SIAM J. NUMER. ANAL.
Vol. 58, No. 1, pp. 519–540

POSITIVE AND ASYMPTOTIC PRESERVING APPROXIMATION OF THE RADIATION TRANSPORT EQUATION[*]

JEAN-LUC GUERMOND[†], BOJAN POPOV[†], AND JEAN RAGUSA[‡]

Abstract. We introduce a (linear) positive and asymptotic-preserving method for solving the one-group radiation transport equation. The approximation in space is discretization agnostic: the space approximation can be done with continuous or discontinuous finite elements (or finite volumes, or finite differences). The method is first-order accurate in space. This type of accuracy is consistent with Godunov's theorem since the method is linear. The two key theoretical results of the paper are Theorem 4.4 and Theorem 4.8. The method is illustrated with continuous finite elements, and it is observed to converge with the rate $\mathcal{O}(h)$ in the L^2-norm on manufactured solutions, and it is $\mathcal{O}(h^2)$ in the diffusion regime. The proposed method does not suffer from overshoots at the interfaces of optically thin and optically thick regions, is positive, and is asymptotic-preserving with respect to the diffusion limit.

Key words. finite element method, radiation transport, diffusion limit, asymptotic-preserving, positivity-preserving

AMS subject classifications. 65N30, 65N22, 82D75, 35Q20

DOI. 10.1137/19M1260785

1. Introduction.

The goal of this paper is to construct approximation techniques for the radiation transport equation that are both positive and asymptotic-preserving in the diffusion limit. Here we adopt the terminology used in the hyperbolic literature (see, e.g., Jin [14]). Letting ψ^ϵ be the solution to a problem that depends on a small parameter ϵ, and letting ψ_h^ϵ be the approximation of ψ^ϵ by some discretization method with mesh-size h, we say that the discretization method is asymptotic-preserving if it is convergent (meaning $\lim_{h\downarrow 0}\psi_{\epsilon,h}=\psi^\epsilon$ for every ϵ) and $\lim_{\epsilon\downarrow 0}\lim_{h\downarrow 0}\psi_h^\epsilon = \lim_{h\downarrow 0}\lim_{\epsilon\downarrow 0}\psi_h^\epsilon$. In the elliptic PDE literature, any numerical method with the above properties is said to be "robust," and a method is said to "lock" if it is not robust, i.e., if it is not asymptotic-preserving (see, e.g., Babuška and Suri [3]). These two terminologies are use interchangeably in the present paper.

In the wake of Reed and Hill [24] and Lesaint and Raviart [19], a dominant paradigm in the kinetic literature to solve the radiation transport equation consists of using the discontinuous Galerkin (dG) technique with the upwind flux. Unfortunately, to the best of our knowledge, there does not exist yet in the literature a dG technique that both is positive and does not lock in the thick diffusion limit. For instance, it was pointed out in Larsen [16] that the finite volume scheme "step scheme" (i.e., piecewise constant dG) with standard upwind locks in the diffusion limit. Several variations of the "step scheme" have been analyzed in Larsen, Morel, and Miller, Jr. [18]: it was shown that the "Lund–Wilson" and the "Castor" variants yield cell-edge angular fluxes that also lock in the diffusion limit. Furthermore, the cell-edge fluxes

[*]Received by the editors May 9, 2019; accepted for publication (in revised form) October 24, 2019; published electronically February 6, 2020.
https://doi.org/10.1137/19M1260785

Funding: The work of the authors was supported by a "Computational R&D in Support of Stockpile Stewardship" grant from Lawrence Livermore National Laboratory, the National Science Foundation grants DMS-1619892 and DMS-1620058, the Air Force Office of Scientific Research, USAF, grant FA9550-15-1-0257, and the Army Research Office grant W911NF-15-1-0517.

[†]Department of Mathematics, Texas A&M University 3368 TAMU, College Station, TX 77843 (guermond@math.tamu.edu, popov@math.tamu.edu).

[‡]Department of Nuclear Engineering, College Station, TX 77843 (jean.ragusa@tamu.edu).

for these schemes cannot reproduce the infinite medium solution. A "new" scheme was proposed in Larsen, Morel, and Miller, Jr. [18] but was subsequently dismissed due to a poor behavior at the boundaries. For many years, the diamond-difference scheme was found to be the best performing finite-difference scheme, even though its cell-edge fluxes lock in the thick diffusion limit. In Larsen and Morel [17], most of the previous schemes have been set aside in favor of the linear discontinuous finite element scheme (the piecewise linear dG technique with standard upwinding).

The cause for locking has been identified in a seminal paper by Adams [2]. The author analyzed multidimensional dG approximations and showed that some dG schemes lock in the diffusion limit because the upwind numerical flux forces the scalar flux, and thus the angular flux, to be continuous across the mesh cells. This observation has been confirmed in Guermond and Kanschat [9], where the equivalence of the limit problem to a mixed discretization for the Laplacian was proved and the nature of boundary layers appearing when the boundary flux is not isotropic was discussed. The asymptotic analysis in [2] and [9] suggests that the problem could be alleviated by modifying the upwind numerical flux. By making the amount of stabilization dependent on the scattering cross section so that the amount of upwinding decreases as the scattering cross section increases, it is shown in Ragusa, Guermond, and Kanschat [23] that locking can indeed be avoided in the thick diffusive limit, including for the dG0 approximation. The dG scheme thus obtained converges robustly for finite element spaces of any polynomial order, including piecewise constant functions (dG0), but, like all the other methods mentioned above, it is not guaranteed to be positive.

The objective of this work is to revisit the approximation theory for the radiation transport equation in heterogeneous media by using the algebraic framework (i.e., discretization agnostic) introduced in Guermond and Popov [10] and Guermond, Popov, and Tomas [12] and by incorporating in a roundabout way some ideas from Gosse and Toscani [8] and Ragusa, Guermond, and Kanschat [23]. We propose a method that is both positivity-preserving and does not lock (i.e., is asymptotic-preserving) in the thick diffusion limit. (The method shares some similarities with the two-dimensional finite volume technique from Buet, Després, and Franck [6, eqs. (18) and (19)].) Being linear, and in compliance with Godunov's theorem, the proposed algorithm is only first-order accurate in space though. This work is the first part of an ongoing project aiming at developing techniques that are high-order accurate, positivity-preserving, and robust (i.e., asymptotic-preserving) in the diffusion limit. The next step will be to increase the accuracy by introducing a nonlinear process; however, since this is not the purpose of the paper, we just mention in passing possible techniques to achieve this goal. This could be done in many ways; for instance, one could invoke a smoothness indicator like in Guermond and Popov [11, sect. 4.3], one could use a limiting technique in the spirit of the flux-corrected-transport method, or one could enforce positivity through inequality constraints like in Hauck and McClarren [13, sect. 4].

The paper is organized as follows. We introduce the model problem and the discrete setting (continuous and discontinuous finite elements) in section 2. The notion of graph viscosity, as defined in [10, 12], is introduced in section 3. We show in this section that the graph viscosity gives a scheme that is positive, but the scheme locks in the diffusion regime. This section is meant to give some perspective on the material introduced in section 4. The positive- and asymptotic-preserving scheme announced above is introduced in section 4. Originality is only claimed for the material presented in this section and the next one; the key results are Theorem 4.4 and Theorem 4.8. In section 5 we report numerical experiments illustrating the performance of the proposed method. The paper finishes with section 6 where we make concluding remarks.

2. Preliminaries.

In this section, we introduce the model problem under investigation and some notation regarding the discretization.

2.1. The model problem.

Let D be an open, bounded, connected Lipschitz domain in $\mathbb{R}^3$, and let $\mathcal{S}$ be the unit sphere in $\mathbb{R}^3$. We denote by $|\mathcal{S}|$ the measure of $\mathcal{S}$, i.e., $|\mathcal{S}| = 4\pi$. The boundary of D is denoted by ∂D, and the outer unit normal is denoted by $\boldsymbol{n}$. We want to solve the linear, one-group, radiation transport equation

$$\boldsymbol{\Omega}\cdot\nabla\psi(\boldsymbol{x},\boldsymbol{\Omega}) + \sigma_t(\boldsymbol{x})\psi(\boldsymbol{x},\boldsymbol{\Omega}) = \sigma_s(\boldsymbol{x})\overline{\psi}(\boldsymbol{x}) + q(\boldsymbol{x},\boldsymbol{\Omega}), \qquad (\boldsymbol{x},\boldsymbol{\Omega}) \in D\times\mathcal{S}, \tag{2.1a}$$

$$\psi(\boldsymbol{x},\boldsymbol{\Omega}) = \alpha(\boldsymbol{x},\boldsymbol{\Omega}), \qquad (\boldsymbol{x},\boldsymbol{\Omega}) \in \partial D_-, \tag{2.1b}$$

$$\overline{\psi}(\boldsymbol{x}) = \frac{1}{|\mathcal{S}|}\int_{\mathcal{S}} \psi(\boldsymbol{x},\boldsymbol{\Omega})\,\mathrm{d}\boldsymbol{\Omega}, \qquad \boldsymbol{x} \in D \tag{2.1c}$$

with $\partial D_- := \{(\boldsymbol{x},\boldsymbol{\Omega}) \in \partial D \times \mathcal{S} \,|\, \boldsymbol{\Omega}\cdot\boldsymbol{n}(\boldsymbol{x}) < 0\}$. The independent variable $(\boldsymbol{x},\boldsymbol{\Omega})$ spans $D\times\mathcal{S}$. The dependent variable $\psi(\boldsymbol{x},\boldsymbol{\Omega})$ is referred to as the angular intensity or angular flux, and the quantity $\overline{\psi}(\boldsymbol{x})$ is called scalar intensity or flux. The symbols $\sigma_t(\boldsymbol{x})$ and $\sigma_s(\boldsymbol{x})$ denote the total and scattering cross sections, respectively.

We want to investigate the approximation of (2.1) using either continuous or discontinuous finite elements. The objective is to construct a method that is asymptotic-preserving in the diffusion limit and positive (assuming that the boundary data, the cross sections, and the source term are nonnegative). In order to do that, we are going to adopt an idea from Gosse and Toscani [8], where a relaxation of the so-called hyperbolic heat equation is introduced, and combine it with an idea from Ragusa, Guermond, and Kanschat [23] where, in addition to the mesh-size, the stabilization parameters of the approximation have been made to depend on the cross sections as well.

2.2. Angular discretization.

In order to simplify the presentation we assume that the discretization in angle is done using a discrete ordinate technique. The (finite) angular quadrature is denoted $(\mu_l, \boldsymbol{\Omega}_l)_{l\in\mathcal{L}}$ and is assumed to satisfy

$$\sum_{l\in\mathcal{L}} \mu_l = |\mathcal{S}|, \quad \sum_{l\in\mathcal{L}} \mu_l\boldsymbol{\Omega}_l = \boldsymbol{0}, \quad \sum_{l\in\mathcal{L}} \boldsymbol{\Omega}_l|\boldsymbol{c}\cdot\boldsymbol{\Omega}_l| = \boldsymbol{0}, \quad \sum_{l\in\mathcal{L}} \mu_l\boldsymbol{\Omega}_l\otimes\boldsymbol{\Omega}_l = \frac{|\mathcal{S}|}{3}\mathbb{I} \tag{2.2}$$

for all $\boldsymbol{c} \in \mathbb{R}^3$, where $\mathbb{I}$ is the 3×3 identity matrix. Recall that $|\mathcal{S}| = 4\pi$. For further reference we also define the set $\mathcal{A}_L := \{\boldsymbol{\Omega}_l \in \mathbb{R}^3,\ l \in \mathcal{L}\}$ with $L := \mathrm{card}(\mathcal{L})$.

2.3. Continuous finite elements.

We describe in this section the Galerkin approximation of (2.1) with continuous finite elements. This technique is not positive and is known to exhibit severe oscillations; it will be appropriately stabilized in section 4.

Let $(\mathcal{T}_h)_{h>0}$ be a shape-regular sequence of unstructured matching meshes. For simplicity we assume that all the elements are generated from a reference element denoted $\widehat{K}$. The geometric transformation mapping $\widehat{K}$ to an arbitrary element $K \in \mathcal{T}_h$ is denoted $T_K : \widehat{K} \longrightarrow K$. We now introduce a reference finite element $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$, which we assume, for simplicity, to be a Lagrange element. We define the following scalar-valued finite element space:

$$P^{\mathrm{g}}(\mathcal{T}_h) = \{v \in \mathcal{C}^0(D;\mathbb{R}) \mid v_{|K}\circ T_K \in \widehat{P}\ \forall K \in \mathcal{T}_h\}. \tag{2.3}$$

The superscript $^{\mathrm{g}}$ is meant to remind us that the space is conforming for the gradient operator, e.g., $P^{\mathrm{g}}(\mathcal{T}_h) \subset H^1(D)$. The global shape functions are denoted by

$\{\varphi_i\}_{i\in\mathcal{V}}$; the associated Lagrange nodes are denoted $\{\boldsymbol{a}_i\}_{i\in\mathcal{V}}$ (here $\mathcal{V}$ is the index set enumerating the shape functions). We recall that the global shape functions satisfy the partition of unity property $\sum_{i\in\mathcal{V}}\varphi_i(\boldsymbol{x})=1$ for all $\boldsymbol{x}\in D$. We assume that they have positive mass

$$m_i := \int_D \varphi_i(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} > 0 \qquad \forall i\in\mathcal{V}. \tag{2.4}$$

For any $i\in\mathcal{V}$, the adjacency list $\mathcal{I}(i)$ is defined by setting $\mathcal{I}(i):=\{j\in\mathcal{V}\mid \varphi_i\varphi_j\not\equiv 0\}$. The approximation space for (2.1) is then defined to be

$$\boldsymbol{P}^{\mathrm{g}}(\mathcal{T}_h,\mathcal{A}_L) := \underbrace{P^{\mathrm{g}}(\mathcal{T}_h)\times\ldots\times P^{\mathrm{g}}(\mathcal{T}_h)}_{L\text{ times}}. \tag{2.5}$$

Let $\sigma_{t,i}$ and $\sigma_{s,i}$ be consistent approximations of σ_t and σ_s at the Lagrange node $\boldsymbol{a}_i$. For instance, let us assume that the mesh $\mathcal{T}_h$ is such that σ_t and σ_s are continuous over each cell K in $\mathcal{T}_h$ (σ_t and σ_s can be discontinuous across some mesh interfaces). Let us denote $\mathcal{T}(i)=\{K\in\mathcal{T}_h\mid \boldsymbol{a}_i\in K\}$. Then we can set $\sigma_{t,i}=\frac{1}{\mathrm{card}(\mathcal{T}(i))}\sum_{K\in\mathcal{T}(i)}\sigma_{t|K}(\boldsymbol{a}_i)$ and $\sigma_{s,i}=\frac{1}{\mathrm{card}(\mathcal{T}(i))}\sum_{K\in\mathcal{T}(i)}\sigma_{s|K}(\boldsymbol{a}_i)$. For further reference we denote the absorption cross section at the node $\boldsymbol{a}_i$ by $\sigma_{a,i}:=\sigma_{t,i}-\sigma_{s,i}$.

