - \frac{\tilde{\alpha}}{\ell_K^i} \right) \right) c_K(V_m) \\
&\quad + \lambda \ell_K^i \left(\frac{\tilde{\alpha}}{\ell_K^i} - \frac{\|\tilde{\boldsymbol{\eta}}_{K_i}\|}{2\|\tilde{x}_{K_i}^{i,\beta} - x_{K_i}^{i,\beta}\|} \right) c_{K_i}^{i,\beta} + \lambda \ell_K^i \frac{\|\tilde{\boldsymbol{\eta}}_{K_i}\|}{2\|\tilde{x}_{K_i}^{i,\beta} - x_{K_i}^{i,\beta}\|} c(\tilde{x}_{K_i}^{i,\beta}).
\end{aligned}$$

Notice that $\|\tilde{\boldsymbol{\eta}}\| \leq \Lambda$. To make $L_{i,\beta} > 0$, we need

$$\tilde{\alpha} \geq \frac{\ell_K^i \Lambda}{2\|\tilde{x}_{K_i}^{i,\beta} - x_{K_i}^{i,\beta}\|}, \quad \lambda \ell_K^i \left(\frac{\tilde{\alpha}}{\ell_K^i} - \frac{\|\tilde{\boldsymbol{\eta}}_K\|}{2\|x_K^{i,\beta} - \tilde{x}_K^{i,\beta}\|} \right) \leq \frac{1}{18} \Phi_K(V_m).$$

It is easy to compute that

$$\frac{\ell_K^i}{\|\tilde{x}_K^{i,\beta} - x_K^{i,\beta}\|} \leq \frac{6}{(3 - \sqrt{3}) \min_j \sin(\theta_K^{i,j})}.$$

and we conclude $L_{i,\beta} > 0$ under the conditions (2.3.22) and (2.3.23). Finally, we can apply the same idea above to estimate L . Similar to (2.3.21), we write

$$c(\tilde{x}_K^{i,\beta}) = \sum_{m=1}^3 \tilde{\mu}_m^{i,\beta} c_K(V_m),$$

with $0 \leq \tilde{\mu}_m^{i,\beta} \leq 1$ and $\sum_{m=1}^3 \tilde{\mu}_m^{i,\beta} = 1$. Then

$$\begin{aligned}
L &= \frac{1}{6} \bar{r}_K^n - \lambda \ell_K^i \sum_{i=1}^3 \sum_{\beta=1}^2 \frac{\|\tilde{\boldsymbol{\eta}}_K\|}{2\|x_K^{i,\beta} - \tilde{x}_K^{i,\beta}\|} c(\tilde{x}_K^{i,\beta}) \\
&= \sum_{m=1}^3 \left(\frac{1}{18} \Phi_K(V_m) - \lambda \ell_K^i \sum_{i=1}^3 \sum_{\beta=1}^2 \frac{\|\tilde{\boldsymbol{\eta}}_K\| \tilde{\mu}_K^{i,\beta}}{2\|x_K^{i,\beta} - \tilde{x}_K^{i,\beta}\|} \right) c_K(V_m) \\
&\geq \sum_{m=1}^3 \left(\frac{1}{18} \Phi_K(V_m) - \lambda \sum_{i=1}^3 \sum_{\beta=1}^2 \frac{(3 + \sqrt{3}) \Lambda}{2 \min_j \sin(\theta_K^{i,j})} \right) c_K(V_m)
\end{aligned}$$

Therefore, we have $L > 0$ under the condition (2.3.23). \square

Base on the above three lemmas, we can state the following theorem.

Theorem 2.3.4. *Suppose $r^n > 0$ ($c^n > 0$), and the parameters α and $\tilde{\alpha}$ satisfy (2.3.19) and (2.3.22), respectively. Then $\bar{r}^{n+1} > 0$ under the conditions (2.3.17), (2.3.20) and (2.3.23).*

Now, we have proved $\bar{r}_j > 0$ for $j = 1, 2, \dots, N - 1$. To obtain $\bar{r}_N > 0$, we need to subtract (2.2.10) from (2.2.8) to obtain

$$\begin{aligned} (r_{N_t}, \zeta) = & (\mathbf{u}c_N - \mathbf{D}(\mathbf{u})\nabla c_N, \nabla \zeta) + (\check{c}_{Nq} - r_N z_N p_t, \zeta) + \sum_{e \in \Gamma_0} \int_e \widehat{\mathbf{u}c_N} \cdot \mathbf{n}_e[\zeta] ds \\ & - \sum_{e \in \Gamma_0} \int_e \left(\{ \mathbf{D}(\mathbf{u})\nabla c_N \cdot \mathbf{n}_e \}[\zeta] + \{ \mathbf{D}(\mathbf{u})\nabla \zeta \cdot \mathbf{n}_e \}[c_N] + \frac{\tilde{\alpha}}{|e|}[c_N][\zeta] \right) ds. \end{aligned} \quad (2.3.24)$$

Here, we have used the fact that the flux for (2.2.10) is consistent with that in (2.2.8). We can observe that the above equation is similar to (2.2.10). Therefore, following the same analysis above with minor changes we have the following theorem.

Theorem 2.3.5. *Suppose $0 \leq r^n \leq \Phi$, and the conditions in Theorem 2.3.4 are satisfied. Moreover, if the fluxes $\widehat{\mathbf{u}c_j}$ and $\hat{\mathbf{u}}$ are consistent, then $\bar{r}^{n+1} \leq \bar{\Phi}$, under the condition*

$$\Delta t \leq \frac{1}{6z_M p_M}, \quad (2.3.25)$$

where p_M is given in (2.3.18) and $z_M = \max_{1 \leq j \leq N} z_j$.

2.4 Bound-preserving technique for high-order scheme

In this section, we will apply the flux limiter to construct high-order bound-preserving technique.

2.4.1 Flux limiter

We use P^k ($k > 2$) polynomials and write (2.3.13) as

$$\bar{r}_K^{n+1} = \bar{r}_K^n + \lambda \sum_{i=1}^3 \hat{F}_{e^i} + \Delta t \bar{s},$$

where

$$\hat{F}_{e^i} = - \int_{e^i} \widehat{\mathbf{u}} \mathbf{c} \cdot \boldsymbol{\nu}_i ds + \int_{e^i} \left(\{ \mathbf{D}(\mathbf{u}) \nabla c \cdot \boldsymbol{\nu}_i \} + \frac{\tilde{\alpha}}{\ell_K^i} [c] \right) ds, \quad \bar{s} = \overline{\tilde{c}q - rz_1 p_t} \quad (2.4.26)$$

are high-order flux and source, respectively. In Section 2.3, we have demonstrated how to treat the source terms. Therefore, we only discuss the modification of the high-order fluxes only. We will apply the flux limiter [25, 38] and combine the high-order flux \hat{F}_{e^i} and the second-order fluxes, which was analyzed in Section 2.3, denoted as \hat{f}_{e^i} . We define the new flux as

$$\tilde{F}_{e^i} = \hat{f}_{e^i} + \theta_{e^i} (\hat{F}_{e^i} - \hat{f}_{e^i}),$$

where θ_{e^i} is a parameter that to be chosen. Then the cell average can be written as

$$\bar{r}_K^{n+1} = \bar{r}_K^n + \lambda \sum_{i=1}^3 \hat{f}_{e^i} + \lambda \sum_{i=1}^3 \theta_{e^i} (\hat{F}_{e^i} - \hat{f}_{e^i}) + \Delta t \bar{s} = \bar{r}_L^{n+1} + \lambda \sum_{i=1}^3 \theta_{e^i} (\hat{F}_{e^i} - \hat{f}_{e^i}),$$

where

$$\bar{r}_L^{n+1} = \bar{r}_K^n + \lambda \sum_{i=1}^3 \hat{f}_{e^i} + \Delta t \bar{s}$$

is the second order cell average which was proved to be positive if Δt is sufficiently small. Notice that, we need the fluxes in (2.2.10) and (2.2.8) to be consistent. Therefore, we have to discuss the fluxes for all components together. We define $\hat{f}_{e^i}^j$ and $\hat{F}_{e^i}^j$ as the second- and high-order fluxes for component j , $j = 1, 2, \dots, N$, respectively, and the cell average \bar{r} for the j th component to be \bar{r}_j . To compute $\hat{f}_{e^i}^j$, we only replace the c_j in $\hat{F}_{e^i}^j$ in (2.4.26) by a second-order approximation. We cannot change \mathbf{u} , since we want $\sum_{j=1}^N \hat{F}_{e^i}^j = \sum_{j=1}^N \hat{f}_{e^i}^j = \hat{\mathbf{u}}_{e^i}$, which due to the flux consistency requirement. To construct the second-order c_j , we can simply apply the second-order L^2 projection to the high-order c_j , and then apply the limiter discussed in 2.4.2 with $k = 1$ and Φ as the second-order L^2 projection of ϕ . We can choose the parameter θ_{e^i} as follows:

1. For any $K \in \Omega_h$, set $\beta_K = 0$.

2. Define $\hat{F}_{e^i}^N = \hat{\mathbf{u}}_{e^i} - \sum_{j=1}^{N-1} \hat{F}_{e^i}^j$, $\hat{f}_{e^i}^N = \hat{\mathbf{u}}_{e^i} - \sum_{j=1}^{N-1} \hat{f}_{e^i}^j$ and $\bar{r}_n = \bar{\Phi} - \sum_{j=1}^{N-1} \bar{r}_j$.

3. For any $j = 1, 2, \dots, N$, if $\hat{F}_{e^i}^j - \hat{f}_{e^i}^j \geq 0$, take $\theta_{K,e^i}^j = 1$, otherwise set

$$\beta_K = \beta_K + \hat{F}_{e^i}^j - \hat{f}_{e^i}^j.$$

4. For those edges e^i with $\hat{F}_{e^i}^j - \hat{f}_{e^i}^j < 0$, we set $\theta_{K,e^i}^j = \min \left\{ -\frac{\bar{r}_{j,L}^{n+1}}{\lambda \beta_K^m}, 1 \right\}$.

5. Take $\theta_{K,e^i} = \min_{1 \leq j \leq N} \theta_{K,e^i}^j$.

6. For any $e \in \Gamma_0$, we can find $K_1, K_2 \in \Omega_h$ such that $K_1 \cap K_2 = e$. We take

$$\theta_e = \min\{\theta_{K_1,e}, \theta_{K_2,e}\}.$$

Following the same analyses in [12], we have $\bar{r}_j^{n+1} \geq 0, j = 1, 2, \dots, N$. Thus, $0 \leq \bar{r}_j^{n+1} \leq \bar{\Phi}$, since we have the relationship $\bar{r}_1^{n+1} + \bar{r}_2^{n+1} + \dots + \bar{r}_N^{n+1} = \bar{\Phi}$.

Remark 2.4.1. *In (2.2.8)-(2.2.10), we do not compute r_N (c_N) directly. Step 2 in the above algorithm is used to compute the fluxes in (2.3.24). Actually, we can simply take $F_{e^i}^N = -\sum_{j=1}^{N-1} F_{e^i}^j, \hat{f}_{e^i}^N = -\sum_{j=1}^{N-1} f_{e^i}^j$, since we only need the difference of the higher order and lower order fluxes. Moreover, step 5 is used to construct consistent fluxes (See definition 2.2.1).*

2.4.2 Slope limiter

In this section, we discuss the limiters to be applied. As discussed in [22], the traditional slope limiter (2.1.4) cannot be applied. In this paper, we will construct a new one. We consider problem with 2 components first and then extend it to N-component ones. The algorithm is given as follows.

1. Define $\hat{S} = \{x \in K : r(x) \leq 0\}$. Take

$$\hat{r}_1 = r_1 + \theta \left(\frac{\bar{r}_1}{\bar{\Phi}} \Phi - r_1 \right), \quad \theta = \max_{y \in \hat{S}} \left\{ \frac{-r_1(y)\bar{\Phi}}{\bar{r}_1\Phi(y) - r_1(y)\bar{\Phi}}, 0 \right\}. \quad (2.4.27)$$

2. Set $r_2 = \Phi - \hat{r}_1$, and repeat the above step for r_2 .
3. Take $\tilde{r}_1 = \Phi - \hat{r}_2$ as the new approximation.

Remark 2.4.2. *In step 1, we can show that $\hat{r}_1 \geq 0$ which further implies $r_2 \leq \Phi$.*

In step 2, we have

$$\hat{r}_2 = r_2 + \theta \left(\frac{\bar{r}_2}{\bar{\Phi}} \Phi - r_2 \right) = (1 - \theta)r_2 + \theta \frac{\bar{r}_2}{\bar{\Phi}} \Phi \leq (1 - \theta)\Phi + \theta\Phi = \Phi, \forall \theta \in [0, 1],$$

which means the property $\hat{r}_2 \leq \Phi$ is inherited naturally from $r_2 \leq \Phi$, no matter which limiter θ is chosen in this step. This fact gives us enough space to choose θ to modify \hat{r}_2 such that $\hat{r}_2 \geq 0$, as we did in step one. Therefore, after step 3, we have $0 \leq \tilde{r}_1 \leq \Phi$.

It is easy to check that $0 \leq \tilde{r}_1 \leq \Phi$ and $\int_K \tilde{r}(x)dx = \int_K r(x)dx$. Moreover, we can also prove that the limiter does not kill the accuracy.

Theorem 2.4.1. *Let $R(x) \in C^{k+1}(K)$ and $r(x), \Phi(x) \in P^k(K)$ with $0 \leq \bar{r} \leq \bar{\Phi}$ and $\|r(x) - R(x)\|_\infty \leq Ch^{k+1}$. Assume there exist two positive constants Φ_m and Φ_M such that $0 < \Phi_m \leq \Phi(x) \leq \Phi_M$, then $\|\tilde{r}(x) - R(x)\|_\infty \leq Ch^{k+1}$.*

Proof. WLOG, we assume $\theta > 0$ in (2.4.27) and need to show the modification in step 1 keeps the accurate: $\|\hat{r}(x) - r(x)\|_\infty \leq Ch^{k+1}$. Denote $r_m = \min_{x \in K} r(x)$, $r_M = \max_{x \in K} r(x)$. Let $y \in K$ be the point at which the maximum in (2.4.27) is achieved and define $r_y = r(y) < 0$, $\Phi_y = \Phi(y)$. Then

$$\theta = \frac{-r_y}{\frac{\bar{r}}{\Phi} \Phi_y - r_y} \leq \frac{-r_y}{\bar{r} \frac{\Phi_m}{\Phi_M} - r_y} \leq \frac{-r_y}{\bar{r} \frac{\Phi_m}{\Phi_M} - r_y \frac{\Phi_m}{\Phi_M}} = \frac{-r_y}{\bar{r} - r_y} \frac{\Phi_M}{\Phi_m} \leq \frac{-r_m}{\bar{r} - r_m} \frac{\Phi_M}{\Phi_m},$$

which further yields

$$|\hat{r} - r| = \theta \left| \frac{\bar{r}}{\Phi} \Phi - r \right| \leq \frac{\Phi_M}{\Phi_m} \frac{-r_m}{\bar{r} - r_m} \left| \frac{\bar{r}}{\Phi} \Phi - r \right| = \frac{\Phi_M}{\Phi_m} (-r_m) \frac{|\frac{\bar{r}}{\Phi} \Phi - r|}{\bar{r} - r_m}.$$

Since $\frac{\Phi_M}{\Phi_m}$ is a constant and $|-r_m| \leq Ch^{k+1}$, we only need to prove that $\frac{|\frac{\bar{r}}{\Phi} \Phi - r|}{\bar{r} - r_m} \leq C$ for some positive constant C independent of x and h . Notice that

$$\bar{r} \frac{\Phi_m}{\Phi_M} - r_M \leq \bar{r} \frac{\Phi}{\Phi} - r \leq \bar{r} \frac{\Phi_M}{\Phi_m} - r_m,$$

we have

$$\left| \bar{r} \frac{\Phi}{\bar{\Phi}} - r \right| \leq \max \left\{ \left| \bar{r} \frac{\Phi_M}{\Phi_m} - r_m \right|, \left| \bar{r} \frac{\Phi_m}{\Phi_M} - r_M \right| \right\},$$

which further yields

$$\frac{\left| \bar{r} \frac{\Phi}{\bar{\Phi}} - r \right|}{\bar{r} - r_m} \leq \max \left\{ \frac{\left| \bar{r} \frac{\Phi_M}{\Phi_m} - r_m \right|}{\bar{r} - r_m}, \frac{\left| \bar{r} \frac{\Phi_m}{\Phi_M} - r_M \right|}{\bar{r} - r_m} \right\}.$$

Next, we will prove the boundedness of $\frac{\left| \bar{r} \frac{\Phi_M}{\Phi_m} - r_m \right|}{\bar{r} - r_m}$, and $\frac{\left| \bar{r} \frac{\Phi_m}{\Phi_M} - r_M \right|}{\bar{r} - r_m}$, respectively. For the first term, we have

$$\frac{\left| \bar{r} \frac{\Phi_M}{\Phi_m} - r_m \right|}{\bar{r} - r_m} = \frac{\bar{r} \frac{\Phi_M}{\Phi_m} - r_m}{\bar{r} - r_m} \leq \frac{\bar{r} \frac{\Phi_M}{\Phi_m} - r_m \frac{\Phi_M}{\Phi_m}}{\bar{r} - r_m} = \frac{\Phi_M}{\Phi_m}.$$

while for the second term

$$\begin{aligned} \frac{\left| \bar{r} \frac{\Phi_m}{\Phi_M} - r_M \right|}{\bar{r} - r_m} &= -\frac{\bar{r} - r_M + \bar{r} \left(\frac{\Phi_m}{\Phi_M} - 1 \right)}{\bar{r} - r_m} \\ &\leq -\frac{\bar{r} - r_M}{\bar{r} - r_m} - \frac{\bar{r} \left(\frac{\Phi_m}{\Phi_M} - 1 \right)}{\bar{r}} \\ &\leq \frac{r_M - \bar{r}}{\bar{r} - r_m} + 1 - \frac{\Phi_m}{\Phi_M}. \end{aligned}$$

In Appendix C of [51], Zhang proved that for any non-constant polynomial of degree k , say $p(x)$, we have

$$\left| \frac{\bar{p} - \max p(x)}{\bar{p} - \min p(x)} \right| \leq C_k,$$

where C_k is a constant only depends on the polynomial degree k . Thus,

$$\frac{\left| \bar{r} \frac{\Phi_m}{\Phi_M} - r_M \right|}{\bar{r} - r_m} \leq C_k + 1 - \frac{\Phi_m}{\Phi_M},$$

and we finish the proof. \square

Remark 2.4.3. *There are two ways to apply this limiter in an N -component system. One way is to compute the parameter θ_j for the j th component, ($j = 1, 2, \dots, N$) and then take $\theta = \max_j \theta_j$. Another way is to modify r_1, r_2, \dots, r_{N-1} one by one such that $r_1 \in [0, \Phi], r_2 \in [0, \Phi - r_1], r_3 \in [0, \Phi - r_1 - r_2], \dots, r_{N-1} \in [0, \Phi - r_1 - r_2 \cdots - r_{N-2}]$.*

2.4.3 High-order time discretization

In this section, we extend the Euler forward time discretization to high-order ones which are convex combinations of Euler forwards. In this paper, we use third-order strong stability preserving (SSP) high-order time discretization to solve the ODE system $\mathbf{u}_t = \mathbf{L}(\mathbf{u})$:

$$\begin{aligned}\mathbf{u}^{(1)} &= \mathbf{u}^n + \Delta t \mathbf{L}(\mathbf{u}, t^n), \\ \mathbf{u}^{(2)} &= \frac{3}{4} \mathbf{u}^n + \frac{1}{4} (\mathbf{u}^{(1)} + \Delta t \mathbf{L}(\mathbf{u}^{(1)}, t^{n+1})), \\ \mathbf{u}^{n+1} &= \frac{1}{3} \mathbf{u}^n + \frac{2}{3} \left(\mathbf{u}^{(2)} + \Delta t \mathbf{L}(\mathbf{u}^{(2)}, t^n + \frac{\Delta t}{2}) \right).\end{aligned}$$

Another choice is third-order SSP multi-step method:

$$\mathbf{u}^{n+1} = \frac{16}{27} (\mathbf{u}^n + 3 \Delta t \mathbf{L}(\mathbf{u}^n, t^n)) + \frac{11}{27} (\mathbf{u}^{n-3} + \frac{12}{11} \Delta t \mathbf{L}(\mathbf{u}^{n-3}, t^{n-3})).$$

More details can be found in [20, 21, 30].