Let $\boldsymbol{\psi}_h := (\psi_{h,1},\ldots,\psi_{h,L})\in\boldsymbol{P}^{\mathrm{g}}(\mathcal{T}_h,\mathcal{A}_L)$ be the discrete ordinate Galerkin approximation of (2.1) with $\psi_{h,k}:=\sum_{j\in\mathcal{V}}\Psi_{jk}\varphi_j\in P(\mathcal{T}_h)$ for all $k\in\mathcal{L}$. The field $\boldsymbol{\psi}_h\in\boldsymbol{P}^{\mathrm{g}}(\mathcal{T}_h,\mathcal{A}_L)$ is obtained by solving the following set of linear equations:

$$\sum_{j\in\mathcal{I}(i)}\Psi_{jk}\int_D(\boldsymbol{\Omega}_k\cdot\nabla\varphi_j)\varphi_i\,\mathrm{d}\boldsymbol{x}+m_i\sigma_{t,i}\Psi_{ik}=m_i\sigma_{s,i}\overline{\Psi}_i+m_iq_{ik}+b^{\partial}_{ik}(\alpha^{\partial}_{ik}-\Psi_{ik}), \tag{2.6a}$$

$$\overline{\Psi}_i=\frac{1}{|\mathcal{S}|}\sum_{k\in\mathcal{L}}\mu_k\Psi_{ik}, \tag{2.6b}$$

where we have lumped the mass matrix, defined $q_{ik}:=\frac{1}{m_i}\int_D\varphi_i(\boldsymbol{x})q(\boldsymbol{x},\boldsymbol{\Omega}_k)\,\mathrm{d}\boldsymbol{x}$, and set

$$b^{\partial}_{ik}=m^{\partial}_i\frac{|\boldsymbol{\Omega}_k\cdot\boldsymbol{n}_i|-\boldsymbol{\Omega}_k\cdot\boldsymbol{n}_i}{2}. \tag{2.7}$$

Here $m^{\partial}_i:=\int_{\partial D}\varphi_i(\boldsymbol{x})\,\mathrm{d}s$, $\boldsymbol{n}_i$ is the unit normal vector (or approximation thereof) at the Lagrange node $\boldsymbol{a}_i$, and $\alpha^{\partial}_{ik}:=\alpha(\boldsymbol{a}_i,\boldsymbol{\Omega}_k)$. To refer to boundary degrees of freedom we introduce the following set of indices:

$$(\mathcal{V}\times\mathcal{L})^{\partial}:=\{(j,l)\in\mathcal{V}\times\mathcal{L}\mid \boldsymbol{\Omega}_l\cdot\boldsymbol{n}_j<0\}. \tag{2.8}$$

For further reference, we introduce

$$\boldsymbol{c}_{ij}:=\int_D\varphi_i(\boldsymbol{x})\nabla\varphi_j(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}. \tag{2.9}$$

With this notation, the discrete system is rewritten as follows for all $(i,k)\in\mathcal{V}\times\mathcal{L}$:

$$\sum_{j\in\mathcal{I}(i)\setminus\{i\}}\boldsymbol{\Omega}_k\cdot\boldsymbol{c}_{ij}(\Psi_{jk}-\Psi_{ik})+m_i\sigma_{t,i}\Psi_{ik}=m_i\sigma_{s,i}\overline{\Psi}_i+m_iq_{ik}+b^{\partial}_{ik}(\alpha^{\partial}_{ik}-\Psi_{ik}). \tag{2.10}$$

Notice that here we have used the partition of unity property which implies that $\sum_{j\in\mathcal{I}(i)}\boldsymbol{c}_{ij}=\boldsymbol{0}$.

Remark 2.1 (boundary conditions). We have imposed the boundary condition weakly in (2.10) by using the penalty technique usually invoked in the context of dG approximations. One can also enforce the boundary conditions strongly; in that case one sets $b^{\partial}_{ik} = 0$, and one adds the equations $\Psi_{ik} = \alpha^{\partial}_{ik}$ to (2.10) for all $(i, k) \in (\mathcal{V}\times\mathcal{L})^{\partial}$. □

As mentioned above, the linear system (2.10) has no positivity property. We are going to remedy this problem in section 3 by introducing some upwinding based on the graph Laplacian. But making the method positivity-preserving by simply introducing upwinding makes it lock. We present in section 4 a modification of the upwind graph viscosity that is both positivity-preserving and asymptotic-preserving.

2.4. Discontinuous finite elements. We briefly describe in this section the dG approximation of (2.1) with the centered numerical flux.

We use the same notation as in section 2.3 for the shape-regular sequence of unstructured matching meshes $(\mathcal{T}_h)_{h>0}$. We also introduce a reference finite element $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$. This may not be a Lagrange element. We define the following scalar-valued broken finite element space:

$$P^{\mathrm{b}}(\mathcal{T}_h) = \{v \in L^1(D;\mathbb{R}) \mid v_{|K}\circ T_K \in \widehat{P} \ \forall K \in \mathcal{T}_h\}. \tag{2.11}$$

The superscript $^{\mathrm{b}}$ is meant to remind us that the space is broken, i.e., the members of $P^{\mathrm{b}}(\mathcal{T}_h)$ can be discontinuous across the mesh interfaces. We denote by $\{\varphi_i\}_{i\in\mathcal{V}}$ the collection of the global shape functions generated from the reference shape functions. The support of each shape function is restricted to one mesh cell only. We assume that all the shape functions have a positive mass

$$m_i := \int_D \varphi_i \,\mathrm{d}x > 0 \quad \forall i \in \mathcal{V}. \tag{2.12}$$

We introduce the following adjacency sets:

$$\mathcal{I}(K) := \{i \in \mathcal{V} \mid \varphi_{i|K} \not\equiv 0\}, \qquad \mathcal{I}(\partial K) := \{i \in \mathcal{V} \mid \varphi_{i|\partial K} \not\equiv 0\}. \tag{2.13}$$

Note that $\mathcal{I}(\partial K)$ not only includes indices of shape functions with support in $\mathcal{I}(K)$, but this set also includes indices of shape functions that do not have support in K. More precisely $\mathcal{I}(\partial K)$ is the union of two disjoint sets $\mathcal{I}(\partial K^{\mathrm{i}})$ and $\mathcal{I}(\partial K^{\mathrm{e}})$ defined as

$$\mathcal{I}(\partial K^{\mathrm{i}}) := \{i \in \mathcal{I}(K) \bigm| \varphi_{i|\partial K} \not\equiv 0\}, \qquad \mathcal{I}(\partial K^{\mathrm{e}}) := \mathcal{I}(\partial K)\backslash\mathcal{I}(\partial K^{\mathrm{i}}). \tag{2.14}$$

For any $i \in \mathcal{V}$, let $K \in \mathcal{T}_h$ be such that $i \in \mathcal{I}(K)$; then we define the adjacency set $\mathcal{I}(i)$ to be the collection of the indices $j \in \mathcal{V}$ such that either $j \in \mathcal{I}(K)$ and $\varphi_i\varphi_j|_K \not\equiv 0$, or $j \in \mathcal{I}(\partial K^{\mathrm{e}})$ and $\varphi_i\varphi_j|_{\partial K} \not\equiv 0$.

Let $K \in \mathcal{T}_h$. We finally assume that the reference finite element is such that the sets of shape functions $\{\varphi_j\}_{j\in\mathcal{I}(K)}$ form a partition of unity over K, and the shape functions $\{\varphi_j\}_{j\in\mathcal{I}(\partial K^{\mathrm{i}})}$, $\{\varphi_j\}_{j\in\mathcal{I}(\partial K^{\mathrm{e}})}$ form partitions of unity over ∂K, i.e.,

$$\sum_{j\in\mathcal{I}(K)} \varphi_{j|K} = 1, \qquad \sum_{j\in\mathcal{I}(\partial K^{\mathrm{i}})} \varphi_{j|\partial K} = 1, \quad \text{and} \quad \sum_{j\in\mathcal{I}(\partial K^{\mathrm{e}})} \varphi_{j|\partial K} = 1. \tag{2.15}$$

Let $i \in V$, $j \in \mathcal{I}(i)$, let us set

$$\boldsymbol{c}^{K}_{ij} := \int_K \varphi_i \nabla\varphi_j \,\mathrm{d}\boldsymbol{x}, \qquad \boldsymbol{c}^{\partial K}_{ij} := \tfrac{1}{2}\int_{\partial K} \varphi_j\varphi_i \boldsymbol{n}_K \,\mathrm{d}s, \tag{2.16}$$

and let us define the vector $\boldsymbol{c}_{ij}$ as follows:

$$\boldsymbol{c}_{ij} := \begin{cases} \boldsymbol{c}_{ij}^{K} & \text{if } j \in \mathcal{I}(K)\backslash\mathcal{I}(\partial K^{\mathsf{i}}), \\ \boldsymbol{c}_{ij}^{K} - \boldsymbol{c}_{ij}^{\partial K} & \text{if } j \in \mathcal{I}(\partial K^{\mathsf{i}}), \\ \boldsymbol{c}_{ij}^{\partial K} & \text{if } j \in \mathcal{I}(\partial K^{\mathsf{e}}). \end{cases} \tag{2.17}$$

The partition of unity property (2.15) implies that $\sum_{j\in\mathcal{I}(i)} \boldsymbol{c}_{ij} = \boldsymbol{0}$ (see, for instance, [12, Lem. 4.1]).

Let us introduce the discrete broken space

$$\boldsymbol{P}^{\mathrm{b}}(\mathcal{T}_h, \mathcal{A}_L) := \underbrace{P^{\mathrm{b}}(\mathcal{T}_h)\times \ldots \times P^{\mathrm{b}}(\mathcal{T}_h)}_{L \text{ times}}. \tag{2.18}$$

Let us denote by $\boldsymbol{\psi}_h := (\psi_{h,1}, \ldots, \psi_{h,L}) \in \boldsymbol{P}^{\mathrm{b}}(\mathcal{T}_h, \mathcal{A}_L)$ the dG approximation of (2.1) using the centered flux with $\psi_{h,k} := \sum_{j\in\mathcal{V}\times\mathcal{L}} \Psi_{jk}\varphi_j \in P^{\mathrm{b}}(\mathcal{T}_h)$. The field $\boldsymbol{\psi}_h \in \boldsymbol{P}^{\mathrm{b}}(\mathcal{T}_h, \mathcal{A}_L)$ is defined to be the solution of

$$\sum_{j\in\mathcal{I}(i)\backslash\{i\}} \boldsymbol{\Omega}_k\cdot\boldsymbol{c}_{ij}(\Psi_{jk} - \Psi_{ik}) + m_i\sigma_{t,i}\Psi_{ik} = m_i\sigma_{s,i}\overline{\Psi}_i + m_i q_{ik} + b_{ik}^{\partial}(\alpha_{ik}^{\partial} - \Psi_{ik}). \tag{2.19}$$

We insist here that we are using the centered flux; there is no upwinding. The proper stabilization will be introduced in section 4.

Remark 2.2 (definition of $\sigma_{t,i}$ and $\sigma_{s,i}$). The definition of the coefficients $\sigma_{t,i}$ and $\sigma_{s,i}$ depend on the definition of the shape functions. If the shape functions are nodal-based (i.e., Lagrange polynomials), then one can take $\sigma_{t,i} = \sigma_{t|K}(\boldsymbol{a}_i)$ and $\sigma_{s,i} = \sigma_{s|K}(\boldsymbol{a}_i)$, where K contains the support of φ_i and $\boldsymbol{a}_i$ is the Lagrange node associated with φ_i. Recall that we denote $\sigma_{a,i} := \sigma_{t,i} - \sigma_{s,i}$. □

3. Graph viscosity, positivity, and locking.

In order to give some perspective, we start by introducing a mechanism that ensures positivity but fails to be robust in the diffusion limit. A correction that makes the method asymptotic-preserving in the diffusion limit is introduced in section 4.

3.1. Positivity.

Our starting point is the algebraic system (2.10) or (2.19), which we call Galerkin, or centered, or inviscid approximation. We are not going to make any distinction between the continuous and the dG approximations. The discrete space are henceforth denoted $P(\mathcal{T}_h)$ and $\boldsymbol{P}(\mathcal{T}_h)$, i.e., we have removed the superscripts $^{\mathrm{g}}$ and $^{\mathrm{b}}$. We consider the following linear system: Find $\boldsymbol{\psi}_h = \sum_{i\in\mathcal{V}}(\Psi_{i1}, \ldots, \Psi_{iL})\varphi_i \in \boldsymbol{P}(\mathcal{T}_h)$ so that the following holds for all $(i, k) \in \mathcal{V}\times\mathcal{L}$:

$$\sum_{j\in\mathcal{I}(i)\backslash\{i\}} \boldsymbol{\Omega}_k\cdot\boldsymbol{c}_{ij}(\Psi_{jk} - \Psi_{ik}) + m_i\sigma_{t,i}\Psi_{ik} = m_i\sigma_{s,i}\overline{\Psi}_i + m_i q_{ik} + b_{ik}^{\partial}(\alpha_{ik}^{\partial} - \Psi_{ik}), \tag{3.1}$$

where we recall that $\sum_{j\in\mathcal{I}(i)} \boldsymbol{c}_{ij} = 0$ for all $i \in \mathcal{V}$. Taking inspiration from Guermond and Popov [11], we introduce the coefficient d_{ij}^{k} defined by setting

$$d_{ij}^{k} = \max(\max(\boldsymbol{\Omega}_k\cdot\boldsymbol{c}_{ij}, 0), \max(\boldsymbol{\Omega}_k\cdot\boldsymbol{c}_{ji}, 0)). \tag{3.2}$$

Then we perturb (3.1) as follows:

$$\sum_{j\in\mathcal{I}(i)\backslash\{i\}} (\boldsymbol{\Omega}_k\cdot\boldsymbol{c}_{ij} - d_{ij}^{k})(\Psi_{jk} - \Psi_{ik}) + m_i\sigma_{t,i}\Psi_{ik} = m_i\sigma_{s,i}\overline{\Psi}_i + m_i q_{ik} + b_{ik}^{\partial}(\alpha_{ik}^{\partial} - \Psi_{ik}). \tag{3.3}$$

The extra term $\sum_{j\in\mathcal{I}(i)\setminus\{i\}} -d_{ij}^k(\Psi_{jk}-\Psi_{ik})$ is a graph viscosity since it acts on the connectivity graph of the degrees of freedom. Notice that this perturbation is first-order consistent since it vanishes if $\Psi_{jk}=\Psi_{ik}$ for all $j\in\mathcal{I}(i)$. In one dimension on a uniform mesh, where the adjacency list is $\{i-1,i,i+1\}$, we have $d_{ij}^k=\frac{|\boldsymbol{\Omega}_k|}{2}$ both for continuous piecewise linear finite elements and for piecewise constant discontinuous elements; as a result, we have $\sum_{j\in\mathcal{I}(i)\setminus\{i\}} -d_{ij}^k(\Psi_{jk}-\Psi_{ik}) = -\frac{|\boldsymbol{\Omega}_k|}{2}(\Psi_{i-1,k}-2\Psi_{ik}+\Psi_{i+1,k})$, which is the expression one expects from an artificial viscosity term. Further insight on the graph viscosity is given in Remark 3.2 in the context of the dG0 setting. The following result is the key motivation for introducing the graph viscosity.