2.5 Numerical experiments

In this section, we provide numerical experiments to test the accuracy and stability of the high-order bound-preserving DG scheme. In all the examples, we

choose $N = 3$, and consider fluid mixture with 3 components. Moreover, we use the third-order SSP Runge-Kutta discretization in time and P^2 element in space. The computational domain is set to be $\Omega = [0, 2\pi] \times [0, 2\pi]$. To construct Ω_h , we first equally divide Ω into $M \times M$ rectangles and the triangles are obtained by equally divide each rectangle into two. See Figure 2.2 for the mesh.

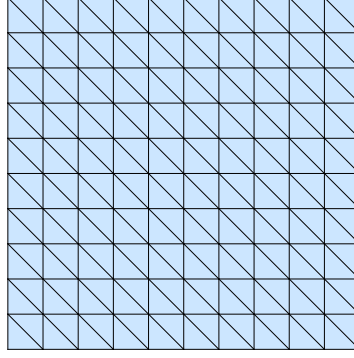


Figure 2.2: Triangular mesh ($M = 10$)

Example 2.5.1. We set the initial conditions as

$$c_{1,0}(x, y) = \frac{1}{6}\left(1 + \frac{1}{2}(\cos x + \cos y)\right), \quad c_{2,0}(x, y) = \frac{1}{3}(1 + \cos x \cos y),$$

$$c_{3,0}(x, y) = 1 - c_{1,0}(x, y) - c_{2,0}(x, y), \quad p_0(x, y) = \cos x \cos y - 1,$$

and the source variables are taken as

$$\tilde{c}_1(x, y, t) = \frac{1}{6}\left(1 + \frac{1}{2}e^{-\gamma t}(\cos x + \cos y - \frac{1}{2}\sin x \cos y - \frac{1}{2}\sin y \cos x)\right),$$

$$\tilde{c}_2(x, y, t) = \frac{1}{3}\left(1 + e^{-2\gamma t}(\cos x \cos y - \frac{1}{2}\sin^2 x \cos^2 y - \frac{1}{2}\cos^2 x \sin^2 y)\right),$$

$$\tilde{c}_3(x, y, t) = 1 - \tilde{c}_1(x, y, t) - \tilde{c}_2(x, y, t), \quad q(x, y, t) = 2e^{-2t}.$$

Other parameters are chosen as

$$\phi(x, y) = \mu(c_1, c_2) = k(x, y) = a(x, y, c_1, c_2) = z_1 = z_2 = z_3 = 1,$$

$$\mathbf{D}(\mathbf{u}) = \text{diag}(\gamma, \gamma).$$

It is easy to verify that the exact solutions are

$$c_1(x, y, t) = \frac{1}{6}(1 + \frac{1}{2}e^{-\gamma t}(\cos x + \cos y)), \quad c_2(x, y, t) = \frac{1}{3}(1 + e^{-2\gamma t} \cos x \cos y),$$

$$c_3(x, y, t) = 1 - c_1(x, y, t) - c_2(x, y, t), \quad p(x, y, t) = e^{-2t}(\cos x \cos y - 1).$$

In the numerical simulation, we choose $\gamma = 0.01$, final time $T = 0.01$ and $\Delta t = 0.001h^2$ to reduce the time error. The computational results are shown in Table 2.1, illustrating the L^2 error and convergence orders for c_1 and c_2 with and without bound-preserving technique. From the table, we observe optimal convergence rates. Therefore, the flux limiter and slope limiter do not degenerate the convergence order.

Example 2.5.2. We choose the initial conditions as

$$c_{1,0}(x, y) = \begin{cases} 1, & x \leq \frac{\pi}{2}, y \leq \frac{\pi}{2}, \\ 0, & \text{otherwise.} \end{cases} \quad c_{2,0}(x, y) = \begin{cases} 1, & x \geq \frac{3\pi}{2}, y \geq \frac{3\pi}{2}, \\ 0, & \text{otherwise.} \end{cases}$$

$$c_{3,0}(x, y) = 1 - c_{1,0}(x, y) - c_{2,0}(x, y) \quad \text{and} \quad p_0(x, y) = \cos\left(\frac{x}{2}\right) + \cos\left(\frac{y}{2}\right).$$

Other parameters are taken as

$$z_1 = z_2 = 1, z_3 = 10, q(x, y, t) = 0, \mathbf{D}(\mathbf{u}) = 0,$$

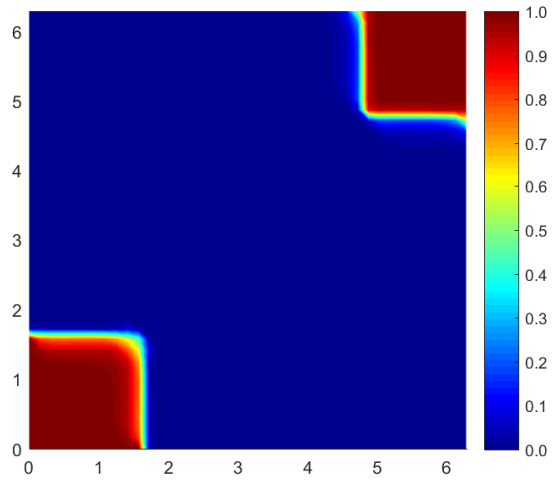
$$\mu(c_1, c_2) = k(x, y) = a(x, y, c_1, c_2) = \phi(x, y) = 1.$$

M	c_1				c_2			
	no limiter		with limiter		no limiter		with limiter	
	L^2 error	order	L^2 error	order	L^2 error	order	L^2 error	order
5	3.02e-3	–	4.61e-3	–	2.12e-2	–	2.39e-2	–
10	5.00e-4	2.59	5.30e-4	3.12	3.29e-3	2.69	3.47e-3	2.78
20	8.85e-5	2.50	8.86e-5	2.58	5.34e-4	2.63	5.34e-4	2.70
40	1.25e-5	2.82	1.25e-5	2.82	7.25e-5	2.88	7.25e-5	2.88
80	1.71e-6	2.87	1.71e-6	2.87	9.41e-6	2.95	9.41e-6	2.95
160	2.02e-7	3.09	2.02e-7	3.09	1.16e-6	3.02	1.16e-6	3.02

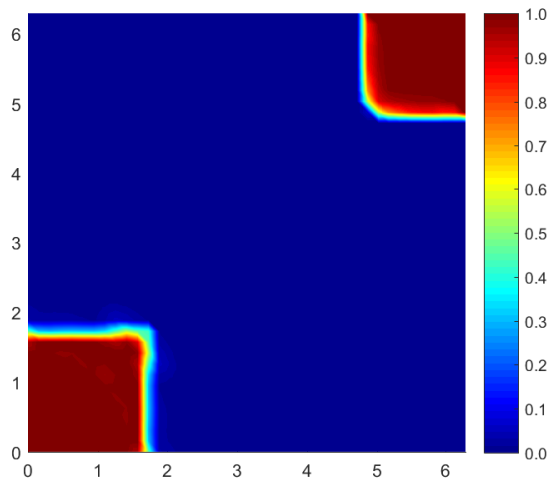
Table 2.1: Example 2.5.1: Accuracy test for c_1 and c_2 with and without bound-preserving technique.

We use this example to demonstrate the stability of the scheme. We choose $\mathbf{D} = \mathbf{0}$, then the diffusion term will not provide any dissipation to the scheme. We compute the components c_1 and c_2 at time $T = 0.1s$ and $T = 0.6s$, respectively, with $M = 40$ and $\Delta t = 0.001h^2$ ($h = \frac{2\pi}{40}$). The numerical results are shown as Figure 2.3. From the figure we can see that the concentrations c_1 and c_2 are between 0 and 1. To test the effectiveness of the bound-preserving technique, we simulate the example without the bound-preserving limiters, and the numerical approximations blow up at about 0.003s even though we take time step size as small as $\Delta t = 0.0001h^2$. In [22], we demonstrated that the reason for the blow-up of the numerical approximations is the ill-posedness of the system. This example demonstrates the necessity of the bound-preserving technique in solving compressible miscible displacements in porous media.

Example 2.5.3. *We investigate the displacement of 3-phase porous media flow in the five-spot arrangement of injection and production wells. The computational domain is a square region taken as quarter-of-a-five-spot pattern. The three phases are light oil c_1 (with low viscosity and high compressibility), heavy oil c_2 (with high viscosity and low compressibility) and water c_3 (with medium viscosity and medium compressibility).*



(a) $T=0.1$ s



(b) $T=0.6$ s

Figure 2.3: Example 2.5.2: Numerical approximations of c_1 and c_2

The initial concentrations of oil (water) are

$$c_{1,0}(x, y) = \begin{cases} 1, & x \leq \frac{\pi}{2}, y \leq \frac{\pi}{2}, \\ 0, & \text{otherwise.} \end{cases}$$

$$c_{2,0}(x, y) = \begin{cases} 0, & x \leq \frac{\pi}{2}, y \leq \frac{\pi}{2}, \\ 1, & \text{otherwise.} \end{cases}$$

$$c_{3,0}(x, y) = 0.$$

Therefore, the lower-left part of the region is light oil enrichment area while the other part is heavy oil enrichment area. Moreover, no water exists initially and the initial pressure is taken as 0 in the whole computational domain. To simulate the random perturbation of porosity and permeability around their average value, we choose the porosity and permeability as

$$\phi(x, y) = 0.5 + 0.05 \sin(5x) \sin(5y) \quad \text{and} \quad k(x, y) = 1.0 + 0.1 \cos(5x) \cos(5y),$$

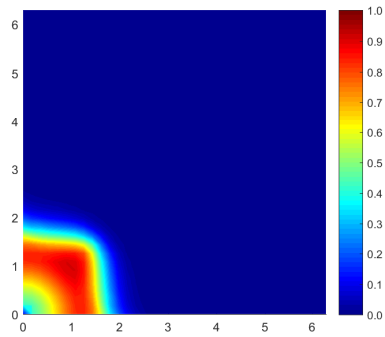
respectively. Other parameters are taken as

$$\mu(c_1, c_2, c_3) = 0.4c_1 + 2.0c_2 + 1.0c_3,$$

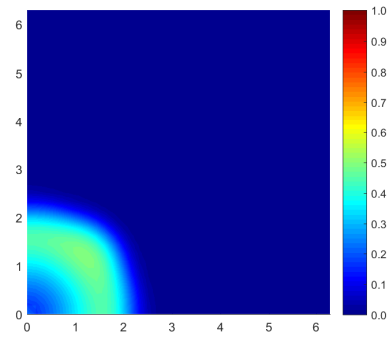
$$z_1 = 1.2, \quad z_2 = 0.8, \quad z_3 = 1.0, \quad \mathbf{D} = \text{diag}(|\mathbf{u}|, |\mathbf{u}|).$$

The injection well is located in lower-left corner and production well is located in upper-right corner, treated as δ sources.

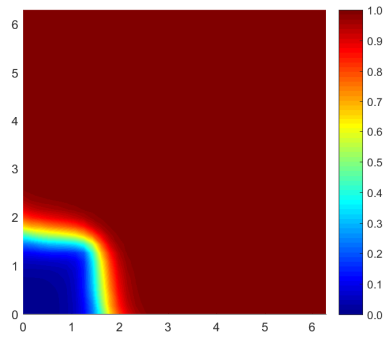
This example is used for petroleum production simulations. We compute the components c_1 and c_2 at time $T = 0.2, 0.8$ with $M = 35$ and $\Delta t = 0.001h^2$ ($h = \frac{2\pi}{35}$). The distributions of c_1 , c_2 and $c_1 + c_2$ at different time are shown in figures



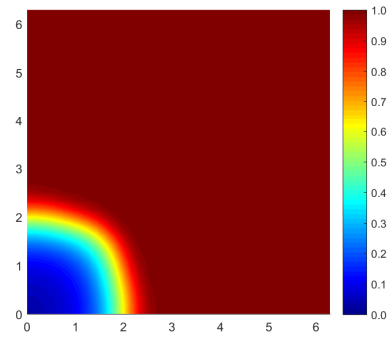
(a) c_1 at $T=0.2$ s



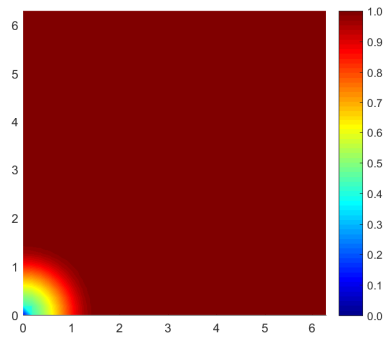
(b) c_1 at $T=0.8$ s



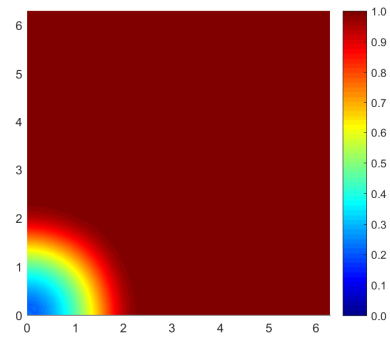
(c) c_2 at $T=0.2$ s



(d) c_2 at $T=0.8$ s



(e) $c_1 + c_2$ at $T=0.2$ s



(f) $c_1 + c_2$ at $T=0.8$ s

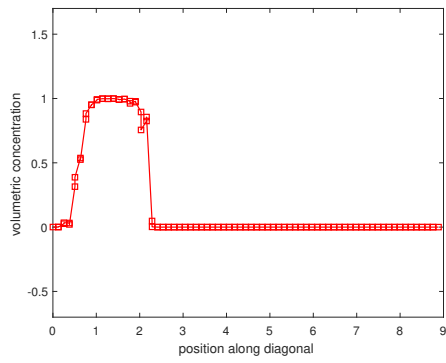
Figure 2.4: Example 2.5.3: Concentrations of c_1 , c_2 and $c_1 + c_2$.

2.4a-2.4f, respectively. From the figure we can see that c_1 , c_2 and $c_1 + c_2$ are all between 0 and 1.

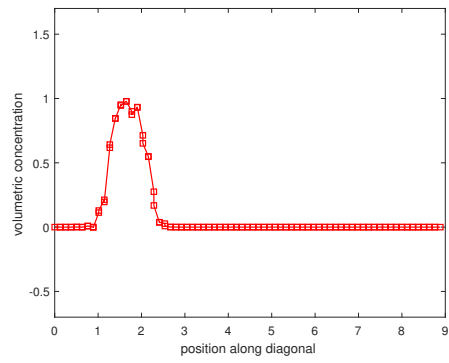
Example 2.5.4. *To show the significance of the bound-preserving technique in real petroleum production simulations, we choose the exact parameters in Example 2.5.3, except $\mathbf{D} = \mathbf{0}$ in order to avoid any dissipation to the scheme which is resulted from the diffusion term.*

This example is used for petroleum production simulations when diffusion effect is negligible. We compute the components c_1 and c_2 at time $T = 0.2, 0.8$ with $M = 35$ and $\Delta t = 0.001h^2 (h = \frac{2\pi}{35})$. The distributions of c_1 , c_2 , and c_3 at different time along diagonal $y = x$ are shown in figures 2.5a-2.5f, respectively. From the figures we can see that the concentrations c_1 , c_2 , and c_3 are between 0 and 1.

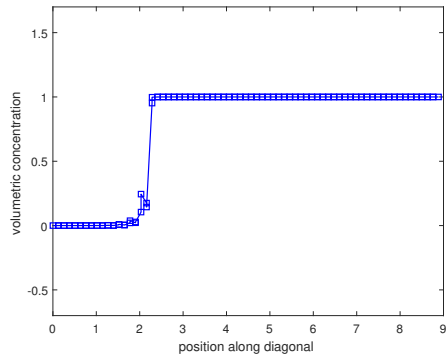
However, the numerical approximations without bound-preserving limiters blow up at about $T = 0.25$ if we take the same time step as before. The distribution of components along diagonal at time $T = 0.1, 0.2$ are shown in figures 2.6a-2.6f, from which we can observe strong oscillations and physically irrelevant values. Further experiments show that, even though we take the time step as small as $\Delta t = 0.0001h^2$, the numerical approximations still blow up at about $T = 0.26$, which implies the necessity of the bound-preserving technique.