LEMMA 3.1 (minimum/maximum principle). *Let d_{ij}^k be defined in* (3.2). *Let $(\Psi_{ik})_{(i,k)\in\mathcal{V}\times\mathcal{L}}$ be the solution to* (3.3). *Let $\Psi^{\min}:=\min_{(i,k)\in\mathcal{V}\times\mathcal{L}}\Psi_{ik}$ and $\Psi^{\max}:=\max_{(i,k)\in\mathcal{V}\times\mathcal{L}}\Psi_{ik}$. Let $(i_0,k_0),(i_1,k_1)\in\mathcal{V}\times\mathcal{L}$ be so that $\Psi_{i_0k_0}=\Psi^{\min}$ and $\Psi_{i_1k_1}=\Psi^{\max}$.*

(i) *Assume that $\min_{(j,l)\in\mathcal{V}\times\mathcal{L}}(\sigma_{a,j}+b_{jl}^\partial)>0$. Then*

$$\frac{m_{i_0}q_{i_0k_0}+b_{i_0k_0}^\partial\alpha_{i_0k_0}^\partial}{m_{i_0}\sigma_{a,i_0}+b_{i_0k_0}^\partial}\le\Psi^{\min}\le\Psi^{\max}\le\frac{m_{i_1}q_{i_1k_1}+b_{i_1k_1}^\partial\alpha_{i_1k_1}^\partial}{m_{i_1}\sigma_{a,i_1}+b_{i_1k_1}^\partial}. \tag{3.4}$$

(ii) *Otherwise, assume that for all $i\in\mathcal{V}$ such that $\sigma_{a,i}=0$ and $b_{ik}^\partial=0$ the definition of d_{ij}^k is slightly modified so that $\boldsymbol{\Omega}_k\cdot\boldsymbol{c}_{ij}<d_{ij}^k$ for all $j\in\mathcal{I}(i)$ (instead of $\boldsymbol{\Omega}_k\cdot\boldsymbol{c}_{ij}\le d_{ij}^k$). If $0\le\min_{(i,k)\in\mathcal{V}\times\mathcal{L}}q_{ik}$ and $0\le\min_{(i,k)\in(\mathcal{V}\times\mathcal{L})^\partial}\alpha_{ik}^\partial$, then $0\le\Psi^{\min}$.*

(iii) *Moreover, under the same assumptions on d_{ij}^k as in* (ii), *if $\max_{(i,k)\in\mathcal{V}\times\mathcal{L}}q_{ik}\le 0$, then $\Psi^{\max}\le\max_{(i,k)\in(\mathcal{V}\times\mathcal{L})^\partial}\alpha_{ik}^\partial$.*

Proof. Proof of (i) assuming that $\min_{(j,l)\in\mathcal{V}\times\mathcal{L}}(\sigma_{a,j}+b_{jl}^\partial)>0$. Let $(i_0,k_0)\in\mathcal{V}\times\mathcal{L}$ be the indices of the degree of freedom where the minimum is attained; that is, $\Psi_{ik}\ge\Psi_{i_0k_0}$ for all $(i,k)\in\mathcal{V}\times\mathcal{L}$. Then using that

$$\boldsymbol{\Omega}_k\cdot\boldsymbol{c}_{ij}-d_{ij}^k\le\max(\boldsymbol{\Omega}_k\cdot\boldsymbol{c}_{ij},0)-d_{ij}^k\le 0,$$

together with $\Psi_{jk_0}-\Psi_{i_0k_0}\ge 0$ for all $j\in\mathcal{I}(i_0)$, and $\Psi_{i_0k_0}\le\overline{\Psi}_{i_0}$, we infer that

$$\begin{aligned}
&m_{i_0}\sigma_{s,i_0}\Psi_{i_0k_0}+m_{i_0}q_{i_0k_0}+b_{i_0k_0}^\partial(\alpha_{i_0k_0}^\partial-\Psi_{i_0k_0})\\
&\qquad\le m_{i_0}\sigma_{s,i_0}\overline{\Psi}_{i_0}+m_{i_0}q_{i_0k_0}+b_{i_0k_0}^\partial(\alpha_{i_0k_0}^\partial-\Psi_{i_0k_0})\\
&\qquad=\sum_{j\in\mathcal{I}(i_0)\setminus\{i_0\}}(\boldsymbol{\Omega}_{k_0}\cdot\boldsymbol{c}_{i_0j}-d_{i_0j}^{k_0})(\Psi_{jk_0}-\Psi_{i_0k_0})+m_{i_0}\sigma_{t,i_0}\Psi_{i_0k_0}\le m_{i_0}\sigma_{t,i_0}\Psi_{i_0k_0}.
\end{aligned}$$

Hence $m_{i_0}q_{i_0k_0}+b_{i_0k_0}^\partial\alpha_{i_0k_0}^\partial\le(m_{i_0}\sigma_{a,i_0}+b_{i_0k_0}^\partial)\Psi_{i_0k_0}$. The assertion follows readily since we assumed that $m_{i_0}\sigma_{a,i_0}+b_{i_0k_0}^\partial>0$. The proof of the other assertion, regarding $\Psi^{\max}$, is analogous.

Proof of (ii) assuming that $0\le\min_{(i,k)\in\mathcal{V}\times\mathcal{L}}q_{ik}$ and $0\le\min_{(i,k)\in(\mathcal{V}\times\mathcal{L})^\partial}\alpha_{ik}^\partial$. From part (i) we have $m_{i_0}q_{i_0}+b_{i_0k_0}^\partial\alpha_{i_0k_0}^\partial\le(m_i\sigma_{a,i_0}+b_{i_0k_0}^\partial)\Psi_{i_0k_0}$. So, we need to prove that $\Psi_{i_0k_0}\ge 0$ only in the case $\sigma_{a,i_0}=0$ and $b_{i_0k_0}^\partial=0$. Assuming that $\sigma_{a,i_0}=0$ and $b_{i_0k_0}^\partial=0$, we have from part (i) the following inequality:

$$0\le m_{i_0}q_{i_0k_0}\le\sum_{j\in\mathcal{I}(i_0)\setminus\{i_0\}}(\boldsymbol{\Omega}_{k_0}\cdot\boldsymbol{c}_{i_0j}-d_{i_0j}^{k_0})(\Psi_{jk_0}-\Psi_{i_0k_0})\le 0.$$

The assumption $\boldsymbol{\Omega}_k\cdot\boldsymbol{c}_{i_0j}-d_{i_0j}^{k_0}<0$ for all $j\in\mathcal{I}(i_0)$ implies that $\Psi_{jk_0}-\Psi_{i_0k_0}=0$ for all $j\in\mathcal{I}(i_0)$. Therefore, we conclude that the global minimum is attained not only at the degree of freedom (i_0,k_0) but also in the whole neighborhood, i.e., for all $j\in\mathcal{I}(i_0)$. Repeating the above argument for a global minimum at (j,k_0) for all $j\in\mathcal{I}(i_0)$, we derive that the global minimum is either nonnegative (if $m_j\sigma_{a,j}+b^{\partial}_{jk_0}>0$) or again attained in the whole neighborhood of j, i.e., for all $s\in\mathcal{I}(j)$. This process can terminate in two ways only: (i) either the global minimum is nonnegative at some j because $m_j\sigma_{a,j}+b^{\partial}_{jk_0}>0$ or (ii) the global minimum is attained at all of the degrees of freedom topologically connected to i_0. In this case we have that $\Psi_{jk_0}=\Psi_{i_0k_0}$ for all j in the same connected component as i_0, which is the entire set $\mathcal{V}$ since $\mathcal{T}_h$ is connected (because D is connected). However, for any fixed k_0 there exists j such that Ψ_{jk_0} is on the inflow boundary for $\boldsymbol{\Omega}_{k_0}$, that is, $b^{\partial}_{jk_0}>0$, and we conclude $\Psi_{i_0k_0}=\Psi_{jk_0}\ge 0$.

Proof of (iii) assuming that $\min_{(i,k)\in\mathcal{V}\times\mathcal{L}} q_{ik}\le 0$. By proceeding as in step (i), we infer that

$$\begin{aligned}&m_{i_1}\sigma_{s,i_1}\Psi_{i_1k_1}+m_{i_1}q_{i_1k_1}+b^{\partial}_{i_1k_1}(\alpha^{\partial}_{i_1k_1}-\Psi_{i_1k_1})\\&\qquad\ge\sum_{j\in\mathcal{I}(i_1)\backslash\{i_1\}}(\boldsymbol{\Omega}_{k_1}\cdot\boldsymbol{c}_{i_1j}-d^{k_1}_{i_1j})(\Psi_{jk_1}-\Psi_{i_1k_1})+m_{i_1}\sigma_{t,i_1}\Psi_{i_1k_1},\end{aligned}$$

i.e., $(m_{i_1}\sigma_{a,i_1}+b^{\partial}_{i_1k_1})\Psi_{i_1k_1}\le m_{i_1}q_{i_1k_1}+b^{\partial}_{i_1k_1}\alpha^{\partial}_{i_1k_1}$, which implies $\Psi_{i_1k_1}\le\alpha^{\partial}_{i_1k_1}$ if $m_{i_1}\sigma_{a,i_1}+b^{\partial}_{i_1k_1}>0$. If $m_{i_1}\sigma_{a,i_1}+b^{\partial}_{i_1k_1}>0$, then $0\ge\sum_{j\in\mathcal{I}(i_1)\backslash\{i_1\}}(\boldsymbol{\Omega}_{k_1}\cdot\boldsymbol{c}_{i_1j}-d^{k_1}_{i_1j})(\Psi_{jk_1}-\Psi_{i_1k_1})\ge 0$ and $\Psi_{jk_1}=\Psi_{i_1k_1}$ for all $j\in\mathcal{I}(i_1)$. Then we proceed as in step (ii) until we reach a degree of freedom j that is on the inflow boundary for $\boldsymbol{\Omega}_{k_1}$, i.e., $b^{\partial}_{jk_1}>0$. The $\Psi^{\max}=\Psi_{i_1k_1}=\Psi_{jk_1}\le\alpha^{\partial}_{j,k_1}$. □

Remark 3.2 (dG0). To give some insight about (3.2) to the reader who is familiar with the dG formulation of the radiation transport equation, we now interpret the graph viscosity in terms of numerical flux. Assume that $P^{\mathrm{b}}(\mathcal{T}_h)$ is composed of piecewise constant polynomials. In this case the indices $i\in\mathcal{V}$ coincide with the enumeration of the cells in $\mathcal{T}_h$. Let $K_i\in\mathcal{T}_h$ be a cell, and let $(K_j)_{j\in\mathcal{I}(i)}$ be all the cells that share a face with K_i; then recalling (2.17), we have $\boldsymbol{c}_{ii}=\int_{K_i}\varphi_i\nabla\varphi_i\,\mathrm{d}\boldsymbol{x}-\frac{1}{2}\int_{\partial K_i}\varphi_i^2\boldsymbol{n}_K\,\mathrm{d}s$ and $\boldsymbol{c}_{ij}=\frac{1}{2}\int_{\partial K_i}\varphi_i\varphi_j\boldsymbol{n}_K\,\mathrm{d}s$ for all $j\in\mathcal{I}(i)\backslash\{i\}$. Let us set $\psi_{h,k}(\boldsymbol{x})=\sum_{j\in\mathcal{V}}\Psi_{jk}\varphi_j\in P^{\mathrm{b}}(\mathcal{T}_h)$. Let us denote $\psi^{\mathrm{e}}_{h,k}$ and $\psi^{\mathrm{i}}_{h,k}$, respectively, the exterior trace and the interior trace of $\psi_{h,k}$ on ∂K_i. Recall that $\psi_{h,k|K}=\psi^{\mathrm{i}}_{h,k}=\Psi_{ik}\varphi_i$ and $\psi^{\mathrm{e}}_{h,k}=\sum_{j\in\mathcal{I}(i)\backslash\{i\}}\Psi_{jk}\varphi_j$. Then

$$\begin{aligned}\sum_{j\in\mathcal{I}(i)}\boldsymbol{\Omega}_k\cdot\boldsymbol{c}_{ij}\Psi_{jk}-\sum_{j\in\mathcal{I}(i)\backslash\{i\}}d^k_{ij}(\Psi_{jk}-\Psi_{ik})&=\int_{K_i}\varphi_i\boldsymbol{\Omega}_k\cdot\nabla\psi_{h,k}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\&\quad+\int_{\partial K_i}\frac{1}{2}(\psi^{\mathrm{e}}_{h,k}-\psi^{\mathrm{i}}_{h,k})\varphi_i\boldsymbol{\Omega}_k\cdot\boldsymbol{n}_K\,\mathrm{d}s-\int_{\partial K_i}\frac{1}{2}(\psi^{\mathrm{e}}_{h,k}-\psi^{\mathrm{i}}_{h,k})\varphi_i|\boldsymbol{\Omega}_k\cdot\boldsymbol{n}_K|\,\mathrm{d}s\\&=-\int_{K_i}\psi_{h,k}(\boldsymbol{x})\boldsymbol{\Omega}_k\cdot\nabla\varphi_i\,\mathrm{d}\boldsymbol{x}\\&\quad+\int_{\partial K_i}\Big(\frac{1}{2}(\psi^{\mathrm{e}}_{h,k}+\psi^{\mathrm{i}}_{h,k})\boldsymbol{\Omega}_k\cdot\boldsymbol{n}_K+\frac{1}{2}(\psi^{\mathrm{i}}_{h,k}-\psi^{\mathrm{e}}_{h,k})|\boldsymbol{\Omega}_k\cdot\boldsymbol{n}_K|\Big)\varphi_i\,\mathrm{d}s.\end{aligned}$$

Hence the dG numerical flux is $\frac{1}{2}(\psi^{\mathrm{e}}_{h,k}+\psi^{\mathrm{i}}_{h,k})\boldsymbol{\Omega}_k\cdot\boldsymbol{n}_K+\frac{1}{2}(\psi^{\mathrm{i}}_{h,k}-\psi^{\mathrm{e}}_{h,k})|\boldsymbol{\Omega}_k\cdot\boldsymbol{n}_K|$, and we recognize the standard upwind flux. In conclusion, in the dG0 context, the system (3.3) with d^k_{ij} defined in (3.2) simply corresponds to the standard upwinding approximation. □

3.2. Locking. Unfortunately, as reported numerous times in the literature, just enforcing positivity in a scheme does not prevent locking. Actually the approximation (3.3) with the graph viscosity defined in (3.2) locks in the diffusive regime.

More precisely, let $\epsilon > 0$, and let us consider the following rescaled version of the problem (2.1):

$$\boldsymbol{\Omega}\cdot\nabla\psi^\epsilon(\boldsymbol{x},\boldsymbol{\Omega}) + \frac{\sigma_t(\boldsymbol{x})}{\epsilon}\psi^\epsilon(\boldsymbol{x},\boldsymbol{\Omega}) = \frac{\sigma_s(\boldsymbol{x})}{\epsilon}\overline{\psi}^\epsilon(\boldsymbol{x}) + \epsilon q(\boldsymbol{x},\boldsymbol{\Omega}), \quad (\boldsymbol{x},\boldsymbol{\Omega}) \in D\times\mathcal{S}, \tag{3.5a}$$

$$\psi^\epsilon(\boldsymbol{x},\boldsymbol{\Omega}) = \alpha(\boldsymbol{x},\boldsymbol{\Omega}), \quad (\boldsymbol{x},\boldsymbol{\Omega}) \in \partial D_- \tag{3.5b}$$

with the additional assumption that $\frac{\sigma_t(\boldsymbol{x})-\sigma_s(\boldsymbol{x})}{\epsilon} = \epsilon\sigma_a(\boldsymbol{x})$. The limit solution to this problem when $\epsilon \downarrow 0$ has been investigated thoroughly in the literature; see, e.g., Chandrasekhar [7] and Malvagi and Pomraning [21]. It is known in particular that $\psi^0 := \lim_{\epsilon\downarrow}\psi^\epsilon$ is isotropic (i.e., does not depend on $\boldsymbol{\Omega}$) and satisfies the following diffusion equation:

$$-\nabla\cdot\left(\frac{1}{3\sigma_s}\nabla\psi^0\right) + \sigma_a\psi^0 = q, \quad \boldsymbol{x} \in D, \tag{3.6a}$$

$$\psi^0(\boldsymbol{x}) = \frac{1}{2\pi}\int_{\partial D_-} W(|\boldsymbol{\Omega}\cdot\boldsymbol{n}(\boldsymbol{x})|)\alpha(\boldsymbol{\Omega},\boldsymbol{x})\,\mathrm{d}\boldsymbol{\Omega}, \quad \boldsymbol{x} \in \partial D, \tag{3.6b}$$

where $W(\mu) = \frac{\sqrt{3}}{2}\mu H(\mu)$ is defined in terms of Chandrasekhar's H-function for isotropic scattering in a conservative medium (see [21] for the asymptotic analysis and [7] for details on the H-function). It is known that the convergence $\psi^\epsilon \to \psi^0$ is not uniform unless $\frac{1}{6}m + \boldsymbol{M}\cdot\boldsymbol{n} = 0$ with

$$m(\boldsymbol{x}) := \frac{1}{\pi}\int_{\partial D_-} \alpha(\boldsymbol{\Omega},\boldsymbol{x})|\boldsymbol{\Omega}\cdot\boldsymbol{n}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{\Omega}, \tag{3.7a}$$

$$\boldsymbol{M}(\boldsymbol{x}) := \frac{1}{4\pi}\int_{\partial D_-} \alpha(\boldsymbol{\Omega},\boldsymbol{x})|\boldsymbol{\Omega}\cdot\boldsymbol{n}(\boldsymbol{x})|\boldsymbol{\Omega}\,\mathrm{d}\boldsymbol{\Omega}. \tag{3.7b}$$

Observe that $m = \alpha$ and $\boldsymbol{M} = -\frac{1}{6}\alpha\boldsymbol{n}$, i.e., $\frac{1}{6}m + \boldsymbol{M}\cdot\boldsymbol{n} = 0$, if α is isotropic (see also Guermond and Kanschat [9, Thms. 5.3 and 5.4]).

Let $\boldsymbol{\psi}_h^\epsilon$ be the discrete ordinate approximation to the solution of (3.5) with d_{ij}^k defined in (3.2):

$$\sum_{j\in\mathcal{I}(i)\setminus\{i\}} (\boldsymbol{\Omega}_k\cdot\boldsymbol{c}_{ij} - d_{ij}^k)(\Psi_{jk}^\epsilon - \Psi_{ik}^\epsilon) + \epsilon m_i\sigma_{a,i}\Psi_{ik}^\epsilon = m_i\frac{\sigma_{s,i}}{\epsilon}(\overline{\Psi}_i^\epsilon - \Psi_{ik}^\epsilon) + \epsilon m_i q_{ik} + b_{ik}^\partial(\alpha_{ik}^\partial - \Psi_{ik}). \tag{3.8}$$

Proposition 3.3 (locking). *Let the graph viscosity d_{ij}^k be defined in (3.2). Assume that $\min_{i,j}\sum_{k\in\mathcal{L}}\mu_k d_{ij}^k > 0$. If the boundary conditions are homogeneous, i.e., $\alpha_{ik}^\partial = 0$, then $\lim_{\epsilon\to 0}\Psi_{jk}^\epsilon = 0$ for all $i, j \in \mathcal{V}$ and all $k \in \mathcal{L}$. Hence there is locking if $q \not\equiv 0$.*

Proof. To avoid losing the reader who is not familiar with functional analysis techniques, we are going to proceed formally. A rigorous proof can be done by proceeding as in Guermond and Kanschat [9, sect. 4]. Using Landau's notation, let us introduce the formal asymptotic expansion $\boldsymbol{\psi}_h^\epsilon = \boldsymbol{\psi}_h^0 + \epsilon\boldsymbol{\psi}_h^1 + \epsilon^2\boldsymbol{\psi}_h^2 + \mathcal{O}(\epsilon^3)$. Inserting this expansion into (3.8) gives

$$0 = m_i\sigma_{s,i}(\overline{\Psi}_i^0 - \Psi_{ik}^0),$$

$$\sum_{j\in\mathcal{I}(i)\setminus\{i\}} (\boldsymbol{\Omega}_k\cdot\boldsymbol{c}_{ij} - d_{ij}^k)(\Psi_{jk}^0 - \Psi_{ik}^0) = m_i\sigma_{s,i}(\overline{\Psi}_i^1 - \Psi_{ik}^1) + b_{ik}^\partial(\alpha_{ik}^\partial - \Psi_{ik}^0).$$

The first equation shows that $\overline{\Psi}_i^0 = \Psi_{ik}^0$. Then integrating the second equation with respect to the angles gives

$$\sum_{j\in\mathcal{I}(i)\backslash\{i\}} (\overline{\Psi}_j^0 - \overline{\Psi}_i^0)\Big(\Big(\sum_{k\in\mathcal{L}} \mu_k \boldsymbol{\Omega}_k\Big)\cdot \boldsymbol{c}_{ij} - \sum_{k\in\mathcal{L}} \mu_k d_{ij}^k\Big) = m_i^\partial \frac{1}{4}\mathfrak{m}_i^\partial - m_i^\partial \delta_i^\partial \overline{\Psi}_i^0,$$

where $\mathfrak{m}_i^\partial := \frac{4}{|\mathcal{S}|}\sum_{k\in\mathcal{L}_i^-} \mu_k \alpha_{ik}^\partial |\boldsymbol{\Omega}_k \cdot \boldsymbol{n}_i|$ and $\delta_i^\partial := \frac{1}{|\mathcal{S}|}\sum_{k\in\mathcal{L}_i^-}\mu_k|\boldsymbol{\Omega}_k\cdot\boldsymbol{n}_i|$ with $\mathcal{L}_i^- := \{k \in \mathcal{L} \mid \boldsymbol{\Omega}_k\cdot\boldsymbol{n}_i < 0\}$. (Note that the continuous counterparts of the coefficients $\mathfrak{m}_i^\partial$ and δ_i^∂ are $\frac{4}{|\mathcal{S}|}\int_{\boldsymbol{\Omega}\cdot\boldsymbol{n}_i<0} \alpha(\boldsymbol{a}_i, \boldsymbol{\Omega})|\boldsymbol{\Omega}\cdot\boldsymbol{n}_i| \,\mathrm{d}\boldsymbol{\Omega}$ and $\frac{1}{|\mathcal{S}|}\int_{\boldsymbol{\Omega}\cdot\boldsymbol{n}_i<0}|\boldsymbol{\Omega}\cdot\boldsymbol{n}_i|\,\mathrm{d}\boldsymbol{\Omega} = \frac{1}{4}$, respectively.) Recalling that the assumptions on the angular quadrature (2.2) imply that $\sum_{k\in\mathcal{L}}\mu_k\boldsymbol{\Omega}_k = \boldsymbol{0}$, setting $\gamma_{ij} := \sum_{k\in\mathcal{L}}\mu_k d_{ij}^k$ we obtain $\sum_{j\in\mathcal{I}(i)\backslash\{i\}}\gamma_{ij}(\overline{\Psi}_j^0 - \overline{\Psi}_i^0) = m_i^\partial(\frac{1}{4}\mathfrak{m}_i^\partial - \delta_i^\partial\overline{\Psi}_i^0)$. Let us assume now that $\alpha_{ik}^\partial = 0$, and let us multiply this equation by $2\overline{\Psi}_i^0$; then

$$\sum_{j\in\mathcal{I}(i)\backslash\{i\}} \gamma_{ij}(\overline{\Psi}_j^0)^2 + \gamma_{ij}(\overline{\Psi}_j^0 - \overline{\Psi}_i^0)^2 - \gamma_{ij}(\overline{\Psi}_i^0)^2 = -2m_i^\partial\delta_i^\partial(\overline{\Psi}_i^0)^2.$$

Now we observe that $\gamma_{ij} = \gamma_{ji}$, and we sum the above identity over $i \in \mathcal{V}$. This yields $\sum_{i\in\mathcal{V}}(2m_i^\partial\delta_i^\partial(\overline{\Psi}_i^0)^2 + \sum_{j\in\mathcal{I}(i)\backslash\{i\}}\gamma_{ij}(\overline{\Psi}_j^0 - \overline{\Psi}_i^0)^2) = 0$. This in turn implies that $\overline{\Psi}_j^0 = \overline{\Psi}_i^0 = 0$ for all $i,j \in \mathcal{V}$ since $\min_{i,j}\gamma_{ij} > 0$. Then we have established that $\lim_{\epsilon\downarrow 0}\boldsymbol{\psi}_{\epsilon,h} = 0$, which in turn immediately implies that $\lim_{h\downarrow 0}\lim_{\epsilon\downarrow 0}\boldsymbol{\psi}_{\epsilon,h} = 0$. Let $R : \mathcal{P}(\mathcal{T}_h) \to L^2(D\times\mathcal{S})$ be some suitable linear reconstruction operator. Then we have $\lim_{\epsilon\downarrow 0}\lim_{\epsilon\downarrow 0} R(\boldsymbol{\psi}_{\epsilon,h}) = 0$. On the other hand (referring the reader to [21, 7] for the details), the limit solution to (3.5), i.e., the solution to (3.6), is not zero because we assumed $q \not\equiv 0$. Moreover, using standard approximation techniques, it can be established that $\lim_{h\downarrow 0} R(\boldsymbol{\psi}_{\epsilon,h}) = \psi^\epsilon$ for every $\epsilon > 0$. Hence, we have proved that

$$\lim_{\epsilon\downarrow 0}\lim_{h\downarrow 0} R(\boldsymbol{\psi}_{\epsilon,h}) = \lim_{\epsilon\downarrow 0}\psi^\epsilon = \psi^0 \not\equiv 0 = \lim_{\epsilon\downarrow 0}\lim_{\epsilon\downarrow 0} R(\boldsymbol{\psi}_{\epsilon,h}),$$

which in turn shows that the approximation is not asymptotic-preserving, i.e., the approximation locks. ∎

4. An asymptotic-preserving scheme.

The goal of this section is to introduce the asymptotic-preserving method mentioned in the introduction of the paper. This scheme is somewhat discretization agnostic since it is solely based on the algebraic formulations (2.10) and (2.19).

4.1. Preliminary notation.

In the rest of the paper we use the following notation:

$$\text{(4.1a)}\qquad \mathcal{L}_i^- := \{k \in \mathcal{L} \mid \boldsymbol{\Omega}_k\cdot\boldsymbol{n}_i < 0\}, \qquad \delta_i^\partial := \frac{1}{|\mathcal{S}|}\sum_{k\in\mathcal{L}_i^-}\mu_k|\boldsymbol{\Omega}_k\cdot\boldsymbol{n}_i|,$$

$$\text{(4.1b)}\qquad \mathfrak{m}_i^\partial := \frac{4}{|\mathcal{S}|}\sum_{k\in\mathcal{L}_i^-}\mu_k\alpha_{ik}^\partial|\boldsymbol{\Omega}_k\cdot\boldsymbol{n}_i|, \qquad \boldsymbol{\mathfrak{M}}_i^\partial := \frac{1}{|\mathcal{S}|}\sum_{k\in\mathcal{L}_i^-}\mu_k\alpha_{ik}^\partial|\boldsymbol{\Omega}_k\cdot\boldsymbol{n}_i|\boldsymbol{\Omega}_k.$$

We also denote $\mathcal{V}^\circ = \{i \in \mathcal{V} \mid \varphi_{i|\partial D} \equiv 0\}$ and $\mathcal{V}^\partial = \mathcal{V}\backslash\mathcal{V}^\circ$. For further reference we define $\mathcal{L}(\boldsymbol{x})^- := \{k \in \mathcal{L} \mid \boldsymbol{\Omega}_k\cdot\boldsymbol{n}(\boldsymbol{x}) < 0\}$, $\mathfrak{m}^\partial(\boldsymbol{x}) := \frac{4}{|\mathcal{S}|}\sum_{k\in\mathcal{L}(\boldsymbol{x})^-}\mu_k\alpha(\boldsymbol{x},\boldsymbol{\Omega}_k)^\partial|\boldsymbol{\Omega}_k\cdot\boldsymbol{n}(\boldsymbol{x})|$, and $\boldsymbol{\mathfrak{M}}^\partial(\boldsymbol{x}) := \frac{1}{|\mathcal{S}|}\sum_{k\in\mathcal{L}(\boldsymbol{x})^-}\mu_k\alpha(\boldsymbol{x},\boldsymbol{\Omega}_k)^\partial|\boldsymbol{\Omega}_k\cdot\boldsymbol{n}(\boldsymbol{x})|\boldsymbol{\Omega}_k$ for all $\boldsymbol{x} \in \partial D$.