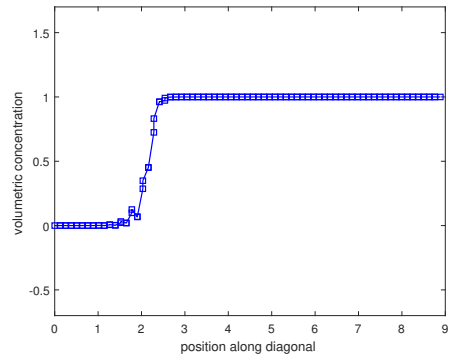
(a) c_1 at $T=0.2$ s



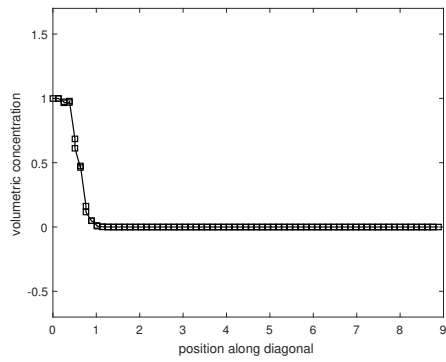
(b) c_1 at $T=0.8$ s



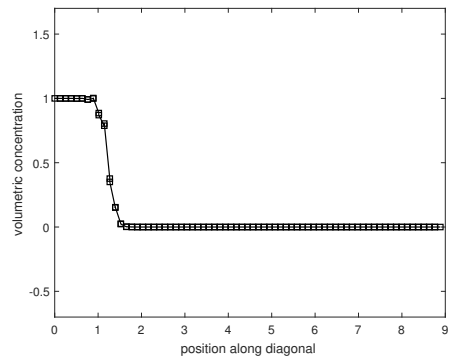
(c) c_2 at $T=0.2$ s



(d) c_2 at $T=0.8$ s

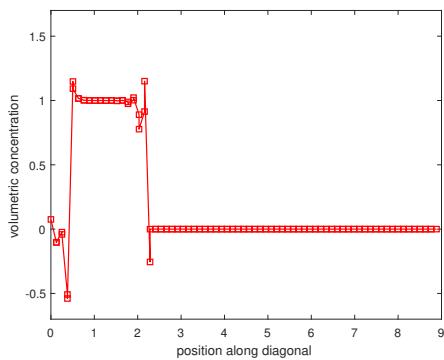


(e) c_3 at $T=0.2$ s

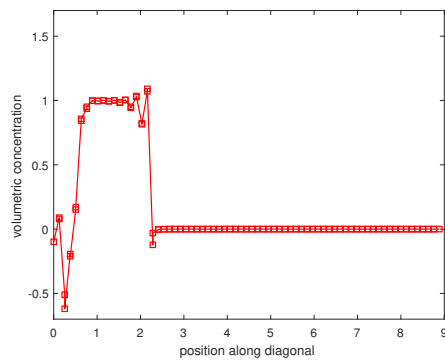


(f) c_3 at $T=0.8$ s

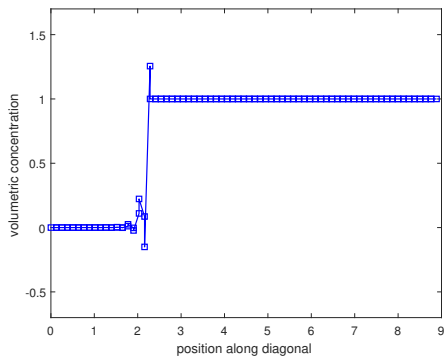
Figure 2.5: Example 2.5.4: Concentrations of c_1 , c_2 and c_3 with limiters



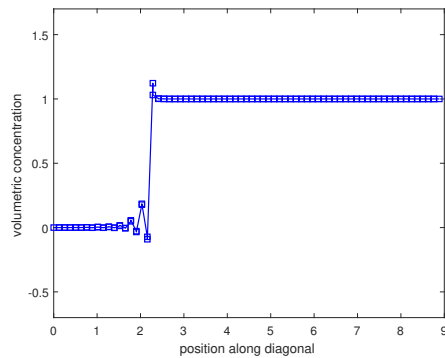
(a) c_1 at $T=0.1$ s



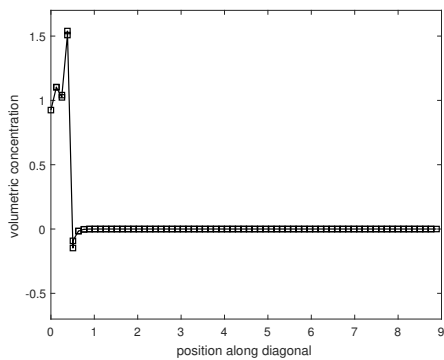
(b) c_1 at $T=0.2$ s



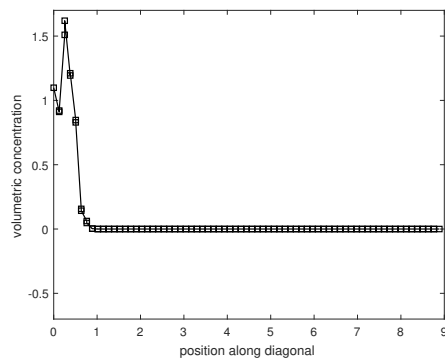
(c) c_2 at $T=0.1$ s



(d) c_2 at $T=0.2$ s



(e) c_3 at $T=0.1$ s



(f) c_3 at $T=0.2$ s

Figure 2.6: Example 2.5.4: Concentrations of c_1 , c_2 and c_3 without limiters

2.6 Concluding remarks

In this paper, we constructed high-order bound-preserving DG methods for compressible miscible displacements in porous media on triangular meshes. We have applied the technique to the problem with multi-component fluid mixtures. Numerical simulations shown the accuracy and necessity of the bound-preserving technique.

Chapter 3

High-Order bound-preserving discontinuous Galerkin methods for wormhole propagation on triangular meshes¹

Abstract

Wormhole propagation, arising in petroleum engineering, is used to describe the distribution of acid and the increase of porosity in carbonate reservoir under dissolution of injected acid. The important physical features of porosity and acid concentration include their boundedness between 0 and 1, as well as the monotone increasing for porosity. How to keep these properties in the simulation is

¹This chapter has been completed as an article to submit to Journal of Computational Physics. Citation: Z. Xu, Y. Yang, H. Guo (2019).

crucial to the robustness of the numerical algorithm. In this paper, we propose high-order bound-preserving discontinuous Galerkin methods to keep these important physical properties. The main technique is to introduce a new variable r to replace the original acid concentration and use a consistent flux pair to deduce a ghost equation such that the positive-preserving technique can be applied on both original and deduced equations. A high-order slope limiter is used to keep a polynomial upper bound which changes over time for r . Moreover, the high-order accuracy is attained by the flux limiter. Numerical examples will be given to demonstrate the high-order accuracy and bound-preserving property of the numerical technique.

Key Words: wormhole propagation, bound-preserving, high-order, discontinuous Galerkin method, triangular meshes, flux limiter

3.1 Introduction

As an important technique of enhanced oil recovery (EOR), acid treatment has been widely practiced in carbonate reservoir to improve the productivity of oil wells. In this technique, acid is injected into wells to dissolve the fines deposited in wellbore and the rock near the wellbore. By doing so, the permeability and porosity of the rock close to a well can be increased prominently, which facilitates oil flow into production well and thereby improves the production rate of oil.

However, the efficiency of this technique has a strong relevance with the dissolution patterns which depend on the injection rate. With a very low injection

rate, the acid only dissolves the face of wellbore since it will be consummated all before they get into deeper region and this scenario is called face dissolution pattern. In contrast, with a very high injection rate, the acid can be pushed uniformly into the wellbore region with certain depth and this results is the so-called uniform dissolution pattern. In addition to the above two extreme cases, with an appropriate injection rate, wormhole patterns can be formed as the injected acid in the rock tends to flow through the paths with high permeability and porosity, which causes the permeability and porosity of these path to be further increased under the dissolution of acid, and facilitate more acid to flow through. Therefore, under optimal injection rate, maximum number of narrow channels with high conductivity will be formed in the rocks after the acidizing process. These highly conducting channels, also known as wormholes because of its shape, can build a good connectivity between reservoir and wellbore, and improve the productivity of oil well enormously. Because of the important role that wormhole plays in improving productivity, a lot of research works have been taken to investigate the formation and propagation of wormholes.

In the early days, researchers investigated the wormhole propagation phenomenon by means of experiments [61, 59]. Later, several mathematical models, such as dimensionless model, capillary tube model, network model, and continuum models, were established to help people understand the process of wormhole propagation. Among these models, the most popular one is the two-scale continuum model developed by Panga et al. in [66], where the authors proposed a partial differential equations (PDE) system to describe the formation

and propagation of wormholes. There were a lot of follow-up works based on this model. In [72], the authors analyzed the front instability of wormhole propagation theoretically and numerically. In [65], Maheshwari et al. presented a 3D simulation for this model. A parallel simulation was conducted by Wu et al. in [68] under a modification of flow equation. In [56], the authors studied the numerical-simulation approach for a modified model. Later, Wei et al. extended this model from single phase to two-phase flow in [67] and discussed the simulation results. Besides the above, many researchers designed specific numerical schemes for this kind of models as well. In [62], the authors constructed a conservative scheme for flow and transport based on mixed finite element method. Later, Li et al. applied finite difference methods to this problem in [63, 64]. Recently, the discontinuous Galerkin (DG) method was applied to this model in [60]. In all the above works, some theoretical works, such as the stability and error estimates, were established under different norms. However, to the best of our knowledge, no works have been undertaken to preserve the boundedness of porosity and concentration of acid without loss of mass conservative. In fact, the boundedness of these variables is essential for the stability of numerical simulations. Firstly, the rate of change of porosity ϕ in this model depends on the concentration of acid c_f . If the exact solutions contain large gradients or even discontinuities, the numerical approximations of c_f can be negative, which further leads to $\phi < 0$ in some regions with low porosity. Secondly, the oscillations of ϕ itself near the wormhole may also result in negative values. Both of the above two cases will bring a negative coefficient in the diffusion term of the

transport equation, leading to ill-posedness of the problems, and finally cause the blow-up of the numerical simulations. We will demonstrate this feasibility by a numerical example in Section 6 and show how bound-preserving technique can prevent the blow-up phenomenon. Moreover, as we will see in the later section, many coefficients in the model appear as functions of ϕ , which require ϕ to take values in the physically relevant range $[0, 1]$ as well. To construct high-order bound-preserving technique, we have to apply suitable limiters to modify the numerical approximations. Therefore, we would like to use DG methods.

The DG methods become increasingly popular due to their good stability, high-order accuracy, and flexibility on h-p adaptivity. In 2010, the genuinely maximum-principle-satisfying high-order DG schemes were constructed for conservation laws on rectangular meshes in [50] by Zhang and Shu. The basic idea is to take the test function to be 1 in each cell to yield an equation satisfied by the cell average of the target variable r , and prove the desired boundedness of the cell average \bar{r} . Then a slope limiter which do not affect accuracy and mass conservation can be used to modify the variable r to obtain a new one $\tilde{r} = \bar{r} + \theta(r - \bar{r})$ such that \tilde{r} has the physically relevant bounds. In the case that the variable r only need a lower bound zero, this technique is also called positivity-preserving technique. The physically positivity-preserving and bound-preserving numerical schemes have been actively studied since then. In 2012, this technique has been successfully extended to triangular meshes in [53], where the general criteria for quadrature rule on triangular elements was proposed. After that, this technique was applied to many problems including compressible Euler

equations with source terms [52], hyperbolic equations involving δ -singularities [44, 70], relativistic hydrodynamics [29], extended MHD equations [55], shallow water equations [37], etc. For convection-diffusion equations, the genuinely second-order maximum-principle-preserving technique were introduced in [54]. Subsequently, the extension to third-order or even higher order bound-preserving techniques for parabolic equations were also developed in [69, 10, 58]. Besides the above, the flux limiter [39, 38, 25] can also be used to obtain the high-order accuracy and maintain the boundedness. However, with the flux limiters we have to modify the numerical fluxes, hence the accuracy is not easy to analyze. Recently, in [22, 57], the authors studied miscible displacements in porous media and applied the techniques introduced in [54, 25, 38, 39] to preserve the two bounds, 0 and 1, of the volumetric fractions. In this paper, we will construct high-order bound-preserving DG schemes for the porosity of the rocks ϕ and the concentration of the acid c_f . However, there are significant differences from most of the previous techniques. First of all, most of the problems in [39, 50] satisfy maximum-principles while the concentration of acid c_f does not. To solve this problem, we derive a ghost equation satisfied by $c = 1 - c_f$ and apply the positivity-preserving technique to both c_f and c . Secondly, the high-order positivity-preserving technique in this paper is based on the flux limiter [38, 25]. The basic idea is to combine higher order and lower order fluxes to construct a new one which can yield positive numerical cell averages. However, for triangular meshes, first-order fluxes are not easy to construct. Therefore, we will consider the second-order flux as the lower order one. Moreover, to

obtain the equation satisfied by the cell averages, we need to numerically approximate $r = \phi c_f$ instead of c_f . By doing so, the upper bound of r is not a constant and the traditional slope limiter may fail to work [22]. Therefore, a new bound-preserving limiter will be introduced. Finally, different from [22, 57], the porosity is increasing and less than 1. Therefore, the upper bound of r is changing during time evolution and special techniques will be introduced to make ϕ to be physically relevant. In summary, the whole algorithm can be separated into four parts. We first apply positivity-preserving technique to obtain positive ϕ_t and use which as another source to find the velocity and pressure. Then apply the positivity-preserving technique again to ϕ and c_f simultaneously to obtain positive numerical cell averages by the flux limiter [38, 25]. Subsequently, we choose consistent flux pair [22, 57] with suitable parameters in the flux limiter in the concentration and pressure equations to obtain the positivity of $1 - c_f$. Finally, we introduce suitable limiters to obtain physically relevant numerical approximations.

The rest of the paper is organized as follows. In Section 2, we introduce the mathematical model of wormhole propagation. In Section 3, we establish the DG scheme used in this paper. In Sections 4 and 5, we construct the second-order bound-preserving scheme and then extend it to high-order spatial and time discretizations. Some numerical examples are given in Section 6. We will end in Section 7 with some concluding remarks.

3.2 Mathematical model

Let the computational domain $\Omega = [0, 2\pi] \times [0, 2\pi]$ and simulation time $J = [0, T]$, the mathematical model of the wormhole propagation is given as follows:

$$\frac{\partial \phi}{\partial t} + \nabla \cdot \mathbf{u} = f, \quad (x, y) \in \Omega, \quad 0 < t \leq T, \quad (3.2.1)$$

$$\mathbf{u} = \frac{-\kappa(\phi)}{\mu} \nabla p, \quad (x, y) \in \Omega, \quad 0 < t \leq T, \quad (3.2.2)$$

$$\frac{\partial(\phi c_f)}{\partial t} + \nabla \cdot (\mathbf{u} c_f) = \nabla \cdot (\phi \mathbf{D} \nabla c_f) + k_c a_v (c_s - c_f) + f_I c_I - f_P c_f \quad (3.2.3)$$

$$\frac{\partial \phi}{\partial t} = \frac{\alpha k_c a_v (c_f - c_s)}{\rho_s}, \quad (x, y) \in \Omega, \quad 0 < t \leq T, \quad (3.2.4)$$

where ϕ is the porosity which is defined as the percentage of the empty space in a rock, κ is the permeability that measures the ability for a rock to allow fluid to pass through it, \mathbf{u} is the Darcy's velocity defined as the volume of flow crossing a unit across-section per unit time, p is the pressure of fluid in porous media, and μ is the viscosity of fluid. $f = f_I - f_P$ is the external volumetric flow rate with $f_I = \max\{f, 0\}$ being the injection flow rate and $f_P = -\min\{f, 0\}$ being the production flow rate. c_f , c_s and c_I are the concentrations of acid in the fluid phase, the fluid-solid interface and in the injected flow, respectively. \mathbf{D} is the dispersion tensor for the acid in porous media and k_c is the local mass-transfer coefficient. a_v is the interfacial area available for reaction, ρ_s is the density of the rock and α is the dissolving constant of the acid, defined as grams of solid dissolved per mole of acid reacted. Moreover, in the case of first order kinetic reaction, the concentration c_s of acid in the fluid-solid interface have a simple

relationship with c_f :

$$c_s = \frac{c_f}{1 + k_s/k_c},$$

where k_s is the kinetic constant for reaction. The coefficients κ and a_v are functions of ϕ defined as

$$\frac{\kappa}{\kappa_0} = \frac{\phi}{\phi_0} \left(\frac{\phi(1 - \phi_0)}{\phi_0(1 - \phi)} \right)^2, \quad \frac{a_v}{a_0} = \frac{1 - \phi}{1 - \phi_0}, \quad (3.2.5)$$

respectively, where κ_0, a_0 , and ϕ_0 are the initial values for κ, a_v, ϕ . Throughout this paper, the values $\mu, k_c, k_s, \alpha, \rho_s$ are positive constants, $\mathbf{D}, f, f_I, f_P, c_I$ are known functions independent of time and ϕ, \mathbf{u}, p, c_f , are unknown time-dependent variables.

We consider impermeable boundary conditions

$$\mathbf{u} \cdot \mathbf{n} = 0, \quad (\mathbf{D}\nabla c - c\mathbf{u}) \cdot \mathbf{n} = 0,$$

where \mathbf{n} is the unit outer normal of the boundary $\partial\Omega$. The initial concentration and porosity are given as

$$c_f(x, y, 0) = c_0(x, y), \quad \phi(x, y, 0) = \phi_0(x, y), \quad (x, y) \in \Omega.$$

Before we finish this section, we would like to make an important reasonable hypothesis for the initial porosity: $0 < \phi_* \leq \phi_0(x, y) \leq \phi^* < 1$.

3.3 The DG scheme

In this section, we will construct the DG scheme for wormhole propagation on triangular meshes. We first demonstrate the notations to be used throughout the paper.

Consider a regular triangulation Ω_h of domain Ω , i.e. $\exists \rho > 0$, such that $\text{diam}(B_K) \geq \rho \text{diam}(K), \forall K \in \Omega_h$, where B_K is the largest ball contained in K . For any triangle $K \in \Omega_h$, we denote the three edges of K to be e_K^i ($i = 1, 2, 3$), with corresponding lengths ℓ_K^i ($i = 1, 2, 3$), unit outer normal vectors $\boldsymbol{\nu}_K^i$ ($i = 1, 2, 3$) and neighboring elements K_i ($i = 1, 2, 3$). We denote $\Gamma = \bigcup_{K \in \Omega_h} \{e | e \in \partial K\}$ to be the set of all cell interfaces and $\Gamma_0 = \Gamma \setminus \partial\Omega_h$ as all the interior ones. Set a predetermined constant unit vector $\boldsymbol{\nu}_0$ which is not parallel to any edge e and define \boldsymbol{n}_e as the unit normal vector of each edge $e \in \Gamma$ such that $\boldsymbol{n}_e \cdot \boldsymbol{\nu}_0 > 0$. For any discontinuous function v (scalar or vector) crossing edge e , let v_e^\pm denote its traces on e evaluated from K or K_i . The ' \pm ' for each edge e_K^i in the cell K is determined by the inner product of $\boldsymbol{\nu}_K^i$ and $\boldsymbol{\nu}_0$ as follows:

$$\begin{aligned} v_e^- &= v_K, \quad v_e^+ = v_{K_i}, \quad \text{if } \boldsymbol{\nu}_0 \cdot \boldsymbol{\nu}_K^i > 0, \\ v_e^+ &= v_K, \quad v_e^- = v_{K_i}, \quad \text{if } \boldsymbol{\nu}_0 \cdot \boldsymbol{\nu}_K^i < 0. \end{aligned}$$

Moreover, we define the jump and average of v (either a scalar or a vector) on the cell interface e as

$$[v]_e = v_e^+ - v_e^-, \quad \{v\}_e = \frac{1}{2}(v_e^+ + v_e^-).$$

The finite element spaces are chosen as

$$V_h = \{v : v|_K \in P^k(K), \forall K \in \Omega_h\} \text{ and } W_h = V_h \times V_h,$$

where $P^k(K)$ denotes the space of polynomials of degree at most k in K . Then the semidiscrete DG scheme for (3.2.1) - (3.2.4) can be written as: find $\phi, r, p \in$

V_h and $\mathbf{u} \in W_h$ such that for any $\zeta, \xi, v \in V_h$ and $\boldsymbol{\eta} \in W_h$, the following equations hold:

$$\left(\frac{\partial \phi}{\partial t}, \zeta \right) = (\mathbf{u}, \nabla \zeta) + \sum_{e \in \Gamma_0} \int_e \hat{\mathbf{u}} \cdot \mathbf{n}_e [\zeta] ds + (f, \zeta), \quad (3.3.6)$$

$$(a(\phi) \mathbf{u}, \boldsymbol{\eta}) = (p, \nabla \cdot \boldsymbol{\eta}) + \sum_{e \in \Gamma} \int_e \hat{p} \mathbf{n}_e \cdot [\boldsymbol{\eta}] ds, \quad (3.3.7)$$

$$\begin{aligned} \left(\frac{\partial r}{\partial t}, \xi \right) &= (\mathbf{u} c_f - \phi D \nabla c_f, \nabla \xi) + \sum_{e \in \Gamma_0} \int_e \widehat{\mathbf{u} c_f} \cdot \mathbf{n}_e [\xi] ds \\ &\quad - \sum_{e \in \Gamma_0} \int_e \left(\{ \phi \mathbf{D}(\mathbf{u}) \nabla c_f \} \cdot \mathbf{n}_e [\xi] + \{ \phi \mathbf{D}(\mathbf{u}) \nabla \xi \} \cdot \mathbf{n}_e [c_f] + \frac{\tilde{\alpha}}{|e|} [c_f] [\xi] \right) ds \\ &\quad + (f_{IC_I} - f_{PC_f} - B_1(\phi) c_f, \xi), \end{aligned} \quad (3.3.8)$$

$$\left(\frac{\partial \phi}{\partial t}, v \right) = (B_2(\phi) c_f, v), \quad (3.3.9)$$

where

$$a(\phi) = \frac{\mu}{k}, B_1(\phi) = \frac{a_0(1-\phi)k_s k_c}{(1-\phi_0)(k_s + k_c)}, B_2(\phi) = \frac{\alpha a_0(1-\phi)k_s k_c}{\rho_s(1-\phi_0)(k_s + k_c)}.$$

Moreover, we use a new variable r instead of ϕc_f on the left hand side of (3.3.8), and define c_f as the L^2 -projection of $\frac{r}{\phi}$ if $k \geq 2$, while take c_f to be the interpolation of $\frac{r}{\phi}$ at the three vertices in each triangle K if $k = 1$.

Following the idea in [22, 57], we take a consistent flux pair $\hat{\mathbf{u}}, \widehat{\mathbf{u} c_f}$ in the sense that $\hat{\mathbf{u}} = \widehat{\mathbf{u} c_f}$ when $c_f = 1$. The consistent flux pair is used in the construction of the bound-preserving techniques. The numerical fluxes $\hat{\mathbf{u}}, \widehat{\mathbf{u} c_f}$ and \hat{p} in (3.3.6)-(3.3.9) are chosen as

$$\begin{aligned} \hat{\mathbf{u}}|_e &= \{\mathbf{u}\}_e, \quad \hat{p}|_e = \{p\}_e, \quad \widehat{\mathbf{u} c_f}|_e = \{\mathbf{u} c_f\}_e - \alpha [c_f]_e \mathbf{n}_e, \quad \text{if } e \in \Gamma_0, \\ \hat{\mathbf{u}}|_e &= 0, \quad \hat{p}|_e = p_K, \quad \widehat{\mathbf{u} c_f}|_e = 0, \quad \text{if } e \in \partial\Omega \cap \partial K. \end{aligned} \quad (3.3.10)$$

In the DG schemes, we introduced two penalty parameters α and $\tilde{\alpha}$. These two parameters will be chosen by the bound-preserving technique.

3.4 Second-order bound-preserving schemes

In this section, we will construct second-order bound-preserving scheme with forward Euler time discretization. High-order time discretizations will be discussed in the next section. At time level n , we assume $\phi^0 < \phi^n < 1$ and $0 \leq r^n \leq \phi^n$, and would like to construct physically relevant numerical approximations at time level $n + 1$, i.e. $\phi^n \leq \phi^{n+1} < 1$ and $0 \leq r^{n+1} \leq \phi^{n+1}$.

At time level n , we will first solve (3.3.9) for ϕ_t^n , then substitute which to the left-hand side of (3.3.6). With forward Euler time discretization, (3.3.6), (3.3.8) and (3.3.9) can be written as

$$\left(\frac{\phi^{n+1} - \phi^n}{\Delta t}, \zeta \right) = (\mathbf{u}, \nabla \zeta) + \sum_{e \in \Gamma_0} \int_e \hat{\mathbf{u}} \cdot \mathbf{n}_e[\zeta] ds + (f, \zeta), \quad (3.4.11)$$

$$\begin{aligned} \left(\frac{r^{n+1} - r^n}{\Delta t}, \xi \right) &= (\mathbf{u}c_f - \phi \mathbf{D} \nabla c_f, \nabla \xi) + \sum_{e \in \Gamma_0} \int_e \widehat{\mathbf{u}c_f} \cdot \mathbf{n}_e[\xi] ds \\ &\quad - \sum_{e \in \Gamma_0} \int_e \left(\{ \phi \mathbf{D}(\mathbf{u}) \nabla c_f \} \cdot \mathbf{n}_e[\xi] + \{ \phi \mathbf{D}(\mathbf{u}) \nabla \xi \} \cdot \mathbf{n}_e[c_f] + \frac{\tilde{\alpha}}{|e|} [c_f][\xi] \right) ds \\ &\quad + (f_I c_I - f_P c_f - B_1(\phi) c_f, \xi), \end{aligned} \quad (3.4.12)$$

$$\left(\frac{\phi^{n+1} - \phi^n}{\Delta t}, v \right) = (B_2(\phi) c_f, v), \quad (3.4.13)$$

with all the superscript n on the right hand sides being omitted for simplicity.

Because of the usage of consistent flux pair $\hat{\mathbf{u}}$ and $\widehat{\mathbf{u}c_f}$, we can get a ghost equation for r_2 by subtracting (3.4.12) from (3.4.11) and introducing ghost vari-

ables $c_2 = 1 - c_f, c_{2I} = 1 - c_I, r_2 = \phi c_2,$

$$\begin{aligned} \left(\frac{r_2^{n+1} - r_2^n}{\Delta t}, \xi \right) &= (\mathbf{u}c_2 - \phi \mathbf{D} \nabla c_2, \nabla \xi) + \sum_{e \in \Gamma_0} \int_e \widehat{\mathbf{u}}_{c_2} \cdot \mathbf{n}_e[\xi] ds \\ &\quad - \sum_{e \in \Gamma_0} \int_e \left(\{ \phi \mathbf{D}(\mathbf{u}) \nabla c_2 \} \cdot \mathbf{n}_e[\xi] + \{ \phi \mathbf{D}(\mathbf{u}) \nabla \xi \} \cdot \mathbf{n}_e[c_2] + \frac{\tilde{\alpha}}{|e|} [c_2][\xi] \right) ds \\ &\quad + (f_I c_{2I} - f_P c_2 + B_1(\phi) c_f, \xi). \end{aligned} \quad (3.4.14)$$

Therefore, though we solve (3.4.11) and (3.4.12) in the real computation, we analyze (3.4.12) and (3.4.14) instead of the former pair as the two forms are equivalent.

The second-order bound-preserving scheme is built and analyzed based on (3.4.12), (3.4.14) and (3.4.13).

In this paper, we use the quadrature rule of order k proposed in [71] to compute the integral in cells, and use the corresponding $k+1$ points Gaussian quadrature rule to evaluate integration on cell interfaces. The quadrature rule of order k has the following crucial properties:

- All of the quadrature points lie in the cells with positive weights,
- The quadrature points contains $k+1$ Gaussian quadrature points in each of its edges,
- It is exact for polynomials up to degree $2k-1$,

The distribution of quadrature points in the case of $k=1$ and $k=2$ are shown in Figure 3.1. We denote $x_{i,\beta}, \beta = 1, 2, \dots, k+1$, as the quadrature points on e_K^i with \tilde{w}_β being the corresponding weights, and denote $x_\gamma, \gamma = 1, 2, \dots, L$, as the

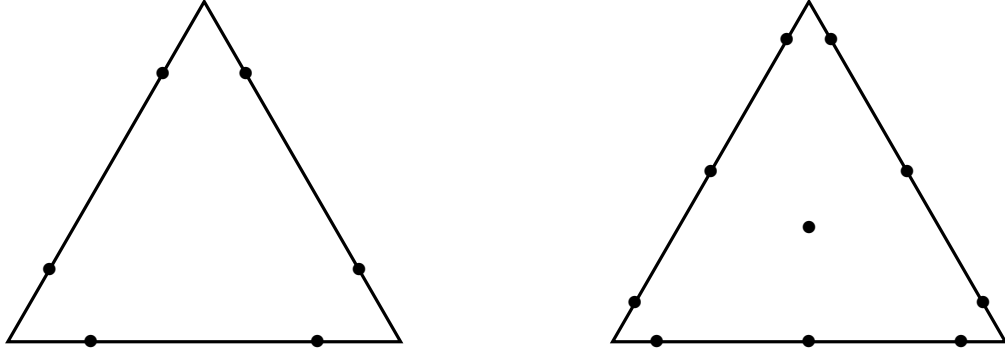


Figure 3.1: Distribution of quadrature points for $k = 1$ and $k = 2$.

quadrature points in cell K with $\hat{\omega}_\gamma$ being the corresponding weights. Moreover, We denote ω_β , $\beta = 1, 2, \dots, k + 1$, as the $k + 1$ Gaussian quadrature weights on the reference interval $[-\frac{1}{2}, \frac{1}{2}]$. Based on the above notations, we define the values of o ($o = r, c, \phi, p, \dots$) at the quadrature points as $o_K^{i,\beta} = o(x_{i,\beta})$ along the boundary of K and $o_K^\gamma = o(x_\gamma)$ in cell K .

In (3.3.9), we take $v = 1$ in K to obtain the equation satisfied by the cell average of ϕ :

$$\bar{\phi}_K^{n+1} = \bar{\phi}_K^n + \Delta t \overline{B_2(\phi^n)} c_f. \quad (3.4.15)$$

We will demonstrate how to preserve the monotonicity and the upper bound of $\bar{\phi}_K^{n+1}$ in the following lemma:

Lemma 3.4.1. *Given $0 \leq r^n \leq \phi^n$ ($0 \leq c_f^n \leq 1$) and $\phi^n < 1$, we have $\bar{\phi}_K^n \leq \bar{\phi}_K^{n+1} < 1$, if the time step satisfies*

$$\Delta t < B_{30}^{-1}, \quad (3.4.16)$$

where B_{30} is a constant defined as

$$B_{30} = \frac{\alpha a_0 k_s k_c}{\rho_s (1 - \phi^*) (k_s + k_c)}.$$

Proof. Define $B_3(x) = \frac{\alpha a_0 k_s k_c}{\rho_s (1 - \phi_0(x)) (k_s + k_c)}$. Then $B_3(x)$ is independent of time t and it's easy to check that $B_2(x, \phi) = B_3(x) \cdot (1 - \phi) \leq B_{30} \cdot (1 - \phi)$.

Applying quadrature rule in [71] to (3.4.15) with enough algebraic order k , we have

$$\begin{aligned} & \bar{\phi}_K^{n+1} \\ &= \bar{\phi}_K^n + \Delta t \overline{B_2(\phi^n) c_f} \\ &= \bar{\phi}_K^n + \Delta t \left(\sum_{i=1}^3 \sum_{\beta=1}^{k+1} \hat{w}_\beta (B_3)_K^{i,\beta} (1 - \phi_K^{i,\beta}) (c_f)_K^{i,\beta} + \sum_{\gamma=1}^L \hat{w}_\gamma (B_3)_K^\gamma (1 - \phi_K^\gamma) (c_f)_K^\gamma \right) \\ &\geq \bar{\phi}_K^n \end{aligned}$$

under the assumption $0 \leq c_f^n \leq 1$ and $\phi_K^n < 1$. Moreover, we have

$$\begin{aligned} & \bar{\phi}_K^{n+1} \\ &= \bar{\phi}_K^n + \Delta t \left(\sum_{i=1}^3 \sum_{\beta=1}^{k+1} \hat{w}_\beta (B_3)_K^{i,\beta} (1 - \phi_K^{i,\beta}) (c_f)_K^{i,\beta} + \sum_{\gamma=1}^L \hat{w}_\gamma (B_3)_K^\gamma (1 - \phi_K^\gamma) (c_f)_K^\gamma \right) \\ &\leq \bar{\phi}_K^n + \Delta t \left(\sum_{i=1}^3 \sum_{\beta=1}^{k+1} \hat{w}_\beta (B_3)_K^{i,\beta} (1 - \phi_K^{i,\beta}) + \sum_{\gamma=1}^L \hat{w}_\gamma (B_3)_K^\gamma (1 - \phi_K^\gamma) \right) \\ &\leq \bar{\phi}_K^n + \Delta t \left(\sum_{i=1}^3 \sum_{\beta=1}^{k+1} \hat{w}_\beta B_{30} (1 - \phi_K^{i,\beta}) + \sum_{\gamma=1}^L \hat{w}_\gamma B_{30} (1 - \phi_K^\gamma) \right) \\ &= \bar{\phi}_K^n + \Delta t B_{30} \left(\sum_{i=1}^3 \sum_{\beta=1}^{k+1} \hat{w}_\beta (1 - \phi_K^{i,\beta}) + \sum_{\gamma=1}^L \hat{w}_\gamma (1 - \phi_K^\gamma) \right) \\ &= \bar{\phi}_K^n + \Delta t B_{30} (1 - \bar{\phi}_K^n). \end{aligned}$$

Thus $\bar{\phi}_K^{n+1} < 1$ under the condition (3.4.16). \square

The bound-preserving property for \bar{r}_K^{n+1} is relatively difficult to derive. Therefore, instead of obtaining $0 \leq \bar{r}_K^{n+1} \leq \bar{\phi}_K^{n+1}$ directly, we apply the positivity-preserving technique to \bar{r}_K^{n+1} and \bar{r}_{2K}^{n+1} in (3.4.12) and (3.4.14), respectively, which further yields $0 \leq \bar{r}_K^{n+1} \leq \bar{\phi}_K^{n+1}$ due to the fact that $\bar{r}_K^{n+1} + \bar{r}_{2K}^{n+1} = \bar{\phi}_K^{n+1}$. To construct the positivity-preserving technique, in (3.4.12), we take $\xi = 1$ in K to obtain the equation satisfied by the cell average of r

$$\bar{r}_K^{n+1} = H_K^c(r, c_f, \mathbf{u}) + H_K^d(r, c_f, \mathbf{u}, \phi) + H_K^s(r, c_f, c_I, f_P, f_I, \phi), \quad (3.4.17)$$

where

$$H_K^c(r, c_f, \mathbf{u}) = \frac{1}{3} \bar{r}_K^n - \lambda \sum_{i=1}^3 \int_{e_K^i} \widehat{\mathbf{u}c_f} \cdot \boldsymbol{\nu}_K^i ds, \quad (3.4.18)$$

$$H_K^d(r, c_f, \mathbf{u}, \phi) = \frac{1}{3} \bar{r}_K^n + \lambda \sum_{i=1}^3 \int_{e_K^i} \left(\{ \mathbf{D}(\mathbf{u}) \nabla c \} \cdot \boldsymbol{\nu}_K^i + \frac{\tilde{\alpha}}{\ell_K^i} [c] \mathbf{n}_e \cdot \boldsymbol{\nu}_K^i \right) ds, \quad (3.4.19)$$

$$H_K^s(r, c_f, c_I, f_P, f_I, \phi) = \frac{1}{3} \bar{r}_K^n + \Delta t \overline{f_{IC_I} - f_{PC_f} - B_1(\phi)c_f}, \quad (3.4.20)$$

with $\lambda = \frac{\Delta t}{|K|}$ being the ratio of time step and area of triangular element K , and $\overline{f_{PC_f} - f_{IC_I} - B_1(\phi)c_f}$ being the cell average of $f_{PC_f} - f_{IC_I} - B_1(\phi)c_f$ on K .

We will demonstrate the positivity-preserving property for \bar{r}_K^{n+1} by collecting the following three lemmas.

Lemma 3.4.2. *Given $r^n > 0$ ($c_f^n > 0$), we have $H_K^s(r, c_f, c_I, f_P, f_I, \phi) > 0$, if the time step satisfies*

$$\Delta t \leq \frac{\phi_\star}{6f_{PM}}, \quad \Delta t \leq \frac{\phi_\star}{6B_1(\phi_\star)}, \quad (3.4.21)$$

where

$$f_{PM} = \max_{i,\beta,\gamma} \{(f_P)_K^{i,\beta}, (f_P)_K^\gamma\}.$$

Proof. We can split (3.4.20) as

$$H_K^s = \Delta t \overline{f_{IC_I}} + \left(\frac{1}{6} \bar{r}_K^n - \Delta t \overline{f_{PC_f}} \right) + \left(\frac{1}{6} \bar{r}_K^n - \Delta t \overline{B_1(\phi)c_f} \right) := L_1 + L_2 + L_3.$$

It is easy to check that $L_1 = \Delta t \overline{f_{IC_I}} \geq 0$. We only need to consider L_2 and L_3 .