Now, depending whether one uses (or prefers using) continuous finite elements or discontinuous finite elements, we introduce two sets of coefficients. In the context of continuous finite elements we set

$$c_{ij}^{\mathrm{g,d}} = \int_D \frac{1}{\widetilde{\sigma}_s(\boldsymbol{x})} \nabla\varphi_i(\boldsymbol{x})\cdot\nabla\varphi_j(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}, \qquad i,j\in\mathcal{V}, \tag{4.2}$$

where $\widetilde{\sigma}_s(\boldsymbol{x}) = \max(\sigma_s(\boldsymbol{x}), \varepsilon\max(\frac{1}{\mathrm{diam}(D)}, \|\sigma_s\|_{L^\infty(D)}))$ with $\varepsilon = 10^{-14}$. The quantity $\widetilde{\sigma}_s$ is introduced to avoid divisions by zero. For discontinuous finite elements of degree 1 or larger we proceed as follows. We assume for simplicity that $\widetilde{\sigma}_s$ is constant over each mesh cell and denote $\widetilde{\sigma}_K := \widetilde{\sigma}_{s|K}$ for all cells K. Let $K\in\mathcal{T}_h$, and let $\mathcal{F}_K^\circ$ be the set of the faces of K that are not on ∂D; that is, $F\in\mathcal{F}_K^\circ$ if there exists $K'\in\mathcal{T}_h$, $K'\neq K$, such that $F := K\cap K'$. For every $F\in\mathcal{F}_K^\circ$, we define $\widetilde{\sigma}_F := \frac{2\widetilde{\sigma}_K\widetilde{\sigma}_{K'}}{\widetilde{\sigma}_K+\widetilde{\sigma}_{K'}}$ and $h_F := \mathrm{diam}(F)$. Let $v\in\boldsymbol{P}^{\mathrm{b}}(\mathcal{T}_h)$, and let v_K, $v_{K'}$ be the restrictions of $\boldsymbol{v}$ on K and K', respectively; we define the weighted average of v across $F\in\mathcal{F}_K^\circ$ as follows: $\{v\} := \frac{\widetilde{\sigma}_K}{\widetilde{\sigma}_K+\widetilde{\sigma}_{K'}} v_{K|F} + \frac{\widetilde{\sigma}_{K'}}{\widetilde{\sigma}_K+\widetilde{\sigma}_{K'}} v_{K'|F}$. The jump of v across $F\in\mathcal{F}_K^\circ$ is defined by setting $[\![v]\!] := v_K - v'_K$. We now define for all $j\in\mathcal{I}(i)$

$$\begin{aligned} c_{ij}^{\mathrm{b,d}} = {} & \int_K \frac{1}{\widetilde{\sigma}_s}\nabla\varphi_i\cdot\nabla(\varphi_{j|K})\,\mathrm{d}\boldsymbol{x} + \gamma \sum_{F\in\mathcal{F}_K^\circ} \frac{1}{\widetilde{\sigma}_F h_F}\int_F [\![\varphi_i]\!][\![\varphi_j]\!]\,\mathrm{d}s \\ & - \sum_{F\in\mathcal{F}_K^\circ}\int_F \left(\left\{\frac{1}{\widetilde{\sigma}_s}\nabla\varphi_i\right\}\cdot\boldsymbol{n}_K[\![\varphi_j]\!] + \left\{\frac{1}{\widetilde{\sigma}_s}\nabla\varphi_j\right\}\cdot\boldsymbol{n}_K[\![\varphi_i]\!]\right)\mathrm{d}s, \end{aligned} \tag{4.3}$$

where γ is a user-defined constant of order 1, and with the convention that $\varphi_{j|K} = 0$ if $j\in\mathcal{I}(\partial K^{\mathrm{e}})$. Denoting by c_{ij}^{d} either $c_{ij}^{\mathrm{g,d}}$ or $c_{ij}^{\mathrm{b,d}}$, depending on the context, and with $v_h := \sum_{j\in\mathcal{V}}\mathsf{V}_j\varphi_j$ and $w_h := \sum_{j\in\mathcal{V}}\mathsf{W}_j\varphi_j$, the bilinear form $a : P(\mathcal{T}_h)\times P(\mathcal{T}_h)\to\mathbb{R}$ defined by

$$a(v_h, w_h) := \frac{1}{3}\sum_{i,j\in\mathcal{V}} c_{ij}^{\mathrm{d}}\mathsf{V}_j\mathsf{W}_i \tag{4.4}$$

is the discrete weak form of the operator $-\nabla\cdot(\frac{1}{3\sigma_s}\nabla v)$ which naturally appears in the diffusion limit of (2.1), i.e., (3.6). Notice that the partition of unity property implies that $\sum_{j\in\mathcal{I}(i)} c_{ij}^{\mathrm{d}} = 0$; hence, we can also write $a(v_h, w_h) = \frac{1}{3}\sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{V}\setminus\{i\}} c_{ij}^{\mathrm{d}}(\mathsf{V}_j - \mathsf{V}_i)\mathsf{W}_i$.

4.2. Description of the method. To avoid repeating the same argument for continuous finite elements and discontinuous finite elements, we denote by c_{ij}^{d} either $c_{ij}^{\mathrm{g,d}}$ or $c_{ij}^{\mathrm{b,d}}$ depending on the context. For any pair $i,j\in\mathcal{V}$ in the same stencil, say, $j\in\mathcal{I}(i)$ (or equivalently $i\in\mathcal{I}(j)$), we define

$$d_{ij}^k := \max(\max(\boldsymbol{\Omega}_k\cdot\boldsymbol{c}_{ij}, 0), \max(\boldsymbol{\Omega}_k\cdot\boldsymbol{c}_{ji}, 0)), \quad \sigma_{s,ij} := \frac{1}{2}(\sigma_{s,i} + \sigma_{s,j}), \tag{4.5a}$$

$$h_{ij} := \frac{3}{\sigma_{s,ij}|c_{ij}^{\mathrm{d}}|}\frac{1}{|\mathcal{S}|}\sum_{k\in\mathcal{L}}\mu_k d_{ij}^k, \quad h_i := \frac{1}{\mathrm{card}(\mathcal{I}(i)) - 1}\sum_{j\in\mathcal{I}(i)\setminus\{i\}} h_{ij}. \tag{4.5b}$$

Notice that $d_{ij}^k = |\boldsymbol{c}_{ij}\cdot\boldsymbol{\Omega}_k|$ if either $i\in\mathcal{V}^\circ$ or $j\in\mathcal{V}^\circ$ since in this case $\boldsymbol{c}_{ij} = -\boldsymbol{c}_{ji}$. The stabilized formulation we consider consists of solving the following set of linear equations:

$$\text{(4.6a)}\quad \sum_{j\in\mathcal{I}(i)\setminus\{i\}} \frac{1}{\sigma_{s,ij}h_{ij}+1}(\boldsymbol{\Omega}_k\cdot\boldsymbol{c}_{ij} - d_{ij}^k)(\Psi_{jk}-\Psi_{ik}) + m_i\sigma_{a,i}\Psi_{ik}$$
$$= m_iq_{ik} + \frac{m_i\sigma_{s,i}}{\sigma_{s,i}h_i+1}\left(-\Psi_{ik}+\overline{\Psi}_i\right) + \frac{1}{\sigma_{s,i}h_i+1}b_{ik}^{\partial}(\beta_{ik}^{\partial}-\Psi_{ik}),$$

$$\text{(4.6b)}\quad \beta_{ij}^{\partial} := \theta_i\alpha_{ik}^{\partial} + (1-\theta_i)\left(\frac{1}{2}\mathfrak{m}_i^{\partial} - 3\boldsymbol{\mathfrak{M}}_i^{\partial}\cdot\boldsymbol{n}_i\right),\qquad \theta_i := \max(1-2\sigma_{s,i}h_i, 0),$$

where it is implicitly understood that $\beta_{ij}^{\partial} = 0$ if $i \in \mathcal{V}^{\circ}$.

Remark 4.1 (consistency). The above formulation coincides with the centered Galerkin approximation (3.1) if $d_{ij}^k = 0$. In the general case, i.e., with d_{ij}^k as defined in (4.5a), we have $d_{ij}^k \sim m_ih^{-1}$, where h is the mesh-size; hence $h_{ij} \sim m_ih^{-1}/(m_ih^{-2}) \sim h$ and $h_i \sim h$. This computation shows that both h_{ij} and h_i scale like the mesh-size (at most). Hence, (4.6a) converges to (3.3) when $\sigma_sh \to 0$. In other words, the solutions to (4.6a) and (3.3) are close when the mesh-size is significantly finer than the mean free path. The above arguments shows that (4.6a) is a consistent approximation of (2.1) (the consistency error is first-order with respect to the mesh-size). □

Remark 4.2 (boundary conditions). The boundary conditions in (4.6) are enforced weakly. Observe that we recover $\beta_{ij}^{\partial} \approx \theta_i\alpha_{ij}^{\partial}$ when the boundary condition at the degree of freedom i is isotropic, and we have equality $\beta_{ij}^{\partial} = \theta_i\alpha_{ij}^{\partial}$ if the angular quadrature satisfies $1 = \frac{4}{|\mathcal{S}|}\sum_{k\in\mathcal{L}_i^-}\mu_k|\boldsymbol{\Omega}_k\cdot\boldsymbol{n}_i|$ (notice that $\theta_i = 1+\mathcal{O}(\sigma_{s,i}h_i)$). When the boundary condition is anisotropic and when the local mesh-size is not small enough to resolve the mean free path, i.e., $2\sigma_{s,i}h(i) \geq 1$, we obtain $\beta_{ij}^{\partial} := \frac{1}{2}\mathfrak{m}_i^{\partial} - 3\boldsymbol{\mathfrak{M}}_i^{\partial}\cdot\boldsymbol{n}_i$. The key motivation for the proposed definition of the boundary condition is based on the following observation: Let $\psi^0 := \lim_{\epsilon\to 0}\psi^{\epsilon}$, where ψ^{ϵ} solves the rescaled problem (3.5). Let $\psi_{dG,h}^{\epsilon}$ be the dG approximation of (3.5) with the upwind numerical flux (assuming that the polynomial degree is larger than or equal to 1), and let $\psi_{dG}^0 := \lim_{h\to 0}\lim_{\epsilon\to 0}\psi_{dG,h}^{\epsilon}$; here the order the two limits are taken is important. Then it is observed in Adams [2, sect. III.D] and proved in Guermond and Kanschat [9, Thm. 5.4] that $\psi_{dG|\partial D}^0 = \frac{1}{2}\mathfrak{m}^{\partial} - 3\boldsymbol{\mathfrak{M}}^{\partial}\cdot\boldsymbol{n}$ (notice that all the arguments in [9] hold true by replacing integrals over the angles by any discrete measure (i.e., quadrature) with the properties stated in (2.2)). If the incoming flux at the boundary is such that $\frac{1}{2}\mathfrak{m}^{\partial} + 3\boldsymbol{\mathfrak{M}}^{\partial}\cdot\boldsymbol{n} \neq 0$, it is known that $\psi^0 \neq \psi_{dG}^0$, but it also known nevertheless that $\frac{1}{2}\mathfrak{m}^{\partial} - 3\boldsymbol{\mathfrak{M}}^{\partial}\cdot\boldsymbol{n}$ is a very good approximation of $\psi_{|\partial D}^0$; see, e.g., discussions in [2, p. 318] and [9, sect. 5.5]. Moreover we have $\psi_{|\partial D}^0 = \psi_{dG|\partial D}^0 = \frac{1}{2}\mathfrak{m}^{\partial} - 3\boldsymbol{\mathfrak{M}}^{\partial}\cdot\boldsymbol{n} = \mathfrak{m}^{\partial}$ when $\frac{1}{2}\mathfrak{m}^{\partial} + 3\boldsymbol{\mathfrak{M}}^{\partial}\cdot\boldsymbol{n} = 0$ (i.e., the incoming flux is isotropic); see, e.g., [9, Thm. 5.3]. □

Remark 4.3 (literature). Let us now show the connection between (4.6) and the technique introduced in Gosse and Toscani [8]. The system solved in this reference is the time-dependent version of (2.1) in one space dimension with two angular directions only: $\rho\partial_t(u,v) + \partial_x(u,-v) + \sigma_s(u,v) = \sigma_s\frac{1}{2}(u+v,u+v)$. Using upwind finite differences (or finite volumes), the proposed scheme is $\rho\partial_t(u_i,v_i) + (\frac{u_i-u_{i-1}}{h}, \frac{v_i-v_{i+1}}{h}) = \frac{\sigma_s}{\sigma_sh+1}(v_i-u_{i-1}, u_i-v_{i+1})$; see equation (6) in [8]. After simple manipulations, we observe that the scheme can be recast as follows: $\rho\partial_t(u_i,v_i) + \frac{1}{\sigma_sh+1}(\frac{u_i-u_{i-1}}{h}, \frac{v_i-v_{i+1}}{h}) + \frac{\sigma_s}{\sigma_sh+1}(u_i,v_i) = \frac{\sigma_s}{\sigma_sh+1}\frac{1}{2}(u_i+v_i,u_i+v_i)$. Hence, the trick introduced in [8] consists of multiplying both the upwind finite differences and the scattering terms by the coefficient $\frac{1}{\sigma_sh+1}$. This is exactly what is done in (4.6a). In our case the upwind finite difference is the term $\sum_{j\in\mathcal{I}(i)\setminus\{i\}}(\boldsymbol{\Omega}_k\cdot\boldsymbol{c}_{ij} - d_{ij}^k)(\Psi_{jk}-\Psi_{ik})$. This trick is now well accepted in the finite volume literature; see, e.g., Buet and Cordier [4, eq. (10)], Buet and Després [5, eq. (31)], Buet, Després, and Franck [6, eq. (19)], Jin and Levermore [15, sect. 2.6], and Li and Wang [20, eq. (2.4)]. Notice that, in addition to our recasting the technique from [8] into a discretization agnostic framework, two other novelties are our handling of the boundary condition, which is inspired from [2, sect. III.D] and [9, sect. 5.5] and the definitions of h_{ij} and h_i; see (4.5b). □

4.3. Diffusion limit expansion. We investigate the diffusion limit of the formulation (4.6) by proceeding as in section 3.2. We rescale the problem as in (3.5) by replacing $\sigma_{s,ij}$, $\sigma_{s,i}$, $\sigma_{a,i}$, and q_{ik} by $\frac{1}{\epsilon}\sigma_{s,ij}$, $\frac{1}{\epsilon}\sigma_{s,i}$, $\epsilon\sigma_{a,i}$, and ϵq_{ik}, respectively. The discrete problem consists of seeking $\boldsymbol{\psi}_h^\epsilon$ such that the following holds true for all $(i,k) \in \mathcal{V}{\times}\mathcal{L}$:

$$\begin{aligned}(4.7)\qquad \sum_{j\in\mathcal{I}(i)\backslash\{i\}} &\frac{\epsilon}{\sigma_{s,ij}h_{ij}}\frac{1}{1+\frac{\epsilon}{\sigma_{s,ij}h_{ij}}}(\boldsymbol{\Omega}_k{\cdot}\boldsymbol{c}_{ij} - d_{ij}^k)(\Psi_{jk}^\epsilon - \Psi_{ik}^\epsilon) + \epsilon m_i\sigma_{a,i}\Psi_{ik}^\epsilon \\ &= \epsilon m_i q_{ik} + \frac{m_i}{h_i}\frac{1}{1+\frac{\epsilon}{\sigma_{s,i}h_i}}\left(-\Psi_{ik}^\epsilon + \overline{\Psi}_i^\epsilon\right) + \frac{\epsilon}{\sigma_{s,i}h_i}\frac{1}{1+\frac{\epsilon}{\sigma_{s,i}h_i}}b_{ik}^\partial(\beta_{ik}^\partial - \Psi_{ik}^\epsilon)\end{aligned}$$

with $\beta_{ik}^\partial := \theta_i^\epsilon\alpha_{ik}^\partial + (1-\theta_i^\epsilon)(\frac{1}{2}\mathfrak{m}_i^\partial - 3\boldsymbol{\mathfrak{M}}_i^\partial{\cdot}\boldsymbol{n}_i)$, $\theta_i^\epsilon := \max(1 - 2\frac{\sigma_{s,i}}{\epsilon}h_i, 0)$.

THEOREM 4.4 (diffusion limit). *Let $\boldsymbol{\psi}_h^\epsilon$ be the solution of the linear system* (4.7). *Assume that the mesh family $(\mathcal{T}_h)_{h>0}$ is such $c_{ij}^{\mathrm{d}} < 0$ for all $i \in \mathcal{V}, j \in \mathcal{I}(i)\backslash\{i\}$. Let $\boldsymbol{\psi}_h^0 = \lim_{\epsilon\to 0}\boldsymbol{\psi}_h^\epsilon$. Then $\boldsymbol{\psi}_h^0$ is isotropic, i.e., $\boldsymbol{\psi}_h^0 = (\psi_h^0,\dots,\psi_h^0)$, and for all $i \in \mathcal{V}$ the scalar field $\psi_h^0 := \sum_{j\in\mathcal{V}}\Psi_j^0\varphi_j$ solves*

$$(4.8)\qquad a(\psi_h^0,\varphi_i) + m_i\sigma_{a,i}\overline{\Psi}_i^0 + \frac{m_i^\partial}{\sigma_{s,i}h_i}\delta_i^\partial\overline{\Psi}_i^0 = m_i\overline{q}_i + \frac{m_i^\partial}{\sigma_{s,i}h_i}\delta_i^\partial\Big(\frac{\mathfrak{m}_i^\partial}{2} - 3\boldsymbol{\mathfrak{M}}_i^\partial{\cdot}\boldsymbol{n}_i\Big).$$

Moreover, setting $\boldsymbol{J}_i^\epsilon := \frac{1}{\epsilon|\mathcal{S}|}\sum_{k\in\mathcal{L}}\mu_k\boldsymbol{\Omega}_k\Psi_{ik}^\epsilon$, and $\boldsymbol{J}_i^0 := \lim_{\epsilon\to 0}\boldsymbol{J}_i^\epsilon$, the vector $\boldsymbol{J}_i^\epsilon$ satisfies the following consistent approximation of Fick's law for all $i \in \mathcal{V}^\circ$:

$$(4.9)\qquad m_i\boldsymbol{J}_i^0 = -\sum_{j\in\mathcal{I}(i)\backslash\{i\}}\frac{h_i}{h_{ij}}\frac{1}{3\sigma_{s,ij}}\boldsymbol{c}_{ij}(\Psi_j^0 - \Psi_i^0).$$

Proof. A rigorous functional analytic argument can be made by proceeding as in [9, sect. 4], but since the mesh-size is fixed and the approximation space is finite-dimensional, there is no fundamental obstacle to proceed formally; hence, we consider the asymptotic expansion $\boldsymbol{\psi}_h = \boldsymbol{\psi}_h^0 + \epsilon\boldsymbol{\psi}_h^1 + \epsilon^2\boldsymbol{\psi}_h^2 + \mathcal{O}(\epsilon^3)$.

Proof of (4.8). Notice first that $\theta_i^\epsilon = 0$ for all $\epsilon \le 2\sigma_{s,i}h_i$; hence, $\beta_{ik}^\partial = \beta_i^\partial := \frac{1}{2}\mathfrak{m}_i^\partial - 3\boldsymbol{\mathfrak{M}}_i^\partial{\cdot}\boldsymbol{n}_i$. Using that $\frac{1}{1+\frac{\epsilon}{\sigma h}} = 1 - \frac{\epsilon}{\sigma h} + \mathcal{O}(\epsilon^2)$, we have

$$\begin{aligned}\sum_{j\in\mathcal{I}(i)\backslash\{i\}} &\frac{\epsilon}{\sigma_{s,ij}h_{ij}}(\boldsymbol{\Omega}_k{\cdot}\boldsymbol{c}_{ij} - d_{ij}^k)(\Psi_{jk} - \Psi_{ik}) + \epsilon m_i\sigma_{a,i}\Psi_{ik} \\ &= \epsilon m_i q_{ik} + \frac{m_i}{h_i}(1 - \frac{\epsilon}{\sigma_{s,i}h_i})\left(-\Psi_{ik} + \overline{\Psi}_i\right) + \frac{\epsilon}{\sigma_{s,i}h_i}b_{ik}^\partial(\beta_i^\partial - \Psi_{ik}) + \mathcal{O}(\epsilon^2).\end{aligned}$$

Inserting now the formal asymptotic expansion $\boldsymbol{\psi}_h = \boldsymbol{\psi}_h^0 + \epsilon\boldsymbol{\psi}_h^1 + \mathcal{O}(\epsilon^2)$ into this equation, we infer that $\overline{\Psi}_i^0 - \Psi_{ik}^0 = 0$ for all $(i,k) \in \mathcal{V}{\times}\mathcal{L}$ and

$$\begin{aligned}(4.10)\qquad \sum_{j\in\mathcal{I}(i)\backslash\{i\}} &\frac{1}{\sigma_{s,ij}h_{ij}}(\boldsymbol{\Omega}_k{\cdot}\boldsymbol{c}_{ij} - d_{ij}^k)(\Psi_{jk}^0 - \Psi_{ik}^0) + m_i\sigma_{a,i}\Psi_{ik}^0 \\ &= m_i q_{ik} + \frac{m_i}{h_i}\left(-\Psi_{ik}^1 + \overline{\Psi}_i^1\right) + \frac{1}{\sigma_{s,i}h_i}b_{ik}^\partial(\beta_i^\partial - \Psi_{ik}^0).\end{aligned}$$

Taking the (weighted) average of the second equation over the discrete ordinates, we obtain

$$\sum_{j\in\mathcal{I}(i)\backslash\{i\}} (\overline{\Psi}^0_j - \overline{\Psi}^0_i)\frac{1}{\sigma_{s,ij}h_{ij}}\frac{1}{|\mathcal{S}|}\sum_{k\in\mathcal{L}} -\mu_k d^k_{ij} + m_i\sigma_{a,i}\overline{\Psi}^0_i = m_i\overline{q}_i + \frac{m^\partial_i}{\sigma_{s,i}h_i}(\delta^\partial_i\beta^\partial_i - \delta^\partial_i\overline{\Psi}^0_i).$$

(Recall that $\delta^\partial_i \approx \frac{1}{4}$). Now we use the definition of h_{ij} (see (4.5b)) and recall that the mesh family $(\mathcal{T}_h)_{h>0}$ is assumed to be such that $c^{\mathrm{d}}_{ij} < 0$ for all $i \in \mathcal{V}$, $j \in \mathcal{I}(i)\backslash\{i\}$; then we obtain

$$\frac{m^\partial_i}{\sigma_{s,i}h_i}\delta^\partial_i\overline{\Psi}^0_i + \sum_{j\in\mathcal{I}(i)\backslash\{i\}}\frac{1}{3}c^{\mathrm{d}}_{ij}(\overline{\Psi}^0_j - \overline{\Psi}^0_i) + m_i\sigma_{a,i}\overline{\Psi}^0_i = m_i\overline{q}_i + \frac{m^\partial_i}{\sigma_{s,i}h_i}\delta^\partial_i\beta^\partial_i.$$

Now using the partition of unity property, i.e., $\sum_{j\in\mathcal{I}(i)} c^{\mathrm{d}}_{ij} = 0$, and recalling the definition of β^∂_i, we infer that

$$\frac{m^\partial_i}{\sigma_{s,i}h_i}\delta^\partial_i\overline{\Psi}^0_i + a(\nabla\psi^0_h,\varphi_i) + m_i\sigma_{a,i}\overline{\Psi}^0_i = m_i\overline{q}_i + \frac{m^\partial_i}{\sigma_{s,i}h_i}\delta^\partial_i(\frac{1}{2}\mathfrak{m}^\partial_i - 3\boldsymbol{\mathfrak{M}}^\partial_i\cdot\boldsymbol{n}_i).$$

Proof of (4.9). Since $\boldsymbol{\psi}^0_h$ is isotropic, we have

$$\boldsymbol{J}^\epsilon_i := \frac{1}{\epsilon|\mathcal{S}|}\sum_{k\in\mathcal{L}}\mu_k\boldsymbol{\Omega}_k\Psi^\epsilon_{ik} = \frac{1}{|\mathcal{S}|}\sum_{k\in\mathcal{L}}\mu_k\boldsymbol{\Omega}_k\Psi^1_{ik} + \mathcal{O}(\epsilon).$$

That is, $\boldsymbol{J}^0_i := \lim_{\epsilon\to 0}\boldsymbol{J}^\epsilon_i = \frac{1}{|\mathcal{S}|}\sum_{k\in\mathcal{L}}\mu_k\boldsymbol{\Omega}_k\Psi^1_{ik}$. We now multiply (4.10) by $\boldsymbol{\Omega}_k$, take the (weighted) average over the discrete ordinates, and recall that the angular quadrature satisfies $\sum_{k\in\mathcal{L}}\mu_k\boldsymbol{\Omega}_k|\boldsymbol{n}\cdot\boldsymbol{\Omega}_k| = \boldsymbol{0}$ for all $\boldsymbol{n}\in\mathbb{R}^3$:

$$\sum_{j\in\mathcal{I}(i)\backslash\{i\}}\frac{1}{3\sigma_{s,ij}h_{ij}}\boldsymbol{c}_{ij}(\Psi^0_j - \Psi^0_i) - \sum_{j\in\mathcal{I}(i)\backslash\{i\}}\frac{(\Psi^0_j - \Psi^0_i)}{\sigma_{s,ij}h_{ij}}\sum_{k\in\mathcal{L}}\frac{\mu_k}{|\mathcal{S}|}\boldsymbol{\Omega}_k d^k_{ij}$$
$$= -\frac{m_i}{h_i}\boldsymbol{J}^0_i + \frac{m^\partial_i}{\sigma_{s,i}h_i}\frac{1}{6}(\beta^\partial_i - \Psi^0_i)\boldsymbol{n}_i,$$

where we used that $\frac{1}{|\mathcal{S}|}\sum_{k\in\mathcal{L}^-_i}|\boldsymbol{\Omega}_k\cdot\boldsymbol{c}|\boldsymbol{\Omega}_k = \frac{1}{6}\boldsymbol{c}$ for any $\boldsymbol{c}\in\mathbb{R}^3$. If $i\in\mathcal{V}^\circ$, then $d^k_{ij} = |\boldsymbol{c}_{ij}\cdot\boldsymbol{\Omega}_k|$, which in turn implies that $\sum_{k\in\mathcal{L}}\mu_k\boldsymbol{\Omega}_k d^k_{ij} = 0$. The assertion follows readily. □

Remark 4.5 (limit problem and boundary conditions). Since h_i behaves like the mesh-size, h, the discrete problem (4.8) is a weak formulation with a penalty on the boundary condition scaling like h^{-1}. The continuous problem associated with the discrete problem (4.8) consists of seeking $\psi^{\lim}\in H^1(D)$ so that

$$(4.11)\qquad -\nabla\cdot\Big(\frac{1}{3\sigma_s}\nabla\psi^{\lim}\Big) + \sigma_a\psi^{\lim} = \overline{q}, \qquad \psi^{\lim}_{|\partial D} = \frac{1}{2}\mathfrak{m}^\partial - 3\boldsymbol{\mathfrak{M}}^\partial\cdot\boldsymbol{n}.$$

This result is coherent with [9, Thm. 5.4]. Recall that in general $\psi^{\lim}\neq\psi^0$ unless $\frac{1}{2}\mathfrak{m}^\partial + 3\boldsymbol{\mathfrak{M}}^\partial\cdot\boldsymbol{n} = 0$; see [9, sect. 5.5]. □

Remark 4.6 (Fick's law). Let us now interpret (4.9). Assume that the mesh is uniform or quasi-uniform in the neighborhood of the Lagrange node $\boldsymbol{a}_i$; then $h_i\approx h_{ij}$ and $\sigma_{s,ij}\approx\sigma_{s,i}$. Hence, $m_i\boldsymbol{J}^0_i \approx -\frac{1}{3\sigma_{s,i}}\sum_{j\in\mathcal{I}(i)}\boldsymbol{c}_{ij}\Psi^0_j$. Owing to the definition of the coefficients $\boldsymbol{c}_{ij}$, this equation is a consistent approximation of Fick's law $\boldsymbol{J} = -\frac{1}{3\sigma_s}\nabla\psi^0$. □

Remark 4.7 (meshes). It is known for simplicial meshes and piecewise linear continuous finite elements that a sufficient condition for the inequality $c^{\mathrm{g,d}}_{ij} < 0$ to hold for all $i \in \mathcal{V}$, $j \in \mathcal{I}(i)\backslash\{i\}$ is that the mesh family $(\mathcal{T}_h)_{h>0}$ satisfies the so-called acute angle condition; see, e.g., Xu and Zikatanov [26, eq. (2.5)]. □

4.4. Positivity. We establish in this section the positivity of the method defined in (4.6) using the definitions in (4.5). We set $\Psi^{\min} := \min_{(j,l)\in\mathcal{V}\times\mathcal{L}} \Psi_{j,l}$ and $\Psi^{\max} := \max_{(j,l)\in\mathcal{V}\times\mathcal{L}} \Psi_{j,l}$.

THEOREM 4.8 (minimum/maximum principle). *Let* $(\Psi_{ik})_{(i,k)\in\mathcal{V}\times\mathcal{L}}$ *be the solution to* (4.6) *with* d^k_{ij} *and all the other parameters defined in* (4.5a)–(4.5b). *Let* (i_0,k_0), $(i_1,k_1) \in \mathcal{V}\times\mathcal{L}$ *be such that* $\Psi_{i_0k_0} = \Psi^{\min}$ *and* $\Psi_{i_1k_1} = \Psi^{\max}$.

(i) *Assume that* $\min_{(j,l)\in\mathcal{V}\times\mathcal{L}}(\sigma_{a,j} + b^{\partial}_{jl}) > 0$. *Then*

$$\frac{m_{i_0}q_{i_0k_0} + \frac{b^{\partial}_{i_0k_0}}{\sigma_{s,i_0}h_{i_0}+1}\beta^{\partial}_{i_0k_0}}{m_{i_0}\sigma_{a,i_0} + \frac{b^{\partial}_{i_0k_0}}{\sigma_{s,i_0}h_{i_0}+1}} \le \Psi^{\min} \le \Psi^{\max} \le \frac{m_{i_1}q_{i_1k_1} + \frac{b^{\partial}_{i_1k_1}}{\sigma_{s,i_1}h_{i_1}+1}\beta^{\partial}_{i_1k_1}}{m_{i_1}\sigma_{a,i_1} + \frac{b^{\partial}_{i_1k_1}}{\sigma_{s,i_1}h_{i_1}+1}}. \tag{4.12}$$

(ii) *Otherwise, assume that for all* $i \in \mathcal{V}$ *such that* $\sigma_{a,i} = 0$ *and* $b^{\partial}_{ik} = 0$ *the definition of* d^k_{ij} *is slightly modified so that* $\boldsymbol{\Omega}_k\cdot\boldsymbol{c}_{ij} < d^k_{ij}$ *for all* $j \in \mathcal{I}(i)$ *(instead of* $\boldsymbol{\Omega}_k\cdot\boldsymbol{c}_{ij} \le d^k_{ij}$*). If* $0 \le \min_{(i,k)\in\mathcal{V}\times\mathcal{L}} q_{ik}$ *and* $0 \le \min_{(i,k)\in(\mathcal{V}\times\mathcal{L})^{\partial}} \alpha^{\partial}_{ik}$, *then* $0 \le \Psi^{\min}$.