Applying quadrature rule in [71] with enough algebraic order k to L_2 and L_3 , respectively, we can get

$$\begin{aligned} L_2 &= \frac{1}{6} \bar{r}_K^n - \Delta t \overline{f_{PC_f}} \\ &= \frac{1}{6} \left(\sum_{i=1}^3 \sum_{\beta=1}^{k+1} \tilde{w}_\beta r_K^{i,\beta} + \sum_{\gamma=1}^L \hat{w}_\gamma r_K^\gamma \right) \\ &\quad - \Delta t \left(\sum_{i=1}^3 \sum_{\beta=1}^{k+1} \tilde{w}_\beta (f_P)_K^{i,\beta} (c_f)_K^{i,\beta} + \sum_{\gamma=1}^L \hat{w}_\gamma (f_P)_K^\gamma (c_f)_K^\gamma \right) \\ &= \sum_{i=1}^3 \sum_{\beta=1}^{k+1} \tilde{w}_\beta \left(\frac{1}{6} r_K^{i,\beta} - \Delta t (f_P)_K^{i,\beta} (c_f)_K^{i,\beta} \right) + \sum_{\gamma=1}^L \hat{w}_\gamma \left(\frac{1}{6} r_K^\gamma - \Delta t (f_P)_K^\gamma (c_f)_K^\gamma \right) \\ &\geq \sum_{i=1}^3 \sum_{\beta=1}^{k+1} \tilde{w}_\beta \left(\frac{1}{6} r_K^{i,\beta} - \Delta t (f_P)_K^{i,\beta} r_K^{i,\beta} \phi_\star^{-1} \right) + \sum_{\gamma=1}^L \hat{w}_\gamma \left(\frac{1}{6} r_K^\gamma - \Delta t (f_P)_K^\gamma r_K^\gamma \phi_\star^{-1} \right) \\ &= \sum_{i=1}^3 \sum_{\beta=1}^{k+1} \tilde{w}_\beta \left(\frac{1}{6} - \Delta t (f_P)_K^{i,\beta} \phi_\star^{-1} \right) r_K^{i,\beta} + \sum_{\gamma=1}^L \hat{w}_\gamma \left(\frac{1}{6} - \Delta t (f_P)_K^\gamma \phi_\star^{-1} \right) r_K^\gamma. \end{aligned}$$

Thus $L_2 > 0$ under the condition (3.4.21).

$$\begin{aligned}
L_3 &= \frac{1}{6} \bar{r}_K^n - \Delta t \overline{B_1(\phi) c_f} \\
&= \frac{1}{6} \left(\sum_{i=1}^3 \sum_{\beta=1}^{k+1} \tilde{w}_\beta r_K^{i,\beta} + \sum_{\gamma=1}^L \hat{w}_\gamma r_K^\gamma \right) \\
&\quad - \Delta t \left(\sum_{i=1}^3 \sum_{\beta=1}^{k+1} \tilde{w}_\beta B_1(\phi_K^{i,\beta})(c_f)_K^{i,\beta} + \sum_{\gamma=1}^L \hat{w}_\gamma B_1(\phi_K^\gamma)(c_f)_K^\gamma \right) \\
&= \sum_{i=1}^3 \sum_{\beta=1}^{k+1} \tilde{w}_\beta \left(\frac{1}{6} r_K^{i,\beta} - \Delta t B_1(\phi_K^{i,\beta})(c_f)_K^{i,\beta} \right) \\
&\quad + \sum_{\gamma=1}^L \hat{w}_\gamma \left(\frac{1}{6} r_K^\gamma - \Delta t B_1(\phi_K^\gamma)(c_f)_K^\gamma \right) \\
&\geq \sum_{i=1}^3 \sum_{\beta=1}^{k+1} \tilde{w}_\beta \left(\frac{1}{6} r_K^{i,\beta} - \Delta t B_1(\phi_K^{i,\beta}) r_K^{i,\beta} \phi_\star^{-1} \right) \\
&\quad + \sum_{\gamma=1}^L \hat{w}_\gamma \left(\frac{1}{6} r_K^\gamma - \Delta t B_1(\phi_K^\gamma) r_K^\gamma \phi_\star^{-1} \right) \\
&= \sum_{i=1}^3 \sum_{\beta=1}^{k+1} \tilde{w}_\beta \left(\frac{1}{6} - \Delta t B_1(\phi_K^{i,\beta}) \phi_\star^{-1} \right) r_K^{i,\beta} \\
&\quad + \sum_{\gamma=1}^L \hat{w}_\gamma \left(\frac{1}{6} - \Delta t B_1(\phi_K^\gamma) \phi_\star^{-1} \right) r_K^\gamma \\
&\geq \sum_{i=1}^3 \sum_{\beta=1}^{k+1} \tilde{w}_\beta \left(\frac{1}{6} - \Delta t B_1(\phi_\star) \phi_\star^{-1} \right) r_K^{i,\beta} \\
&\quad + \sum_{\gamma=1}^L \hat{w}_\gamma \left(\frac{1}{6} - \Delta t B_1(\phi_\star) \phi_\star^{-1} \right) r_K^\gamma.
\end{aligned}$$

Thus $L_3 > 0$ under the condition (3.4.21). To sum up, $H_K^s(r, c_f, c_I, f_P, f_I, \phi) = L_1 + L_2 + L_3 > 0$ under the condition (3.4.21). \square

In the following two lemmas, we only consider second order scheme, i.e. we

use P^1 element, and apply quadrature rule in [71] with $k = 1$ in cell and the corresponding 2 point Gaussian quadrature rule on cell interface. Note that in this case, $\hat{\omega}_\beta = \frac{1}{3}\omega_\beta$.

Lemma 3.4.3. *Given $r^n > 0$ ($c_f^n > 0$), we have $H_K^c(r, c_f, \mathbf{u}) > 0$, if α satisfies*

$$\alpha \geq \max_{i,\beta,K} \{|\mathbf{u}_K^{i,\beta}|\}, \quad (3.4.22)$$

and the time step satisfies

$$\Delta t \leq \min_{i,\beta,m} \left\{ \frac{|K|\phi(V_m)}{9\ell_K^i(|\mathbf{u}_K^{i,\beta}| + \alpha)} \right\}, \quad (3.4.23)$$

where $\phi(V_m)$, $m = 1, 2, 3$ are the values of ϕ at vertices $V_m \in K$ at time level n .

Proof. Applying quadrature rule for $k = 1$, we can rewrite (3.4.18) as

$$\begin{aligned}
H_K^c &= \frac{1}{3} r_K^n - \lambda \sum_{i=1}^3 \int_{e_K^i} \widehat{\mathbf{u}}_{c_f} \cdot \boldsymbol{\nu}_K^i ds \\
&= \frac{1}{9} \left(\sum_{i=1}^3 \sum_{\beta=1}^2 w_\beta r_K^{i,\beta} \right) - \lambda \sum_{i=1}^3 \sum_{\beta=1}^2 w_\beta \ell_K^i (\widehat{\mathbf{u}}_{c_f})_{e_K^i}^{i,\beta} \cdot \boldsymbol{\nu}_K^i \\
&= \sum_{i=1}^3 \sum_{\beta=1}^2 w_\beta \left(\frac{1}{9} r_K^{i,\beta} - \lambda \ell_K^i (\widehat{\mathbf{u}}_{c_f})_{e_K^i}^{i,\beta} \cdot \boldsymbol{\nu}_K^i \right) \\
&= \sum_{i=1}^3 \sum_{\beta=1}^2 w_\beta \left(\frac{1}{9} r_K^{i,\beta} - \lambda \ell_K^i \left(\frac{1}{2} \mathbf{u}_{K_i}^{i,\beta} \cdot \boldsymbol{\nu}_K^i (c_f)_{K_i}^{i,\beta} + \frac{1}{2} \mathbf{u}_K^{i,\beta} \cdot \boldsymbol{\nu}_K^i (c_f)_K^{i,\beta} \right. \right. \\
&\quad \left. \left. - \alpha (c_f)_{K_i}^{i,\beta} + \alpha (c_f)_K^{i,\beta} \right) \right) \\
&= \sum_{i=1}^3 \sum_{\beta=1}^2 w_\beta \left\{ \left(\frac{1}{18} r_K^{i,\beta} - \frac{1}{2} \lambda \ell_K^i \left(\mathbf{u}_{K_i}^{i,\beta} \cdot \boldsymbol{\nu}_K^i (c_f)_{K_i}^{i,\beta} - \alpha (c_f)_{K_i}^{i,\beta} + \alpha (c_f)_K^{i,\beta} \right) \right) \right. \\
&\quad \left. + \left(\frac{1}{18} r_K^{i,\beta} - \frac{1}{2} \lambda \ell_K^i \left(\mathbf{u}_K^{i,\beta} \cdot \boldsymbol{\nu}_K^i (c_f)_K^{i,\beta} - \alpha (c_f)_{K_i}^{i,\beta} + \alpha (c_f)_K^{i,\beta} \right) \right) \right\} \\
&:= \sum_{i=1}^3 \sum_{\beta=1}^2 w_\beta \left(L_1^{i,\beta} + L_2^{i,\beta} \right).
\end{aligned}$$

Since c_f and r are both approximated by linear functions, they can be represented as a linear combination of their values on three vertices $\{V_1, V_2, V_3\}$ of K , i.e. for any point $x_K^\rho \in K$,

$$(c_f)_K^\rho = \sum_{m=1}^3 \mu_m^\rho c_f(V_m), \quad r_K^\rho = \sum_{m=1}^3 \mu_m^\rho r(V_m) = \sum_{m=1}^3 \mu_m^\rho \phi(V_m) c_f(V_m), \quad (3.4.24)$$

where $0 \leq \mu_1^\rho, \mu_2^\rho, \mu_3^\rho \leq 1$ and $\mu_1^\rho + \mu_2^\rho + \mu_3^\rho = 1$ are the barycentric coordinates

of x_K^ρ in K . Then we have

$$\begin{aligned}
L_1^{i,\beta} &= \frac{1}{18}r_K^{i,\beta} - \frac{1}{2}\lambda\ell_K^i \left(\mathbf{u}_{K_i}^{i,\beta} \cdot \boldsymbol{\nu}_K^i (c_f)_{K_i}^{i,\beta} - \alpha(c_f)_{K_i}^{i,\beta} + \alpha(c_f)_K^{i,\beta} \right) \\
&= \sum_{m=1}^3 \frac{1}{18}\mu_m^{i,\beta} \phi(V_m) c_f(V_m) \\
&\quad - \frac{1}{2}\lambda\ell_K^i \left(\mathbf{u}_{K_i}^{i,\beta} \cdot \boldsymbol{\nu}_K^i (c_f)_{K_i}^{i,\beta} - \alpha(c_f)_{K_i}^{i,\beta} + \alpha \sum_{m=1}^3 \mu_m^{i,\beta} c_f(V_m) \right) \\
&= \sum_{m=1}^3 \mu_m^{i,\beta} \left(\frac{1}{18}\phi(V_m) - \frac{1}{2}\lambda\ell_K^i \alpha \right) c_f(V_m) + \frac{1}{2}\lambda\ell_K^i (\alpha - \mathbf{u}_{K_i}^{i,\beta} \cdot \boldsymbol{\nu}_K^i) (c_f)_{K_i}^{i,\beta},
\end{aligned}$$

and

$$\begin{aligned}
L_2^{i,\beta} &= \frac{1}{18}r_K^{i,\beta} - \frac{1}{2}\lambda\ell_K^i \left(\mathbf{u}_K^{i,\beta} \cdot \boldsymbol{\nu}_K^i (c_f)_K^{i,\beta} - \alpha(c_f)_{K_i}^{i,\beta} + \alpha(c_f)_K^{i,\beta} \right) \\
&= \sum_{m=1}^3 \frac{1}{18}\mu_m^{i,\beta} \phi(V_m) c_f(V_m) \\
&\quad - \frac{1}{2}\lambda\ell_K^i \left(\mathbf{u}_K^{i,\beta} \cdot \boldsymbol{\nu}_K^i \sum_{m=1}^3 \mu_m^{i,\beta} c_f(V_m) - \alpha(c_f)_{K_i}^{i,\beta} + \alpha \sum_{m=1}^3 \mu_m^{i,\beta} c_f(V_m) \right) \\
&= \sum_{m=1}^3 \mu_m^{i,\beta} \left(\frac{1}{18}\phi(V_m) - \frac{1}{2}\lambda\ell_K^i (\mathbf{u}_K^{i,\beta} \cdot \boldsymbol{\nu}_K^i + \alpha) \right) c_f(V_m) + \frac{1}{2}\lambda\ell_K^i \alpha (c_f)_{K_i}^{i,\beta}.
\end{aligned}$$

Therefore $L_1^{i,\beta}, L_2^{i,\beta} > 0$ under the conditions (3.4.22) and (3.4.23), respectively, which further yields $H_K^c > 0$. \square

Lemma 3.4.4. *Given $r^n > 0$ ($c_f^n > 0$), we have $H_K^d(r, c_f, \mathbf{u}, \phi) > 0$, if $\tilde{\alpha}$ satisfies*

$$\tilde{\alpha} \geq \frac{(3 + \sqrt{3})\Lambda}{2\rho}, \tag{3.4.25}$$

and the time step satisfies

$$\Delta t \leq \min_m \left\{ \frac{|K|\phi(V_m)}{18\tilde{\alpha}} \right\}, \quad \Delta t \leq \min_m \left\{ \frac{\rho|K|\phi(V_m)}{27(3 + \sqrt{3})\Lambda} \right\}, \tag{3.4.26}$$

where $\phi(V_m)$, $m = 1, 2, 3$ are the values of ϕ at the vertices $V_m \in K$ at time level n , ρ is the parameter used in the definition of regularity of Ω_h , and Λ is the largest spectral radius of \mathbf{D} in K 's.

Proof. For the diffusion part

$$H_K^d(r, c_f, \mathbf{u}, \phi) = \frac{1}{3} \bar{r}_K^n + \lambda \sum_{i=1}^3 \int_{e_K^i} \left(\{\mathbf{D}(\mathbf{u}) \nabla c_f\} \cdot \boldsymbol{\nu}_K^i + \frac{\tilde{\alpha}}{\ell_K^i} [c_f] \mathbf{n}_e \cdot \boldsymbol{\nu}_K^i \right) ds.$$

Since \mathbf{D} is symmetric, following [54], we can rewrite the diffusion term as a directional derivative

$$\mathbf{D} \nabla c_f \cdot \boldsymbol{\nu}_K^i = \mathbf{D} \boldsymbol{\nu}_K^i \cdot \nabla c_f = S^i \frac{\partial c_f}{\partial \mathbf{l}^i},$$

where $S^i = \|\mathbf{D} \boldsymbol{\nu}_K^i\| \leq \Lambda$ and $\mathbf{l}^i = \mathbf{D} \boldsymbol{\nu}_K^i / \|\mathbf{D} \boldsymbol{\nu}_K^i\|$. Define $S_K^i = S^i|_K$, $S_{K_i}^i = S^i|_{K_i}$ and $\mathbf{l}_K^i = \mathbf{l}^i|_K$, $\mathbf{l}_{K_i}^i = \mathbf{l}^i|_{K_i}$. For each quadrature point $x_K^{i,\beta}$ on the edge e_K^i , we can find the corresponding points $\tilde{x}_K^{i,\beta} \in \partial K$ and $\tilde{x}_{K_i}^{i,\beta} \in \partial K_i$ such that $\overrightarrow{\tilde{x}_K^{i,\beta} x_K^{i,\beta}}$ and $\overrightarrow{x_K^{i,\beta} \tilde{x}_{K_i}^{i,\beta}}$ are the same direction with \mathbf{l}_K^i and $\mathbf{l}_{K_i}^i$, respectively. See Figure 3.2 for an illustration. At the quadrature point $x = x_K^{i,\beta}$, we have

$$\begin{aligned} \{\mathbf{D}(\mathbf{u}) \nabla c_f\}_{e_K^i}^{i,\beta} \cdot \boldsymbol{\nu}_K^i &= \frac{1}{2} \mathbf{D}(\mathbf{u}_{K_i}^{i,\beta}) \nabla (c_f)_{K_i}^{i,\beta} \cdot \boldsymbol{\nu}_K^i + \frac{1}{2} \mathbf{D}(\mathbf{u}_K^{i,\beta}) \nabla (c_f)_K^{i,\beta} \cdot \boldsymbol{\nu}_K^i \\ &= \frac{1}{2} S_K^{i,\beta} \frac{(c_f)_K^{i,\beta} - c_f(\tilde{x}_K^{i,\beta})}{\|x_K^{i,\beta} - \tilde{x}_K^{i,\beta}\|} + \frac{1}{2} S_{K_i}^{i,\beta} \frac{c_f(\tilde{x}_{K_i}^{i,\beta}) - (c_f)_{K_i}^{i,\beta}}{\|\tilde{x}_{K_i}^{i,\beta} - x_K^{i,\beta}\|} \\ &= \frac{S_K^{i,\beta}}{2 \|x_K^{i,\beta} - \tilde{x}_K^{i,\beta}\|} (c_f)_K^{i,\beta} - \frac{S_{K_i}^{i,\beta}}{2 \|\tilde{x}_{K_i}^{i,\beta} - x_K^{i,\beta}\|} (c_f)_{K_i}^{i,\beta} \\ &\quad - \frac{S_K^{i,\beta}}{2 \|x_K^{i,\beta} - \tilde{x}_K^{i,\beta}\|} c_f(\tilde{x}_K^{i,\beta}) + \frac{S_{K_i}^{i,\beta}}{2 \|\tilde{x}_{K_i}^{i,\beta} - x_K^{i,\beta}\|} c_f(\tilde{x}_{K_i}^{i,\beta}). \end{aligned}$$

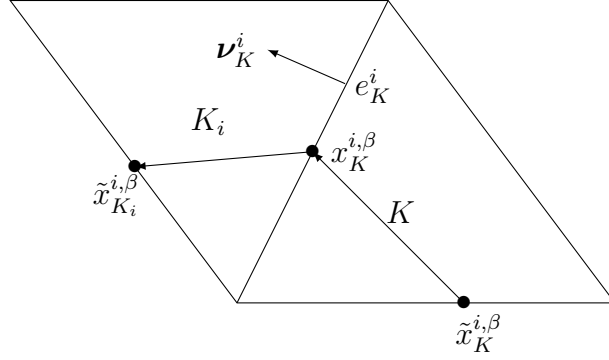


Figure 3.2: The points chosen to evaluate directional derivative in the diffusion part.