(iii) *Moreover, under the same assumptions on* d^k_{ij} *as in* (ii), *if* $\max_{(i,k)\in\mathcal{V}\times\mathcal{L}} q_{ik} \le 0$, *then* $\Psi^{\max} \le \max_{(i,k)\in(\mathcal{V}\times\mathcal{L})^{\partial}} \beta^{\partial}_{ik}$

Proof. We proceed as in the proof of Lemma 3.1. We start with the proof of (i) and assume that $\min_{(j,l)\in\mathcal{V}\times\mathcal{L}}(\sigma_{a,j} + b^{\partial}_{jl}) > 0$. Let $(i_0,k_0) \in \mathcal{V}\times\mathcal{L}$ be the indices of the degree of freedom where the minimum is attained; that is, $\Psi_{ik} \ge \Psi_{i_0k_0}$ for all $(i,k) \in \mathcal{V}\times\mathcal{L}$. Then using that $\boldsymbol{\Omega}_k\cdot\boldsymbol{c}_{ij} - d^k_{ij} \le \max(\boldsymbol{\Omega}_k\cdot\boldsymbol{c}_{ij}, 0) - d^k_{ij} \le 0$, together with $\Psi_{jk_0} - \Psi_{i_0k_0} \ge 0$ for all $j \in \mathcal{I}(i_0)$, and $\Psi_{i_0k_0} \le \overline{\Psi}_{i_0}$, we infer that

$$\begin{aligned} m_{i_0}q_{i_0k_0} + \frac{b^{\partial}_{i_0k_0}}{\sigma_{s,i_0}h_{i_0}+1}\beta^{\partial}_{i_0k_0} &= \sum_{j\in\mathcal{I}(i_0)\backslash\{i_0\}} \frac{\boldsymbol{\Omega}_{k_0}\cdot\boldsymbol{c}_{i_0j} - d^{k_0}_{i_0j}}{\sigma_{s,i_0j}h_{i_0j}+1}(\Psi_{jk_0} - \Psi_{i_0k_0}) \\ &\quad + \frac{m_{i_0}\sigma_{s,i_0}}{\sigma_{s,i_0}h_{i_0}+1}(\Psi_{i_0k_0} - \overline{\Psi}_{i_0}) + m_{i_0}\sigma_{a,i_0}\Psi_{i_0k_0} + \frac{b^{\partial}_{i_0k_0}}{\sigma_{s,i_0}h_{i_0}+1}\Psi_{i_0k_0} \\ &\le m_{i_0}\sigma_{a,i_0}\Psi_{i_0k_0} + \frac{b^{\partial}_{i_0k_0}}{\sigma_{s,i_0}h_{i_0}+1}\Psi_{i_0k_0}. \end{aligned}$$

Hence $m_{i_0}q_{i_0k_0} + \frac{b^{\partial}_{i_0k_0}}{\sigma_{s,i_0}h_{i_0}+1}\beta^{\partial}_{i_0k_0} \le (m_{i_0}\sigma_{a,i_0} + \frac{b^{\partial}_{i_0k_0}}{\sigma_{s,i_0}h_{i_0}+1})\Psi_{i_0k_0}$. The assertion follows readily. The proof of the other assertion regarding $\Psi^{\max}$ is analogous.

Proof of (ii) assuming that $0 \le \min_{(i,k)\in\mathcal{V}\times\mathcal{L}} q_{ik}$ and $0 \le \min_{(i,k)\in(\mathcal{V}\times\mathcal{L})^{\partial}} \alpha^{\partial}_{ik}$. From part (i) we conclude that we need to prove $\Psi_{i_0k_0} \ge 0$ only in the case $\sigma_{a,i_0} = 0$ and $b^{\partial}_{i_0k_0} = 0$. Assuming that $\sigma_{a,i_0} = 0$ and $b^{\partial}_{i_0k_0} = 0$, we have from part (i) the following inequality:

$$0 \le m_{i_0}q_{i_0k_0} \le \sum_{j\in\mathcal{I}(i_0)\backslash\{i_0\}} \frac{\boldsymbol{\Omega}_{k_0}\cdot\boldsymbol{c}_{i_0j} - d^{k_0}_{i_0j}}{\sigma_{s,i_0j_0}h_{i_0j}+1}(\Psi_{jk_0} - \Psi_{i_0k_0}) \le 0.$$

The assumption $\boldsymbol{\Omega}_k\cdot\boldsymbol{c}_{i_0j} - d^{k_0}_{i_0j} < 0$ for all $j \in \mathcal{I}(i_0)$ implies that $\Psi_{jk_0} - \Psi_{i_0k_0} = 0$ for all $j \in \mathcal{I}(i_0)$. Therefore, we conclude that the global minimum is attained not only at the degree of freedom (i_0, k_0) but also in the whole neighborhood, i.e., for all $j \in \mathcal{I}(i_0)$. Repeating the above argument for a global minimum at (j, k_0) for all $j \in \mathcal{I}(i_0)$, we derive that the global minimum is either nonnegative (if $m_j\sigma_{a,j} + b^\partial_{jk_0} > 0$) or again attained in the whole neighborhood of j, i.e., for all $s \in \mathcal{I}(j)$. This process can terminate in two ways: (i) either the global minimum is nonnegative at some j because $m_j\sigma_{a,j} + b^\partial_{jk_0} > 0$ or (ii) the global minimum is attained at all of the degrees of freedom topologically connected to i_0. In this case we have that $\Psi_{jk_0} = \Psi_{i_0k_0}$ for all j in the same connected component as i_0, which is the entire set $\mathcal{V}$ since $\mathcal{T}_h$ is connected (because D is connected). However, for any fixed k_0 there exists j such that Ψ_{jk_0} is on the inflow boundary for $\boldsymbol{\Omega}_{k_0}$. That is, we have $b^\partial_{jk_0} > 0$, and conclude (see (4.6)) that $\Psi_{i_0k_0} \geq 0$ because $\beta^\partial_{ij} = \theta_i\alpha^\partial_{ik} + (1-\theta_i)(\frac{1}{2}\mathfrak{m}^\partial_i - 3\boldsymbol{\mathfrak{M}}^\partial_i{\cdot}\boldsymbol{n}_i) \geq 0$ on the the inflow boundary (notice that $\mathfrak{m}^\partial_i \geq 0$ and $\boldsymbol{\mathfrak{M}}^\partial_i{\cdot}\boldsymbol{n}_i \leq 0$).

Proof of (iii) assuming that $\min_{(i,k)\in\mathcal{V}\times\mathcal{L}} q_{ik} \leq 0$. By proceeding as in step (i), we infer that

$$m_{i_1}q_{i_1k_1} + \frac{b^\partial_{i_1k_1}}{\sigma_{s,i_1}h_{i_1}+1}(\beta^\partial_{i_1k_1} - \Psi_{i_1k_1}) \\ \geq \sum_{j\in\mathcal{I}(i_1)\setminus\{i_1\}} \frac{\boldsymbol{\Omega}_{k_1}{\cdot}\boldsymbol{c}_{i_1j} - d^{k_1}_{i_1j}}{\sigma_{s,i_1j}h_{i_1j}+1}(\Psi_{jk_1} - \Psi_{i_1k_1}) + m_{i_1}\sigma_{a,i_1}\Psi_{i_1k_1} \geq 0,$$

i.e., $(m_{i_1}\sigma_{a,i_1} + \frac{b^\partial_{i_1k_1}}{\sigma_{s,i_1}h_{i_1}+1})\Psi_{i_1k_1} \leq m_{i_1}q_{i_1k_1} + \frac{b^\partial_{i_1k_1}}{\sigma_{s,i_1}h_{i_1}+1}\beta^\partial_{i_1k_1}$, which implies $\Psi_{i_1k_1} \leq \beta^\partial_{i_1k_1}$ if $b^\partial_{i_1k_1} > 0$. Hence we just need to consider the case $b^\partial_{i_1k_1} = 0$. In that case $0 \geq \sum_{j\in\mathcal{I}(i_1)\setminus\{i_1\}} \frac{\boldsymbol{\Omega}_{k_1}\cdot\boldsymbol{c}_{i_1j} - d^{k_1}_{i_1j}}{\sigma_{s,i_1j}h_{i_1j}+1}(\Psi_{jk_1} - \Psi_{i_1k_1}) \geq 0$ and $\Psi_{jk_1} = \Psi_{i_1k_1}$ for all $j \in \mathcal{I}(i_1)$. Then we proceed as in step (ii) until we reach a degree of freedom j that is on the inflow boundary for $\boldsymbol{\Omega}_{k_1}$, i.e., $b^\partial_{jk_1} > 0$. Then $\Psi^{\max} = \Psi_{i_1k_1} = \Psi_{jk_1} \leq \beta^\partial_{j,k_1}$. □

5. Numerical illustrations.

We present in this section numerical results to illustrate the positive- and asymptotic-preserving algorithm (4.6) described in section 4.2. We compare this technique in various regimes with the standard dG1 technique using the upwind flux.

5.1. Numerical details.

The positive- and asymptotic-preserving algorithm defined in (4.6) is implemented with piecewise linear continuous finite elements on simplices. We use the same code for one-dimensional and two-dimensional tests. The meshes in one dimension are uniform. The meshes in two space dimension are nonuniform, are composed of triangles, and have the Delaunay property. Nothing special is done to make the triangulations satisfy the acute angle condition, i.e., the condition may not be satisfied for a few pairs of vertices. In one dimension we use the Gauss–Legendre quadrature for the angular discretization: the x_1-component of the angles are the quadrature points of the Gaussian quadrature over $[-1, 1]$, and the weights are the weights of the Gaussian quadrature. In two dimensions we use the standard triangular S_N quadrature (Gauss–Legendre quadrature along the polar axis and equidistributed angles along the azimuth with $\frac{1}{8}N(N+2)$ angles per octant). Since the size of the problems involved here is small (at most $2{\times}10^6$ degrees of freedom), we assemble the sparse matrix defined in (4.6) using the compressed sparse row format and solve it using Pardiso (see, e.g., Petra, Schenk, and Anitescu [22]). More sophisticated techniques involving source iterations and synthetic acceleration could be used for significantly larger systems. We do not discuss this issue since it is out of the scope of the paper.

In order to assess the asymptotic-preserving approach, we compare it against a state-of-the-art technique. More specifically, (2.1) is solved using dG1 with the upwind numerical flux and the same triangular S_N quadrature as above. The linear system is solved by iterating on the scattering source (see, e.g., Adams and Larsen [1]); for instance, starting with some guess $\overline{\boldsymbol{\psi}}_h^{(0)}$, one constructs a sequence $\boldsymbol{\psi}_h^{(0)},\dots,\boldsymbol{\psi}_h^{(\ell)},\dots$ Given some state $\boldsymbol{\psi}_h^{(\ell)}$ we compute an intermediate state $\boldsymbol{\psi}_h^{(\ell+\frac{1}{2})}$ such that

$$\sum_{j\in\mathcal{I}(i)} A^k_{ij}\Psi^{(\ell+\frac{1}{2})}_{jk} + m_i\sigma_{t,i}\Psi^{(\ell+\frac{1}{2})}_{ik} + b^{\partial}_{ik}\Psi^{(\ell+\frac{1}{2})}_{ik} = m_i\sigma_{s,i}\overline{\Psi}^{(\ell)}_i + m_i q_{ik} + b^{\partial}_{ik}\alpha^{\partial}_{ik}, \tag{5.1a}$$

$$A^k_{ij} := \begin{cases} \displaystyle\int_K (\boldsymbol{\Omega}_k\cdot\nabla\varphi_j)\varphi_i \,\mathrm{d}\boldsymbol{x}, & j\in\mathcal{I}(K)\backslash\mathcal{I}(\partial K^{\mathrm{i}}),\\ \displaystyle\int_K (\boldsymbol{\Omega}_k\cdot\nabla\varphi_j)\varphi_i \,\mathrm{d}\boldsymbol{x} + \int_{\partial K}\varphi_i\varphi_j(\boldsymbol{\Omega}_k\cdot\boldsymbol{n}_K)_-\,\mathrm{d}\boldsymbol{x}, & j\in\mathcal{I}(\partial K^{\mathrm{i}}),\\ \displaystyle -\int_{\partial_K}\frac{|\boldsymbol{\Omega}_k\cdot\boldsymbol{n}_K|-\boldsymbol{\Omega}_k\cdot\boldsymbol{n}_K}{2}\varphi_i\varphi_j\,\mathrm{d}\boldsymbol{x}, & j\in\mathcal{I}(\partial K^{\mathrm{e}}) \end{cases} \tag{5.1b}$$

with $z_- := \frac{1}{2}(|z|-z)$. For each direction k, (5.1a) is solved cell-by-cell by sweeping through the mesh from the inflow boundary to the outflow boundary defined by the angle $\boldsymbol{\Omega}_k$ (a process termed "transport sweep" in the radiation transport community). Without synthetic acceleration, we set $\boldsymbol{\psi}_h^{(\ell+1)} = \boldsymbol{\psi}_h^{(\ell+1/2)}$, and the new source iteration ($\ell \leftarrow \ell+1$) can proceed. However, in highly diffusive configurations, a diffusion synthetic accelerator is invoked to compute a correction $\delta\boldsymbol{\psi}_h^{\ell+1}$ to improve the scalar flux iterate; at the end of the process we set $\boldsymbol{\psi}_h^{(\ell+1)} = \boldsymbol{\psi}_h^{(\ell+1/2)} + \delta\boldsymbol{\psi}_h^{\ell+1}$. Here, we use a dG compatible diffusion synthetic accelerator based on an interior penalty technique; see, e.g., Wang and Ragusa [25] for additional details.

5.2. Manufactured solution.

We first test our piecewise linear, continuous finite element implementation of the algorithm described in section 4.2 on a manufactured solution. The domain is $D=(0,1)^2\times\mathbb{R}$ with $\sigma_t=\sigma_s=1$, and the solution is $\boldsymbol{\psi} := (\psi_1,\dots,\psi_L)$ with

$$\psi_k(\boldsymbol{x}) = 2+\sin(\boldsymbol{\Omega}_k\cdot\boldsymbol{x})+\sin(\pi x_1)\sin(\pi x_2), \tag{5.2}$$

where $k\in\mathcal{L}$, $\boldsymbol{x} := (x_1,x_2)\in D$. The source term $q(\boldsymbol{x},\boldsymbol{\Omega}_k)$ is computed accordingly with $\overline{\psi}(\boldsymbol{x}) := \frac{1}{|\mathcal{S}|}\sum_{k\in\mathcal{K}}\psi_k(\boldsymbol{x})$.