Therefore, we can rewrite $H_K^d(r, c_f, \mathbf{u}, \phi)$ as

$$\begin{aligned}
H_K^d &= \frac{1}{6}\bar{r}_K^n + \frac{1}{6}\bar{r}_K^n + \lambda \sum_{i=1}^3 \int_{e_K^i} \left(\{\mathbf{D}(\mathbf{u})\nabla c_f\} \cdot \boldsymbol{\nu}_K^i + \frac{\tilde{\alpha}}{\ell_K^i} [c_f] \mathbf{n}_e \cdot \boldsymbol{\nu}_K^i \right) ds \\
&= \frac{1}{6}\bar{r}_K^n + \frac{1}{18} \sum_{i=1}^3 \sum_{\beta=1}^2 w_\beta r_K^{i,\beta} \\
&\quad + \lambda \sum_{i=1}^3 \sum_{\beta=1}^2 w_\beta \ell_K^i \left(\{\mathbf{D}(\mathbf{u})\nabla c_f\}_{e_K^i}^{i,\beta} \cdot \boldsymbol{\nu}_K^i + \frac{\tilde{\alpha}}{\ell_K^i} (c_f)_{K_i}^{i,\beta} - \frac{\tilde{\alpha}}{\ell_K^i} (c_f)_K^{i,\beta} \right) \\
&= \frac{1}{6}\bar{r}_K^n + \frac{1}{18} \sum_{i=1}^3 \sum_{\beta=1}^2 \sum_{m=1}^3 w_\beta \mu_m^{i,\beta} \phi(V_m) c_f(V_m) \\
&\quad + \lambda \sum_{i=1}^3 \sum_{\beta=1}^2 w_\beta \ell_K^i \left(\{\mathbf{D}(\mathbf{u})\nabla c_f\}_{e_K^i}^{i,\beta} \cdot \boldsymbol{\nu}_K^i + \frac{\tilde{\alpha}}{\ell_K^i} (c_f)_{K_i}^{i,\beta} - \frac{\tilde{\alpha}}{\ell_K^i} (c_f)_K^{i,\beta} \right) \\
&= \frac{1}{6}\bar{r}_K^n + \sum_{i=1}^3 \sum_{\beta=1}^2 w_\beta \left(\frac{1}{18} \sum_{m=1}^3 \mu_m^{i,\beta} \phi(V_m) c_f(V_m) \right. \\
&\quad \left. + \lambda \ell_K^i \left(\{\mathbf{D}(\mathbf{u})\nabla c_f\}_{e_K^i}^{i,\beta} \cdot \boldsymbol{\nu}_K^i + \frac{\tilde{\alpha}}{\ell_K^i} (c_f)_{K_i}^{i,\beta} - \frac{\tilde{\alpha}}{\ell_K^i} (c_f)_K^{i,\beta} \right) \right) \\
&:= \sum_{i=1}^3 \sum_{\beta=1}^2 w_\beta L_1^{i,\beta} + L_2,
\end{aligned}$$

where

$$\begin{aligned}
L_1^{i,\beta} &= \frac{1}{18} \sum_{m=1}^3 \mu_m^{i,\beta} \phi(V_m) c_f(V_m) + \lambda \ell_K^i \left[\left(\frac{S_K^{i,\beta}}{2\|x_K^{i,\beta} - \tilde{x}_K^{i,\beta}\|} - \frac{\tilde{\alpha}}{\ell_K^i} \right) (c_f)_{K_i}^{i,\beta} \right. \\
&\quad \left. + \left(\frac{\tilde{\alpha}}{\ell_K^i} - \frac{S_{K_i}^{i,\beta}}{2\|\tilde{x}_{K_i}^{i,\beta} - x_{K_i}^{i,\beta}\|} \right) (c_f)_{K_i}^{i,\beta} + \frac{S_{K_i}^{i,\beta}}{2\|\tilde{x}_{K_i}^{i,\beta} - x_{K_i}^{i,\beta}\|} c_f(\tilde{x}_{K_i}^{i,\beta}) \right], \\
L_2 &= \frac{1}{6} \bar{r}_K^n - \lambda \ell_K^i \sum_{i=1}^3 \sum_{\beta=1}^2 \frac{\omega_\beta S_K^{i,\beta}}{2\|x_K^{i,\beta} - \tilde{x}_K^{i,\beta}\|} c_f(\tilde{x}_K^{i,\beta}).
\end{aligned}$$

We need to make $L_1^{i,\beta}, L_2 > 0$. In fact

$$\begin{aligned}
L_1^{i,\beta} &= \frac{1}{18} \sum_{m=1}^3 \mu_m^{i,\beta} \phi(V_m) c_f(V_m) + \lambda \ell_K^i \left(\frac{S_K^{i,\beta}}{2\|x_K^{i,\beta} - \tilde{x}_K^{i,\beta}\|} - \frac{\tilde{\alpha}}{\ell_K^i} \right) (c_f)_{K_i}^{i,\beta} \\
&\quad + \lambda \ell_K^i \left(\frac{\tilde{\alpha}}{\ell_K^i} - \frac{S_{K_i}^{i,\beta}}{2\|\tilde{x}_{K_i}^{i,\beta} - x_{K_i}^{i,\beta}\|} \right) (c_f)_{K_i}^{i,\beta} + \lambda \ell_K^i \frac{S_{K_i}^{i,\beta}}{2\|\tilde{x}_{K_i}^{i,\beta} - x_{K_i}^{i,\beta}\|} c_f(\tilde{x}_{K_i}^{i,\beta}) \\
&= \sum_{m=1}^3 \mu_m^{i,\beta} \left(\frac{1}{18} \phi(V_m) + \lambda \ell_K^i \left(\frac{S_K^{i,\beta}}{2\|x_K^{i,\beta} - \tilde{x}_K^{i,\beta}\|} - \frac{\tilde{\alpha}}{\ell_K^i} \right) \right) c_f(V_m) \\
&\quad + \lambda \ell_K^i \left(\frac{\tilde{\alpha}}{\ell_K^i} - \frac{S_{K_i}^{i,\beta}}{2\|\tilde{x}_{K_i}^{i,\beta} - x_{K_i}^{i,\beta}\|} \right) (c_f)_{K_i}^{i,\beta} + \lambda \ell_K^i \frac{S_{K_i}^{i,\beta}}{2\|\tilde{x}_{K_i}^{i,\beta} - x_{K_i}^{i,\beta}\|} c_f(\tilde{x}_{K_i}^{i,\beta}).
\end{aligned}$$

Since $S_K^{i,\beta}, S_{K_i}^{i,\beta} \leq \Lambda$, to make $L_1^{i,\beta} > 0$, we need

$$\tilde{\alpha} \geq \frac{\ell_K^i \Lambda}{2\|\tilde{x}_{K_i}^{i,\beta} - x_{K_i}^{i,\beta}\|}, \quad \lambda \ell_K^i \left(\frac{\tilde{\alpha}}{\ell_K^i} - \frac{S_K^{i,\beta}}{2\|x_K^{i,\beta} - \tilde{x}_K^{i,\beta}\|} \right) \leq \frac{1}{18} \phi(V_m).$$

It's easy to compute that

$$\frac{\ell_K^i}{\|\tilde{x}_{K_i}^{i,\beta} - x_{K_i}^{i,\beta}\|} \leq \frac{3 + \sqrt{3}}{\min_j \sin(\theta_{K_i}^j)},$$

where the $\theta_{K_i}^j$ is the angle in triangle K_i which is opposite to the edge $e_{K_i}^j$. From

Figure 3.3 and regularity assumption of Ω_h , for all angle $\theta_{K_i}^j$ in K_i , we have

$$\sin \theta_{K_i}^j = \frac{h}{s} \geq \frac{\text{diam}(B_{K_i})}{\text{diam}(K_i)} \geq \rho.$$

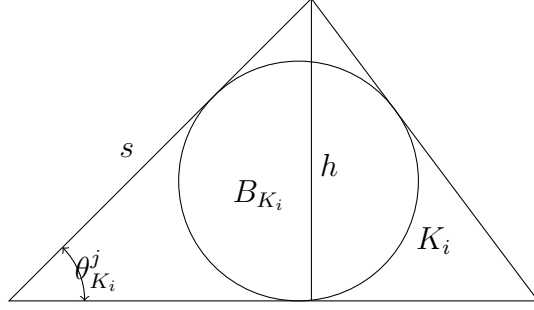


Figure 3.3: Triangle K_i and its sine

Therefore $L_1^{i,\beta} > 0$ under the conditions (3.4.25) and (3.4.26). As for L_2 , similar to (3.4.24), we write

$$c_f(\tilde{x}_K^{i,\beta}) = \sum_{m=1}^3 \tilde{\mu}_m^{i,\beta} c_f(V_m),$$

with $0 \leq \tilde{\mu}_m^{i,\beta} \leq 1$ and $\sum_{m=1}^3 \tilde{\mu}_m^{i,\beta} = 1$, and use the fact that $\omega_1 = \omega_2 = \frac{1}{2}$. Then

$$\begin{aligned} L_2 &= \frac{1}{6} \bar{r}_K^n - \lambda \ell_K^i \sum_{i=1}^3 \sum_{\beta=1}^2 \sum_{m=1}^3 \frac{\tilde{\mu}_m^{i,\beta} S_K^{i,\beta}}{4 \|x_K^{i,\beta} - \tilde{x}_K^{i,\beta}\|} c_f(V_m) \\ &= \sum_{m=1}^3 \left(\frac{1}{18} \phi(V_m) - \lambda \ell_K^i \sum_{i=1}^3 \sum_{\beta=1}^2 \frac{\tilde{\mu}_m^{i,\beta} S_K^{i,\beta}}{4 \|x_K^{i,\beta} - \tilde{x}_K^{i,\beta}\|} \right) c_f(V_m) \\ &\geq \sum_{m=1}^3 \left(\frac{1}{18} \phi(V_m) - \lambda \sum_{i=1}^3 \sum_{\beta=1}^2 \frac{(3 + \sqrt{3})\Lambda}{4\rho} \right) c_f(V_m) \\ &= \sum_{m=1}^3 \left(\frac{1}{18} \phi(V_m) - \lambda \frac{3(3 + \sqrt{3})\Lambda}{2\rho} \right) c_f(V_m). \end{aligned}$$

Thus, $L_2 > 0$ under the condition (3.4.26). Therefore we have $H_K^d(r, c_f, \mathbf{u}, \phi) > 0$ under the conditions (3.4.25) and (3.4.26). \square

Collecting the three lemmas above, we have the following Lemma:

Lemma 3.4.5. *Given $r^n > 0$, and the parameters α and $\tilde{\alpha}$ satisfy (3.4.22) and (3.4.25), respectively. Then $\bar{r}_K^{n+1} > 0$ under the conditions (3.4.21), (3.4.23) and (3.4.26).*

Compare the equation (3.4.14) with (3.4.12), we can observe that the equation for r_2 is almost the same as that for r , except that its source term contains a positive term $+B_1(\phi)c_f$ instead of $-B_1(\phi)c_f$, which will benefit its positivity. Therefore, we can get a similar lemma for r_2 :

Lemma 3.4.6. *Given $r_2^n > 0$, and the parameters α and $\tilde{\alpha}$ satisfy (3.4.22) and (3.4.25), respectively. Then $\bar{r}_{2K}^{n+1} > 0$ under the conditions (3.4.21), (3.4.23) and (3.4.26).*

Combine Lemmas 3.4.5, 3.4.6 and 3.4.1, and use the fact that $r^n + r_2^n = \phi^n, r^{n+1} + r_2^{n+1} = \phi^{n+1}$, we finally reach our main theorem:

Theorem 3.4.7. *Given $0 \leq r^n \leq \phi^n < 1$, if we chose consistent flux pair $\hat{\mathbf{u}}, \widehat{\mathbf{u}}c_f$ and the penalty parameters α and $\tilde{\alpha}$ satisfying (3.4.22) and (3.4.25), respectively, then $\bar{\phi}_K^n \leq \bar{\phi}_K^{n+1} < 1$ and $0 \leq \bar{r}_K^{n+1} \leq \bar{\phi}_K^{n+1}$ under the conditions (3.4.16), (3.4.21), (3.4.23), and (3.4.26).*

3.5 Bound-preserving technique for high-order schemes

In this section, we proceed to discuss the high-order bound-preserving technique.

3.5.1 High-order spatial discretization

In Lemmas 3.4.3 and 3.4.4, our proofs are based on P^1 elements. To attain high-order accuracy, we use P^k ($k > 2$) polynomials and apply the flux limiters following [25, 38].

We write (3.4.17) as

$$\bar{r}_K^{n+1} = \bar{r}_K^n + \lambda \sum_{i=1}^3 \left(\theta_{e_K^i} \hat{F}_{e_K^i} + (1 - \theta_{e_K^i}) \hat{f}_{e_K^i} \right) + \Delta t \bar{s},$$

where

$$\begin{aligned} \hat{F}_{e_K^i} &= - \int_{e_K^i} \widehat{\mathbf{u}c_f} \cdot \nu_i ds + \int_{e_K^i} \left(\{ \mathbf{D}(\mathbf{u}) \nabla c \} \cdot \nu_K^i + \frac{\tilde{\alpha}}{\ell_K^i} [c_f] \right) ds, \\ \bar{s} &= \overline{f_{IC_I} - f_{PC_f} - B_1(\phi)c_f} \end{aligned} \quad (3.5.27)$$

are high-order flux and source, respectively, and $\hat{f}_{e_K^i}$ is the second-order bound-preserving flux analyzed in Section 3.4. In lemma 3.4.2, we considered high-order approximations of the source term. Therefore, we only discuss the modification of the high-order fluxes in this section, which is implemented by choosing an appropriate parameter $\theta_{e_K^i}$. The cell average can be written as

$$\begin{aligned} \bar{r}_K^{n+1} &= \bar{r}_K^n + \lambda \sum_{i=1}^3 \hat{f}_{e_K^i} + \lambda \sum_{i=1}^3 \theta_{e_K^i} (\hat{F}_{e_K^i} - \hat{f}_{e_K^i}) + \Delta t \bar{s} \\ &= \bar{r}_L^{n+1} + \lambda \sum_{i=1}^3 \theta_{e_K^i} (\hat{F}_{e_K^i} - \hat{f}_{e_K^i}), \end{aligned}$$

where

$$\bar{r}_L^{n+1} = \bar{r}_K^n + \lambda \sum_{i=1}^3 \hat{f}_{e_K^i} + \Delta t \bar{s}$$

is the second-order cell average which was proved to be physically relevant if Δt is sufficiently small. To compute $\hat{f}_{e_K^i}$, we replace the high-order c_f in $\hat{F}_{e_K^i}$ by a

second-order approximation $\check{c}_f \in [0, 1]$. To construct it, we can simply apply the second-order L^2 projection to the high-order r^n to get \check{r}_K^n and high-order ϕ^n to get $\check{\phi}_K^n$, and then apply the limiter discussed at the end of this section with $k = 1$ to obtain $0 \leq \check{r}_K^n \leq \check{\phi}_K^n$. The \check{c}_f can be obtained as the linear interpolation of $\frac{\check{r}_K^n}{\check{\phi}_K^n}$ at the three vertices in each cell K . We choose the parameter $\theta_{e_K^i}$ as follows:

1. For any $K \in \Omega_h$, set $\beta_K^1, \beta_K^2 = 0$.
2. For each edge e_K^i , if $\hat{F}_{e_K^i} - \hat{f}_{e_K^i} \leq 0$, set $\beta_K^1 = \beta_K^1 + \hat{F}_{e_K^i} - \hat{f}_{e_K^i}$, otherwise set $\beta_K^2 = \beta_K^2 + \hat{F}_{e_K^i} - \hat{f}_{e_K^i}$.
3. Take $\theta_{K, e_K^i} = \min \left\{ -\frac{\bar{r}_L^{n+1}}{\lambda \beta_K^1}, \frac{\bar{\phi}_K^{n+1} - \bar{r}_L^{n+1}}{\lambda \beta_K^2}, 1 \right\}$.
4. For any $e \in \Gamma_0$, we can find $K_1, K_2 \in \Omega_h$ such that $K_1 \cap K_2 = e$. We take $\theta_e = \min\{\theta_{K_1, e_K^i}, \theta_{K_2, e_K^i}\}$.

The above algorithm can guarantee the monotone increasing and bound-preserving properties for the cell averages of ϕ and r : if $0 \leq r^n \leq \phi^n < 1$, then $\bar{\phi}_K^n \leq \bar{\phi}_K^{n+1} < 1$ and $0 \leq \bar{r}_K^{n+1} \leq \bar{\phi}_K^{n+1}$, under the appropriate penalty parameters $\alpha, \tilde{\alpha}$ and sufficiently small time step Δt . It remains to use proper slope limiter to modify ϕ_K^{n+1} and r_K^{n+1} such that $\phi_K^n \leq \phi_K^{n+1} < 1$ and $0 \leq r_K^{n+1} \leq \phi_K^{n+1}$ without loss of cell average and accuracy. As discussed in [22], the traditional slope limiter [50] cannot be applied since the bounds of ϕ_K^{n+1}, r_K^{n+1} are not constants but polynomials changing overtime. In this paper, we extend the limiter introduced in [57] and the algorithm can be described as follows: For polynomials $u(x), U(x) \in P^k(K)$ such that $0 \leq \bar{u} \leq \bar{U}$ and $U_* \leq U(x) \leq U^*$,

where U_*, U^* are two positive constants. We obtain a modified $\tilde{u}(x)$ in the following way:

1. Define $\hat{S} = \{x \in K : u(x) < 0\}$. Take

$$\hat{u} = u + \theta \left(\frac{\bar{u}}{\bar{U}} U - u \right), \quad \theta = \max_{y \in \hat{S}} \left\{ \frac{-u(y)\bar{U}}{\bar{u}U(y) - u(y)\bar{U}}, 0 \right\}. \quad (3.5.28)$$

2. Set $v = U - \hat{u}$, and repeat the above step for v to get \hat{v} .
3. Take $\tilde{u} = U - \hat{v}$ as the new approximation for $u(x)$.

This limiter is proved in [57] to have the following three necessary properties:

- boundness: $0 \leq \tilde{u}(x) \leq U(x), \forall x \in K$,
- average: $\int_K \tilde{u} dx = \int_K u dx$,
- accuracy: $\|u(x) - \tilde{u}(x)\|_\infty \leq Ch^{k+1}, h = \text{diam}(K)$.

We use such a slope limiter in the following way: To obtain $\phi_K^n \leq \tilde{\phi}_K^{n+1} < 1$, we take $u = \phi_K^{n+1} - \phi_K^n$ and $U = 1 - \phi_K^n$ in the limiter, and then $\tilde{\phi}_K^{n+1} = \tilde{u} + \phi_K^n$; To obtain $0 \leq \tilde{r}_K^{n+1} \leq \tilde{\phi}_K^{n+1}$, we take $u = r_K^{n+1}$ and $U = \tilde{\phi}_K^{n+1}$ to apply this limiter directly.

3.5.2 High-order time discretization

In the previous subsection, we only discussed the bound-preserving technique based on Euler forward time discretization. The technique can be extended to high-order time integrations that are convex combinations of Euler forwards. In

this paper, we use third-order strong stability preserving (SSP) time discretization to solve the ODE system $\mathbf{u}_t = \mathbf{L}(\mathbf{u})$:

$$\begin{aligned}\mathbf{u}^{(1)} &= \mathbf{u}^n + \Delta t \mathbf{L}(\mathbf{u}, t^n), \\ \mathbf{u}^{(2)} &= \frac{3}{4} \mathbf{u}^n + \frac{1}{4} (\mathbf{u}^{(1)} + \Delta t \mathbf{L}(\mathbf{u}^{(1)}, t^{n+1})), \\ \mathbf{u}^{n+1} &= \frac{1}{3} \mathbf{u}^n + \frac{2}{3} \left(\mathbf{u}^{(2)} + \Delta t \mathbf{L}(\mathbf{u}^{(2)}, t^n + \frac{\Delta t}{2}) \right).\end{aligned}$$

Another choice is to use third-order SSP multi-step method which is also a convex combination of forward Euler:

$$\mathbf{u}^{n+1} = \frac{16}{27} (\mathbf{u}^n + 3 \Delta t \mathbf{L}(\mathbf{u}^n, t^n)) + \frac{11}{27} (\mathbf{u}^{n-3} + \frac{12}{11} \Delta t \mathbf{L}(\mathbf{u}^{n-3}, t^{n-3})).$$

More details can be found in [20, 21, 30].

3.6 Numerical experiments

In this section, we provide numerical experiments to show the performance of the high-order bound-preserving DG scheme. In all the examples, we use third-order SSP Runge-Kutta discretization in time and P^2 element in space unless otherwise stated. To construct Ω_h , for simplicity, we first equally divide Ω into $N \times N$ rectangles and then obtain a uniform triangular mesh by equally dividing every rectangle into two. See Figure 3.4 for an illustration. However, the algorithms can be applied for any unstructured triangular meshes.

Example 3.6.1. *We first test the accuracy of the high-order bound-preserving DG scheme. Because of the restriction $0 \leq c_I \leq 1$, $f = f_I - f_P$ and $f_I, f_P \geq 0$ of*

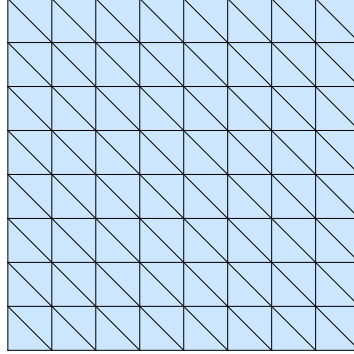


Figure 3.4: Triangular mesh ($N = 8$)

right hand side, it's difficult to find a suitable exact solution. Therefore, we use periodic boundary condition and predetermine the Darcy's velocity $u = (1, 1)^T$ in order to use spectral method to give a reference solution. Initial conditions are given as

$$c_f(x, y, 0) = 0.5 + 0.5 \cos(x) \cos(y) \quad \phi(x, y, 0) = 0.5 + 0.4 \sin(x) \sin(y),$$

The source functions are taken as

$$f_I = 2\phi_t, \quad f_P = -\phi_t, \quad c_I = 1,$$

where ϕ_t is obtained in the computation. The parameters are taken as:

$$D = 0.1\|\mathbf{u}\|, \quad k_c = k_s = a_0 = \frac{\alpha}{\rho_s} = 1$$

We use the uniform triangular meshes with $N = 4, 8, 16, 32, 64$, respectively, over the computational domain $\Omega = [0, 2\pi] \times [0, 2\pi]$ and set $\Delta t = 0.001h^2$ to reduce the time error. Moreover, the reference solution is obtained by spectral method on 64×64 equispaced grid points with fourth-order Runge–Kutta

time discretization. The computational results at $T = 0.01$ are shown in Table 3.1, illustrating the error and convergence order of c_f and ϕ , with and without bound-preserving technique respectively. From the table, we can observe optimal convergence rates. Therefore, the flux limiter and slope limiter do not degenerate the convergence order.

	c_f				ϕ			
	no limiter		with limiter		no limiter		with limiter	
N	L^2 error	order	L^2 error	order	L^2 error	order	L^2 error	order
4	2.90e-1	–	9.48e-2	–	1.82e-1	–	1.11e-1	–
8	2.46e-2	3.56	1.21e-2	2.97	2.50e-2	2.87	1.50e-2	2.89
16	1.83e-3	3.74	1.16e-3	3.39	3.18e-3	2.97	1.91e-3	2.97
32	1.40e-4	3.71	1.47e-4	2.98	4.00e-4	2.99	2.39e-4	2.99
64	1.29e-5	3.47	1.53e-5	3.26	5.00e-5	3.00	2.99e-5	3.00
N	L^∞ error	order	L^∞ error	order	L^∞ error	order	L^∞ error	order
4	1.52e-1	–	2.45e-2	–	5.01e-2	–	2.64e-2	–
8	7.08e-3	4.42	3.61e-3	2.76	7.11e-3	2.82	3.98e-3	2.73
16	7.21e-4	3.29	6.21e-4	2.54	9.16e-4	2.96	5.20e-4	2.94
32	1.12e-4	2.69	1.33e-4	2.22	1.15e-4	2.99	6.56e-5	2.99
64	9.66e-6	3.53	1.53e-5	3.13	1.44e-5	3.00	8.22e-6	3.00

Table 3.1: Example 3.6.1: Accuracy test for c_f and ϕ with and without bound-preserving technique.

Example 3.6.2. *We take the initial conditions with large gradients*

$$c_{f0} = \frac{\text{sign}(\sin(2x)\sin(2y)) + 1}{2}, \quad \phi_0 = 0.9 \frac{\text{sign}(\sin(x)\sin(y)) + 1}{2} + 0.05.$$

The source functions are taken as

$$f_I = (1 + \frac{\pi^2}{2}\bar{\phi}_t) \max\{\sin(2x)\sin(2y), 0\}, \quad f_P = -\min\{\sin(2x)\sin(2y), 0\}, \quad c_I = 0,$$

where $\bar{\phi}_t$ is the average of ϕ_t over the whole computational domain. Other parameters are chosen as

$$\mu = k_0 = k_s = k_c = 1, \quad a_0 = 0.5, \quad D(\mathbf{u}) = 0.01.$$

This example is used to demonstrate the necessity of the bound-preserving technique. The simulation will blow up without the technique due to the negativity of ϕ in some region while the bound-preserving scheme performs well.

We take $N = 40$ over the computational domain $\Omega = [0, 2\pi] \times [0, 2\pi]$. Moreover, we use P^1 element in this example since it is more suitable to demonstrate the stability than higher order ones, and set the time-step as small as $\Delta t = 0.001h^2$. All these effort is made to prevent the simulation without limiter from blowing up.

However, numerical simulation shows that the simulation without bound-preserving technique blows up at about $T = 0.0155$. The distributions of c_f before blow-up is shown in Figure 3.5.

While with the settings exactly the same, the simulation with bound-preserving technique is stable. The distribution of c_f with time evolution in this case is given in Figure 3.6. We can see that the numerical approximations are high

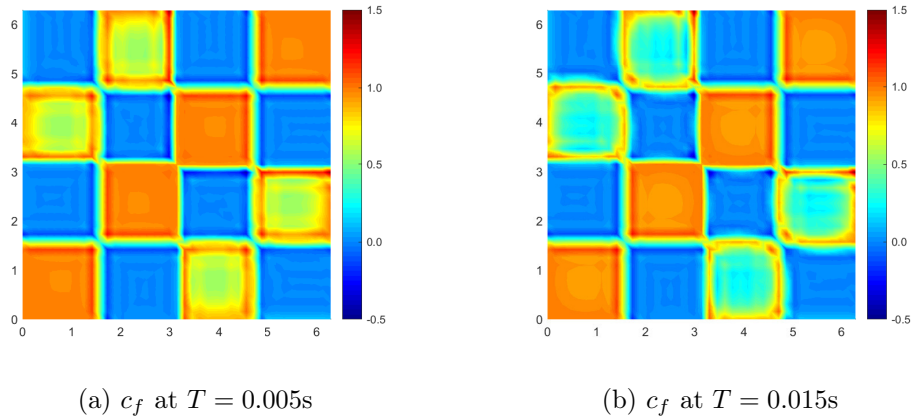
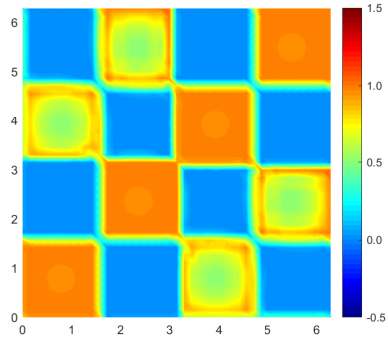


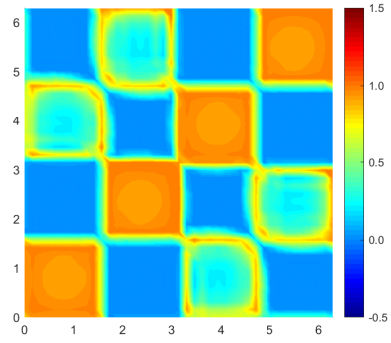
Figure 3.5: Example 3.6.2: Evolution of c_f without limiter

oscillatory. This is because the bound-preserving technique only preserves the bound but cannot suppress the oscillations. Some suitable limiters such as TVD, TVB and WENO limiters can suitably smooth the numerical approximations. Though oscillatory, the numerical simulation did not blow up. Therefore, with the bound-preserving technique, the numerical scheme is quite stable. What's more, we plot the evolution of extreme value of c_f and ϕ in Ω along simulation time in Figure 3.7 to illustrate the effectiveness of bound-preserving technique more clearly. We can observe that without the bound-preserving limiter, the concentration of acid c_f can be negative and greater than 1, and the porosity ϕ can also be negative, leading to ill-posed problems. With the bound-preserving technique, all the numerical approximations are within the physical bounds.

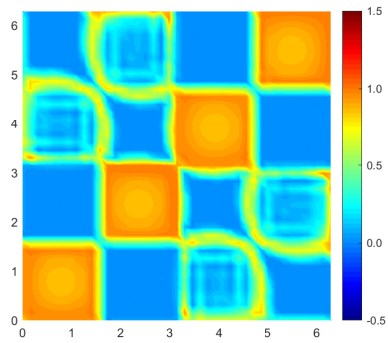
Example 3.6.3. *We simulate a single wormhole propagation experiment in rectangular rock tube, from which we can observe the formation and propagation of*



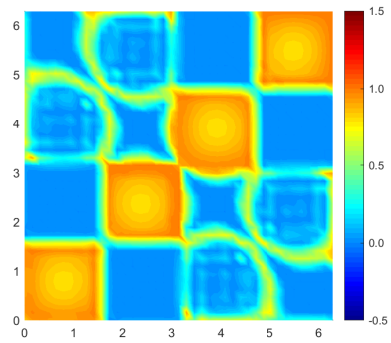
(a) c_f at $T = 0.005\text{s}$



(b) c_f at $T = 0.015\text{s}$

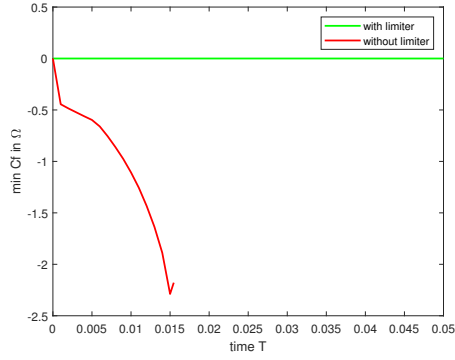


(c) c_f at $T = 0.030\text{s}$

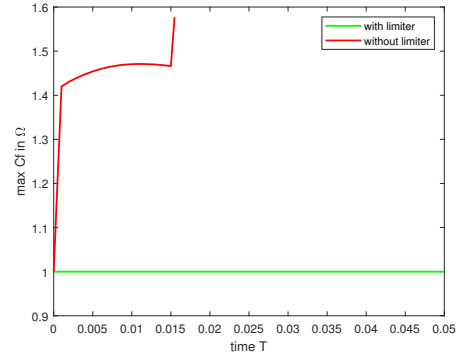


(d) c_f at $T = 0.050\text{s}$

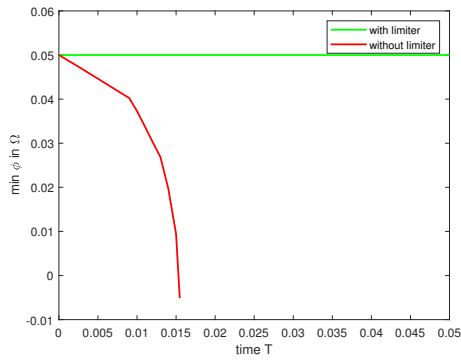
Figure 3.6: Example 3.6.2: Evolution of c_f with limiter.



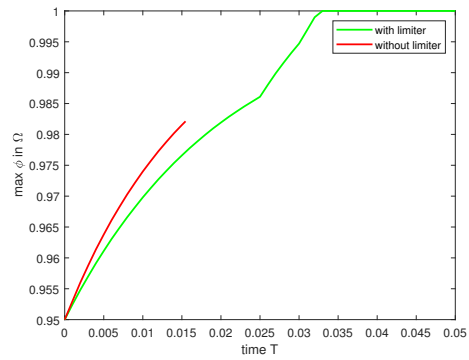
(a) Evolution of $\min_{x \in \Omega} c_f(x, t)$



(b) Evolution of $\max_{x \in \Omega} c_f(x, t)$



(c) Evolution of $\min_{x \in \Omega} \phi(x, t)$



(d) Evolution of $\max_{x \in \Omega} \phi(x, t)$

Figure 3.7: Example 3.6.2: Evolution of extreme value of c_f and ϕ in Ω .

a wormhole starting from a singular point. The parameters are taken as

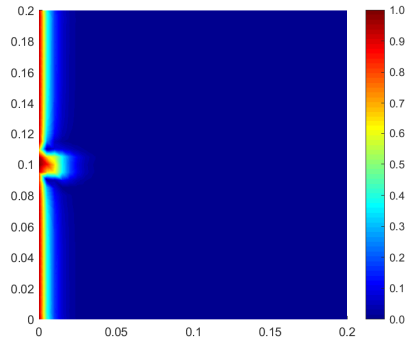
$$\begin{aligned} \mathbf{D} &= 0, \quad K_0 = 10^{-9}\text{m}^2, \quad T = 15\text{s}, \\ \alpha &= 100\text{kg/mol}, \quad k_c = 1\text{m/s}, \quad k_s = 10\text{m/s}, \\ \mu &= 10^{-2}\text{Pa}\cdot\text{s}, \quad f_I = f_p = 0, \quad c_{f0} = 0, \\ a_0 &= 2\text{m}^{-1}, \quad \rho_s = 2500\text{kg/m}^2, \quad \phi_0 = 0.2, \end{aligned}$$

Moreover, the computational domain is $\Omega = [0, 0.2\text{m}] \times [0, 0.2\text{m}]$. To investigate the phenomenon of wormhole propagation, we set a singular area with high porosity $\phi = 0.4$ and corresponding permeability determined by (3.2.5) which is about 10^{-8}m^2 on the middle of the left boundary with size $0.01\text{m} \times 0.01\text{m}$. The left and right boundary of the domain are Dirichlet conditions with pressure $p_d = 10000\text{Pa}$ and $p_d = -10000\text{Pa}$, respectively. The upper and lower boundaries of the domain are impermeable, i.e. $\mathbf{u} \cdot \mathbf{n} = 0$. The acid flows into the rock from the left boundary with a concentration of $c_I = 1\text{mol/m}^2$ and drained out of it from the right boundary.

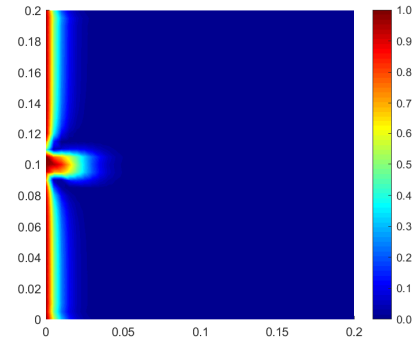
The contour plots of the concentration of acid and porosity of the rock at different time are shown in Figures 3.8-3.9, from which we can observe $c_f, \phi \in [0, 1]$ and the phenomenon of wormhole propagation along the whole simulation.

3.7 Concluding remarks

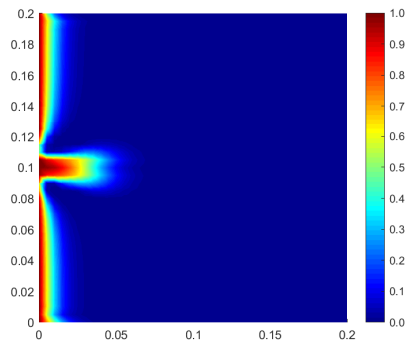
In this paper, we constructed high-order bound-preserving DG methods for wormhole propagation on triangular meshes. We have obtained the bound-preserving and monotone-increasing properties for concentration and porosity,



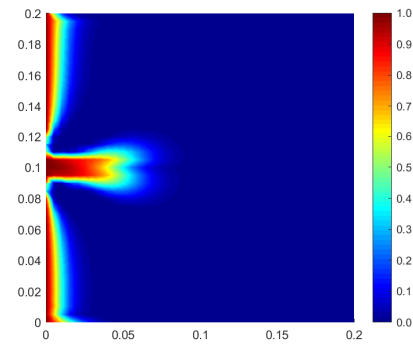
(a) c_f at $T = 1s$



(b) c_f at $T = 5s$

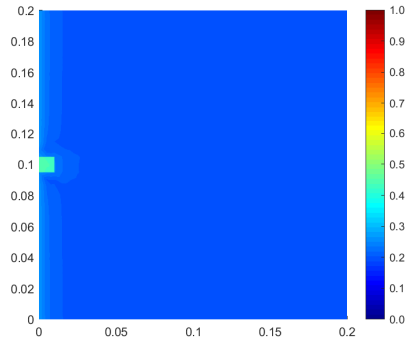


(c) c_f at $T = 10s$

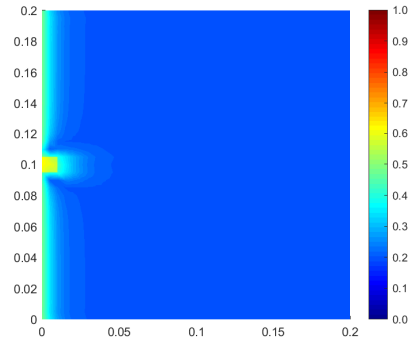


(d) c_f at $T = 15s$

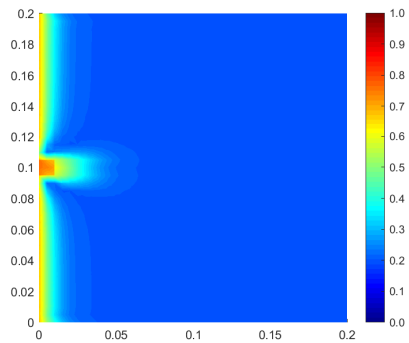
Figure 3.8: Example 3.6.3: Concentration of acid with time evolution.



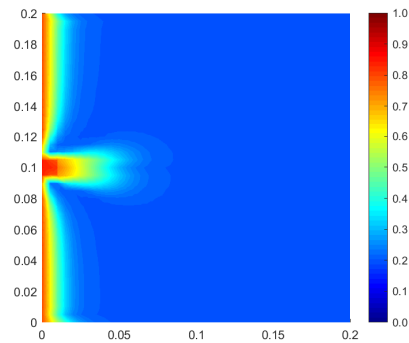
(a) ϕ at $T = 1$ s



(b) ϕ at $T = 5$ s



(c) ϕ at $T = 10$ s



(d) ϕ at $T = 15$ s

Figure 3.9: Example 3.6.3: Porosity of rock with time evolution.

respectively, with high-order accuracy. Numerical experiments have shown the accuracy and effectiveness of the bound-preserving technique.

Chapter 4

Conclusion

In the first work, we constructed the high-order bound-preserving discontinuous Galerkin method to multi-component compressible miscible displacements problem in porous media on triangular meshes. By introducing a new slope limiter and applying flux limiter, we attained the high-order accuracy without loss of bound-preserving properties. Numerical examples shown the accuracy and effectiveness of our scheme.

In our second work, we extended the idea in previous work to the wormhole propagation problems on triangular meshes. The properties of monotone increasing for porosity, and boundedness for porosity and concentration were guaranteed simultaneously. Numerical simulations shown the accuracy and necessity of our bound-preserving technique.

Bibliography

- [1] W.H. Reed and T. R. Hill, *Triangular mesh methods for the neutron transport equation*, Los Alamos Scientific Laboratory Report LA-UR-73-479, Los Alamos, NM, 1973.
- [2] B. Cockburn, S. Hou, C. W. Shu, *The Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws. IV: The multi-dimensional case*, Math. Comp. 54 (1990) 545-581.
- [3] B. Cockburn, S. Y. Lin, C. W. Shu, *TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws. III: One-dimensional systems*, J. Comput. Phys. 84 (1989) 90-113.
- [4] B. Cockburn, C. W. Shu, *TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws. II: General framework*, Math. Comp. 52 (1989) 411-435.
- [5] B. Cockburn, C. W. Shu, *The Runge-Kutta discontinuous Galerkin method for conservation laws. V: Multidimensional systems*, J. Comput. Phys. 141 (1998) 199-224.

- [6] B. Cockburn, C.-W. Shu, *The local discontinuous Galerkin method for time-dependent convection-diffusion systems*, SIAM Journal on Numerical Analysis, 35 (1998), 2440-2463.
- [7] S. Bartels, M. Jensen and R. Müller, *Discontinuous Galerkin finite element convergence for incompressible miscible displacement problem of low regularity*, SIAM Journal on Numerical Analysis, 47 (2009), 3720-3743.
- [8] P. Bastian, *A fully-coupled discontinuous Galerkin method for two-phase flow in porous media with discontinuous capillary pressure*, Computational Geosciences, 18 (2014), 779-796.
- [9] H.-Z. Chen and H. Wang, *An optimal-order error estimate on an H^1 -Galerkin mixed method for a nonlinear parabolic equation in porous medium flow*, Numerical Methods for Partial Differential Equations, 26 (2010), 188-205.
- [10] Z. Chen, H. Huang and J. Yan, *Third order Maximum-principle-satisfying direct discontinuous Galerkin methods for time dependent convection diffusion equations on unstructured triangular meshes*, Journal of Computational Physics, 308 (2016), 198-217.
- [11] S.-H. Chou and Q. Li, *Mixed finite element methods for compressible miscible displacement in porous media*, Mathematics of Computation, 57 (1991), 507-527.

- [12] A. Christlieb, Y. Liu, Q. Tang and Z. Xu, *Parametrized Maximum-principle-preserving and positivity-preserving flux limiter for WENO schemes on unstructured meshes*, Journal of Computational Physics, 281 (2015), 334-351.
- [13] M. Cui, *A combined mixed and discontinuous Galerkin method for compressible miscible displacement problem in porous media*, Journal of Computational and Applied Mathematics, 198 (2007), 19-34.
- [14] M. Cui, *Analysis of a semidiscrete discontinuous Galerkin scheme for compressible miscible displacement problem*, Journal of Computational and Applied Mathematics, 214 (2008), 617-636.
- [15] J. Douglas, Jr., R.E. Ewing and M.F. Wheeler, *A time-discretization procedure for a mixed finite element approximation of miscible displacement in porous media*, R.A.I.R.O. Analyse numérique, 17 (1983), 249-256.
- [16] J. Douglas, Jr., R.E. Ewing and M.F. Wheeler, *The approximation of the pressure by a mixed method in the simulation of miscible displacement*, R.A.I.R.O. Analyse numérique, 17 (1983), 17-33.
- [17] J. Douglas, Jr. and J. Roberts, *Numerical methods for a model for compressible miscible displacement in porous media*, Mathematics of Computation, 41 (1983), 441-459.
- [18] A. Ern, I. Mozolevski and L. Schuh, *Discontinuous Galerkin approximation of two-phase flows in heterogeneous porous media with discontinuous capillary*

- pressures*, Computer Methods in Applied Mechanics and Engineering, 199 (2010), 1491-1501.
- [19] A. Ern, I. Mozolevski and L. Schuh, *Accurate velocity reconstruction for Discontinuous Galerkin approximations of two-phase porous media flows*, Comptes Rendus Mathematique, 347 (2009), 551-554.
- [20] S. Gottlieb, D. Ketcheson and C.-W. Shu, *High order strong stability preserving time discretizations*, Journal of Scientific Computing, 38 (2009), 251-289.
- [21] S. Gottlieb, C.-W. Shu and E. Tadmor, *Strong stability-preserving high-order time discretization methods*, SIAM review 43.1 (2001), pp. 89-112.
- [22] H. Guo and Y. Yang, *Bound-Preserving Discontinuous Galerkin Method for Compressible Miscible Displacement in Porous Media*, SIAM Journal on Scientific Computing, 39 (2017), A1969-A1990.
- [23] H. Guo and Q. Zhang, *Error analysis of the semi-discrete local discontinuous Galerkin method for compressible miscible displacement problem in porous media*, Applied Mathematics and Computation, 259 (2015), 88-105.
- [24] H. Guo, Q. Zhang and Y. Yang, *A combined mixed finite element method and local discontinuous Galerkin method for miscible displacement problem in porous media*, Science China Mathematics, 57 (2014), 2301-2320.

- [25] L. Guo and Y. Yang, *Positivity preserving high-order local discontinuous Galerkin method for parabolic equations with blow-up solutions*, Journal of Computational Physics, 289 (2015), 181-195.
- [26] S. Kumar, *A mixed and discontinuous Galerkin finite volume element method for incompressible miscible displacement problems in porous media*, Numerical Methods for Partial Differential Equations, 28 (2012), 1354-1381.
- [27] N. Ma, D. Yang and T. Lu, *L^2 -norm error bounds of characteristics collocation method for compressible miscible displacement in porous media*, International Journal of Numerical Analysis and Modeling, 2 (2005), 28-42.
- [28] B. Rivière, *Discontinuous Galerkin finite element methods for solving the miscible displacement problem in porous media*, Ph.D. Thesis, The University of Texas at Austin, 2000.
- [29] T. Qin, C.-W. Shu and Y. Yang, *Bound-preserving discontinuous Galerkin methods for relativistic hydrodynamics*, Journal of Computational Physics, 315 (2016), 323-347.
- [30] C.-W. Shu, *Total-variation-diminishing time discretizations*, SIAM Journal on Scientific and Statistical Computing 9 (1988), 1073-1084.
- [31] S. Sun, B. Rivière and M.F. Wheeler, *A combined mixed finite element and discontinuous Galerkin method for miscible displacement problem in porous media*, Recent Progress in Computational and Applied PDEs, Tony Chan et

- al. (Eds.), Kluwer Academic Publishers, Plenum Press, Dordrecht, NewYork, 2002, 323-351.
- [32] S. Sun and M.F. Wheeler, *Discontinuous Galerkin methods for coupled flow and reactive transport problems*, Applied Numerical Mathematics, 52 (2005), 273-298.
- [33] S. Sun and M.F. Wheeler, *Symmetric and nonsymmetric discontinuous Galerkin methods for reactive transport in porous media*, SIAM Journal on Numerical Analysis, 43 (2005), 195-219.
- [34] H. Wang, D. Liang, R.E. Ewing, S.L. Lyons and G. Qin, *An approximation to miscible fluid flows in porous media with point sources and sinks by an Eulerian-Lagrangian localized adjoint method and mixed finite element methods*, SIAM Journal on Scientific Computing, 22 (2000), 561-581.
- [35] H. Wang, D. Liang, R.E. Ewing, S.L. Lyons and G. Qin, *An accurate approximation to compressible flow in porous media with wells*, Numerical Treatment of Multiphase Flows in Porous Media, Lecture Notes in Physics, 552 (2000), 324-332.
- [36] M. F. Wheeler and B. L. Darlow, *Interiori penalty Galerkin methods for miscible displacement problems in porous media*, Computational Methods in Nonlinear Mechanics, North-Holland, Amsterdam, 1980, 458-506.

- [37] Y. Xing, X. Zhang and C.-W. Shu, *Positivity preserving high order well balanced discontinuous Galerkin methods for the shallow water equations*, Advances in Water Resources, 33 (2010), 1476-1493.
- [38] T. Xiong, J.-M. Qiu and Z. Xu, *High order maximum-principle-preserving discontinuous Galerkin method for convection-diffusion equations*, SIAM Journal on Scientific Computing, 37 (2015), A583-A608.
- [39] Z. Xu, *Parametrized maximum principle preserving flux limiters for high order schemes solving hyperbolic conservation laws: One-dimensional scalar problem*, Mathematics of Computation, 83 (2014), 310-331.
- [40] D. Yang *A splitting positive definite mixed element method for miscible displacement of compressible flow in porous media*, Numerical Methods for Partial Differential Equations, 17 (2001), 229-249.
- [41] J. Yang and Y. Chen, *A priori error estimates of a combined mixed finite element and discontinuous Galerkin method for compressible miscible displacement with molecular diffusion and dispersion*, Journal of Computational Mathematics, 28 (2010), 1005-1022.
- [42] J. Yang, *A posteriori error of a discontinuous Galerkin scheme for compressible miscible displacement problems with molecular diffusion and dispersion*, International Journal for Numerical Methods in Fluids, 65 (2011), 781-797.

- [43] J. Yang and Y. Chen, *A priori error analysis of a discontinuous Galerkin approximation for a kind of compressible miscible displacement problems*, Science China Mathematics, 53 (2010), 2679-2696.
- [44] Y. Yang and C.-W. Shu, *Discontinuous Galerkin method for hyperbolic equations involving δ -singularities: Negative-order norm error estimates and applications*, Numerische Mathematik, 124,(2013),753-781.
- [45] Y. Yang, D. Wei and C.-W. Shu, *Discontinuous Galerkin method for Krause's consensus models and pressureless Euler equations*, Journal of Computational Physics, 252 (2013), 109-127.
- [46] F. Yu, H. Guo, N. Chuenjarern and Y. Yang, *Conservative local discontinuous Galerkin method for compressible miscible displacements in porous media*, Journal of Scientific Computing, 73 (2017), 1249-1275.
- [47] Y. Yuan, *The characteristic finite difference fractional steps methods for compressible two-phase displacement problem*, Science in China Series A-Mathematics, 42 (1999), 48-57.
- [48] Y. Yuan, *The upwind finite difference fractional steps methods for two-phase compressible flow in porous media*, Numerical Methods Partial Differential Equations, 19 (2003), 67-88.
- [49] Y. Yuan, *The modified upwind finite difference fractional steps method for compressible two-phase displacement problem*, Acta Mathematicae Applicatae Sinica, 20 (2004), 381-396.

- [50] X. Zhang and C.-W. Shu, *On maximum-principle-satisfying high order schemes for scalar conservation laws*, Journal of Computational Physics, 229 (2010), 3091-3120.
- [51] X. Zhang and C.-W. Shu, *On positivity preserving high order discontinuous Galerkin schemes for compressible Euler equations on rectangular meshes*, Journal of Computational Physics, 229 (2010), 8918-8934.
- [52] X. Zhang and C.-W. Shu, *Positivity-preserving high order discontinuous Galerkin schemes for compressible Euler equations with source terms*, Journal of Computational Physics, 230 (2011), 1238-1248.
- [53] X. Zhang, Y. Xia and C.-W. Shu, *Maximum-Principle-Satisfying and Positivity-Preserving High Order Discontinuous Galerkin Schemes for Conservation Laws on Triangular Meshes*, Journal of Scientific Computing, (2012), 50: 29-32.
- [54] Y. Zhang, X. Zhang and C.-W. Shu, *Maximum-principle-satisfying second order discontinuous Galerkin schemes for convection-diffusion equations on triangular meshes*, Journal of Computational Physics, 234 (2013), 295-316.
- [55] X. Zhao, Y. Yang and C. Seyler, *A positivity-preserving semi-implicit discontinuous Galerkin scheme for solving extended magnetohydrodynamics equations*, Journal of Computational Physics, 278 (2014), 400-415.
- [56] O.O. Akanni, H.A. Nasr-El-Din and D. Gusain, *A Computational Navier-Stokes Fluid-Dynamics-Simulation Study of Wormhole Propagation in*

- Carbonate-Matrix Acidizing and Analysis of Factors Influencing the Dissolution Process*, SPE Journal, 22 (2017), 187962.
- [57] N. Chuenjarern, Z. Xu and Y. Yang *High-order bound-preserving discontinuous Galerkin methods for compressible miscible displacements in porous media on triangular meshes*, Journal of Computational Physics, 378 (2019), 110-128.
- [58] J. Du and Y. Yang, *Maximum-principle-preserving third-order local discontinuous Galerkin methods on overlapping meshes*, Journal of computational physics, 377 (2019), 117-141.
- [59] C.N. Fredd and H.S. Fogler, *Influence of transport and reaction on wormhole formation in porous media*, AIChE journal, 44 (1998), 1933-1949.
- [60] H. Guo, L. Tian, Z. Xu, et. al., *High-order local discontinuous Galerkin method for simulating wormhole propagation*, Journal of Computational and Applied Mathematics, 350 (2019), 247-261.
- [61] M.L. Hoefner and H.S. Fogler, *Pore evolution and channel formation during flow and reaction in porous media*, AIChE Journal, 34 (1988), 45-54.
- [62] J. Kou, S. Sun and Y. Wu, *Mixed finite element-based fully conservative methods for simulating wormhole propagation*, Computer Methods in Applied Mechanics and Engineering, 298 (2016), 279-302.

- [63] X. Li and H. Rui, *Characteristic block-centered finite difference method for simulating incompressible wormhole propagation*, Computers and Mathematics with Applications, 73 (2017), 2171-2190.
- [64] X. Li and H. Rui, *Block-Centered Finite Difference Method for Simulating Compressible Wormhole Propagation*, Journal of Scientific Computing, 74 (2018), 1115-1145.
- [65] P. Maheshwari and V. Balakotaiah, *3D Simulation of carbonate acidization with HCl: comparison with experiments* In SPE Production and Operations Symposium. Society of Petroleum Engineers, 2013.
- [66] M.K. Panga, M. Ziauddin and V. Balakotaiah, *Two-scale continuum model for simulation of wormholes in carbonate acidization* AIChE journal, 51 (2005), 3231-3248.
- [67] W. Wei, A. Varavei and K. Sepehrnoori, *Modeling and Analysis on the Effect of Two-Phase Flow on Wormhole Propagation in Carbonate Acidizing*, SPE Journal, 22 (2017), 186111.
- [68] Y. Wu, A. Salama and S. Sun, *Parallel simulation of wormhole propagation with the Darcy–Brinkman–Forchheimer framework*, Computers and Geotechnics, 69 (2015), 564-577.
- [69] S. Srinivasan, J. Poggie and X. Zhang, *A positivity-preserving high order discontinuous Galerkin scheme for convection-diffusion equations*, Journal of Computational Physics, 366 (2018), 120-143.

- [70] Y. Yang, D. Wei and C.-W. Shu, *Discontinuous Galerkin method for Krause's consensus models and pressureless Euler equations*, Journal of Computational Physics, 252 (2013), 109–127.
- [71] L. Zhang, T. Cui and H. Liu, *A set of symmetric quadrature rules on triangles and tetrahedra*, Journal of Computational Mathematics, 27 (2009), 89-96.
- [72] C. Zhao, B.E. Hobbs, P. Hornby, et. al., *Theoretical and numerical analyses of chemical-dissolution front instability in fluid-saturated porous rocks*, International Journal for Numerical and Analytical Methods in Geomechanics, 32 (2008), 1107-1130.

Appendix A

Copyright documentations

A.1 Copyright documentation of Chapter 2

Order detail ID: 71854353
ISSN: 0021-9991
Publication Type: Journal
Volume:
Issue:
Start page:
Publisher: ACADEMIC PRESS

Permission Status:  **Granted**
Permission type: Republish or display content
Type of use: Thesis/Dissertation
Order License Id: 4551961331812

Hide details

Requestor type	Academic institution
Format	Print
Portion	chapter/article
Number of pages in chapter/article	19
The requesting person/organization	Ziyao Xu
Title or numeric reference of the portion(s)	Entire article
Title of the article or chapter the portion is from	High-Order Bound-preserving Discontinuous Galerkin Method for The Compressible Miscible Displacement Problem in Porous Media on Triangular Mesh
Editor of portion(s)	N/A
Author of portion(s)	Nattaporn Chuenjarern,Ziyao Xu,Yang Yang
Volume of serial or monograph	378
Page range of portion	110-128

Publication date of portion	7 November 2018
Rights for	Main product
Duration of use	Life of current edition
Creation of copies for the disabled	no
With minor editing privileges	no
For distribution to	Worldwide
In the following language(s)	Original language of publication
With incidental promotional use	no
Lifetime unit quantity of new product	Up to 4,999
Title	High-Order Bound-preserving Discontinuous Galerkin Method and its Application in Petroleum Engineering
Institution name	Michigan technological University
Expected presentation date	Apr 2019

A.2 Copyright documentation of Chapter 3

High-Order bound-preserving discontinuous Galerkin methods for wormhole propagation on triangular meshes has been completed as an article to submit to Journal of Computational Physics. As it has not yet been published (at the time of this report is made), we (Ziyao Xu, Yang Yang, and Hui Guo) still retain the copyright and as such, there is no need to obtain copyright documentation.