The relative errors in the L^2-norm, L^∞-norm, and H^1-seminorm are calculated on five nonuniform meshes composed of triangles with 140, 507, 1927, 7545, and 29870 Lagrange nodes, respectively; the corresponding mesh-sizes are approximately $h\approx 0.1$, 0.5, 0.025, 0.125, and 0.00625. We define the error $\boldsymbol{e} := (e_1,\dots,e_L)$ with $e_k := \psi_{h,k} - \Pi_h^{\mathrm{L}}(\psi_k)$, where $\Pi_h^{\mathrm{L}}(\psi_k)$ is the Lagrange interpolant of ψ_k in $P^{\mathrm{g}}(\mathcal{T}_h)$, and we set

$$\|\boldsymbol{e}\|_{L^2}^2 = \sum_{k\in\mathcal{L}}\mu_k\|e_k\|^2_{L^2(D)}, \quad \|\boldsymbol{e}\|_{L^\infty} = \max_{k\in\mathcal{L}}\|e_k\|_{L^\infty(D)}. \tag{5.3}$$

The relative errors are denoted and defined as follows: $\mathrm{rel}(\|\boldsymbol{e}\|_{L^2}) = \|\boldsymbol{e}\|_{L^2}/\|\boldsymbol{\psi}\|_{L^2}$, $\mathrm{rel}(\|\boldsymbol{e}\|_{L^\infty}) = \|\boldsymbol{e}\|_{L^\infty}/\|\boldsymbol{\psi}\|_{L^\infty}$, $\mathrm{rel}(\|\nabla\boldsymbol{e}\|_{\boldsymbol{L}^2}) = \|\nabla\boldsymbol{e}\|_{\boldsymbol{L}^2}/\|\nabla\boldsymbol{\psi}\|_{\boldsymbol{L}^2}$. The results for the S_6 and S_{10} quadratures are reported in Table 1. We observe that, as expected, the method is first-order accurate in space in the L^2-norm, and it is $\mathcal{O}(h^{\frac{1}{2}})$ in the L^∞-norm and in the H^1-seminorm. These results are compatible with the best theoretical

Table 1
Convergence tests with respect to mesh-size with solution (5.2) *and quadrature* S_6 *and* S_{10}*.*

	#DOFs	rel($\|\boldsymbol{e}\|_{L^2}$)	Rate	rel($\|\boldsymbol{e}\|_{L^\infty}$)	Rate	rel($\|\nabla\boldsymbol{e}\|_{\boldsymbol{L}^2}$)	Rate
S_6	140	5.20E-02	–	2.89E-01	–	3.07E-01	–
	507	2.70E-02	1.02	2.08E-01	0.51	2.01E-01	0.66
	1927	1.37E-02	1.01	1.48E-01	0.51	1.36E-01	0.59
	7545	6.93E-03	1.00	1.05E-01	0.50	9.38E-02	0.54
	29870	3.48E-03	1.00	7.48E-02	0.50	6.55E-02	0.52
S_{10}	140	5.19E-02	–	2.91E-01	–	3.07E-01	–
	507	2.69E-02	1.02	2.08E-01	0.52	2.01E-01	0.66
	1927	1.37E-02	1.01	1.48E-01	0.51	1.37E-01	0.58
	7545	6.93E-03	1.00	1.09E-01	0.45	9.48E-02	0.54
	29870	3.48E-03	1.00	8.22E-02	0.42	6.64E-02	0.52

error estimates known for the approximation of the linear transport equation using first-order viscosities.

5.2.1. Diffusion limit with constant cross sections. We consider the two-dimensional domain $D = (0,1)^2\times\mathbb{R}$ with constant cross sections $\sigma_t = \sigma_s = \frac{1}{\epsilon}$ and source term $q(\boldsymbol{x}) = \epsilon\frac{2}{3}\pi^2\sin(\pi x_1)\sin(\pi x_2)$. The diffusion limit corresponding to $\epsilon \to 0$ is $\psi^0(\boldsymbol{x}) = \sin(\pi x_1)\sin(\pi x_2)$. We solve (2.1) with continuous linear finite elements and the algorithm described in section 4.2. The meshes are nonuniform and composed of triangles. To estimate the convergence we use five meshes with 140, 507, 1927, 7545, and 29870 Lagrange nodes, respectively; the corresponding mesh-sizes are approximately $h \approx 0.1$, 0.5, 0.025, 0.125, and 0.00625. We use the S_6 angular quadrature.

The results for $\epsilon \in \{10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\}$ are reported in Table 2. We show in this table the relative L^2-norm and the relative H^1-seminorm of the difference $\overline{\boldsymbol{\psi}}_h - \Pi_h^{\mathrm{L}}(\boldsymbol{\psi}^0)$, where $\Pi_h^{\mathrm{L}}(\boldsymbol{\psi}^0)$ is the Lagrange interpolant of $\boldsymbol{\psi}^0$. We clearly observe that, just like proved in [9, Thm. 5.3] for the upwind dG1 approximation, the scalar flux $\overline{\boldsymbol{\psi}}_h$ converges optimally to $\boldsymbol{\psi}_h^0$ when ϵ is significantly smaller than the mesh-size. The convergence order is $\mathcal{O}(h^2)$ in the L^2-norm. It seems that some supercloseness phenomenon occurs in the H^1-seminorm since $\|\nabla(\overline{\boldsymbol{\psi}}_h - \Pi_h^{\mathrm{L}}(\boldsymbol{\psi}^0))\|_{\boldsymbol{L}^2}$ converges like $\mathcal{O}(h^{1.5})$.

5.3. One-dimensional results. We now perform four one-dimensional tests and compare the positive- asymptotic-preserving method (with piecewise linear continuous finite elements) with the upwind dG1 approximation. We use the S_8 angular quadrature (8 discrete directions in one dimension) for all the cases. The angles are

Table 2
Convergence test on $\boldsymbol{e} := \overline{\boldsymbol{\psi}}_h - \Pi_h^{\mathrm{L}}(\boldsymbol{\psi}^0)$ *with respect to the mesh-size and* ϵ*.*

ϵ	#DOFs	rel($\|\boldsymbol{e}\|_{L^2}$)	Rate	rel($\|\nabla\boldsymbol{e}\|_{\boldsymbol{L}^2}$)	Rate
10^{-3}	140	2.01E-02	–	9.68E-03	–
	507	2.15E-03	2.34	8.00E-03	1.44
	1927	2.91E-03	-.45	6.62E-03	0.28
	7545	3.11E-03	-.10	7.75E-03	-.23
	29870	3.17E-03	-.03	8.84E-03	-.19
10^{-4}	140	1.92E-02	–	1.20E-02	–
	507	2.87E-03	2.22	5.85E-03	1.85
	1927	5.43E-04	2.49	2.16E-03	1.49
	7545	2.01E-04	1.45	1.21E-03	0.85
	29870	2.53E-04	-.33	1.33E-03	-.13

ϵ	#DOFs	rel($\|\boldsymbol{e}\|_{L^2}$)	Rate	rel($\|\nabla\boldsymbol{e}\|_{\boldsymbol{L}^2}$)	Rate
10^{-5}	140	1.92E-02	–	1.22E-02	–
	507	3.12E-03	2.12	5.76E-03	1.87
	1927	7.59E-04	2.12	1.99E-03	1.59
	7545	1.72E-04	2.18	7.17E-04	1.49
	29870	3.28E-05	2.41	2.73E-04	1.40
10^{-6}	140	1.91E-02	–	1.22E-02	–
	507	3.14E-03	2.11	5.75E-03	1.87
	1927	7.84E-04	2.08	1.98E-03	1.60
	7545	1.93E-04	2.06	7.07E-04	1.51
	29870	4.64E-05	2.07	2.35E-04	1.60

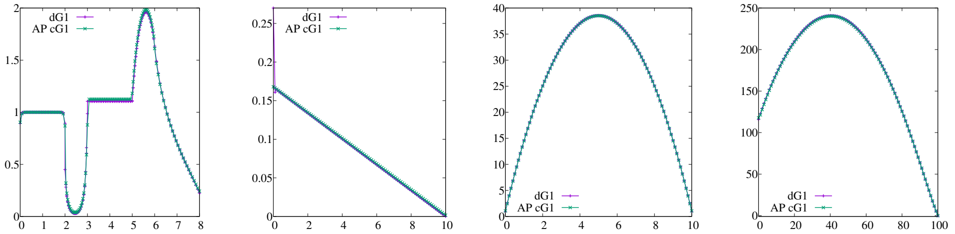
TABLE 3
Data for the one-dimensional test cases.

Case 1	#Zones	5				
	Length	2.0	1.0	2.0	1.0	2.0
	σ_s	0.0	0.0	0.0	0.9	0.9
	σ_t	50.0	5.0	0.0	1.0	1.0
	q	50.0	0.0	0.0	1.0	1.0
	#DOFs	25	25	25	25	25
	B.C.	Vac.				

Case 2	#Zones	1
	Length	10.0
	σ_s	100.0
	σ_t	100.0
	q	0.0
	#DOFs	100
	B.C.	$\psi_5(0) = 0$

Case 3	#Zones	1
	Length	10.0
	σ_s	10.0
	σ_t	10.0
	q	0.1
	#DOFs	100
	B.C.	Vac.

Case 4	#Zones	1
	Length	100.0
	σ_s	0.09999
	σ_t	0.1
	q	1.0
	#DOFs	100
	B.C.	Vac.



(a) Case 1, $\overline{\psi}_h$ (b) Case 2, $\overline{\psi}_h$ (c) Case 3, $\overline{\psi}_h$ (d) Case 4, $\psi_{h,1}$

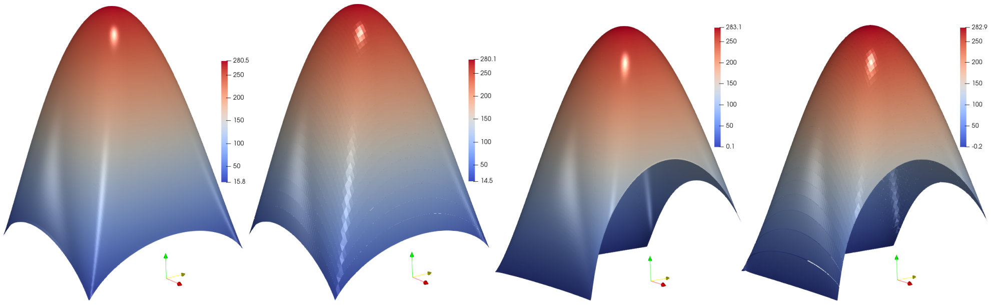
FIG. 1. *Comparison between the (first-order) positive-, asymptotic-preserving cG*1 *method and the (second-order) upwind dG*1 *method.*

enumerated in increasing order from 1 to 8. The data for the four cases are reported in Table 3. The boundary condition (B.C.) for cases 1, 3, and 4 is $\boldsymbol{\psi}_{h|\partial D_-} = 0$ (this is the so-called vacuum boundary condition (Vac.)). The boundary conditions for case 2 are $\boldsymbol{\psi}_{h,k} = 0$ for $k \neq 5$, $1 \leq k \leq 8$, and $\psi_{h,5}(0) = 1.0$.

The results are reported in Figure 1. We show in Figure 1(a)–(c) the scalar flux for the dG1 approximation (labeled dG1) and for the positive and asymptotic-preserving technique (labeled AP cG1). We observe a fair agreement between the two methods given the number of grid points. Figure 1(d) shows the angular flux $\psi_{h,1}$ for case 4. For this case the dG1 approximation gives negatives values at $x = 100$ on the angular fluxes 1, 2, and 3 (the values are -0.24, -0.22, -0.066, respectively (approximated to 2 digits)). In all the cases the asymptotic-preserving technique is always nonnegative.

5.4. Boundary effects. We consider the problem (2.1) in the two-dimensional domain $D = (0, 100)^2 \times \mathbb{R}$ with uniform cross sections $\sigma_t(\boldsymbol{x}) = 0.1$, $\sigma_s(\boldsymbol{x}) = 0.0999$ and uniform source term $q(\boldsymbol{x}, \boldsymbol{\Omega}) = 1$ for all $(\boldsymbol{x}, \boldsymbol{\Omega}) \in D \times \mathcal{S}$. The boundary condition is set to zero $\alpha(\boldsymbol{x}, \boldsymbol{\Omega}) = 0$ for all $(\boldsymbol{x}, \boldsymbol{\Omega}) \in \partial D_-$. (This is the two-dimensional counterpart of the one-dimensional case 4 discussed in section 5.3.) We use the S_6 quadrature for the discrete ordinates (24 directions in two dimensions). The approximation in space for the asymptotic-preserving method is done on a nonuniform grid composed of 151294 triangles with 76160 grid points (i.e., 1 829 520 degrees of freedom in total). The dG1 approximation is done with 64×64 cells, that is, 16384 space degrees of freedom (i.e., 393 216 degrees of freedom in total).

We show in Figure 2 the scalar flux and the angular flux corresponding to the first angle $\boldsymbol{\Omega}_1$. We have verified that the angular fluxes for the asymptotic-preserving method are all nonnegative, as expected, but the upwind dG1 approximation gives negative angular fluxes. In particular, we observe in Figure 2(d) that the minimum value of the first angular flux of the dG1 approximation is equal to -0.2 (1 digit approximation.)

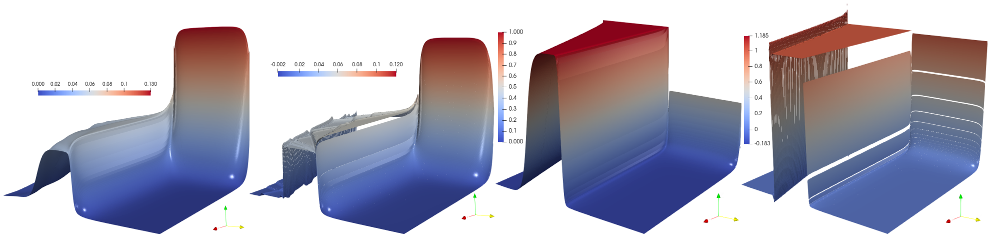


(a) $\overline{\psi}_h$, AP scheme (b) $\overline{\psi}_h$, dG1 scheme (c) $\psi_{h,1}$, AP scheme (d) $\psi_{h,1}$, dG1 scheme

FIG. 2. *Scalar $\overline{\psi}_h$ and angular flux $\psi_{h,1}$.*

5.5. Reflection effects. We now consider the two-dimensional problem with reflection effects. The domain is $D = (0,1)^2 \times \mathbb{R}$ with uniform cross sections $\sigma_t(\boldsymbol{x}) = 100$, $\sigma_s(\boldsymbol{x}) = 99$ if $x_2 \geq 0.5$ (optically thick and diffusive zone), and $\sigma_t(\boldsymbol{x}) = \sigma_s(\boldsymbol{x}) = 0$ if $x_2 \leq 0.5$ (void). We use the S_6 quadrature. The left boundary is illuminated with intensity 1 along the first direction of the quadrature $\boldsymbol{\Omega}_1 := (0.93802334, 0.25134260, 0.23861919)$ (eight digits truncation). The incoming flux is set to 0 along the bottom boundary for $\boldsymbol{\Omega}_1$. For all the other angular fluxes we set $\psi_{h,k|\partial D_-} = 0$, $k \in \mathcal{L}\backslash\{1\}$. The approximation in space for the asymptotic-preserving method is done on a nonuniform grid composed of 151434 triangles with 76230 grid points (i.e., 1 829 520 degrees of freedom in total). The dG1 computation is done with 256×256 cells to ascertain the accuracy of the solution since it is our reference; that makes 262144 degrees of freedom for the space approximations (i.e., 6 291 456 degrees of freedom in total).

We show in Figure 3 the scalar flux and the angular flux corresponding to the first angle $\boldsymbol{\Omega}_1$. The angular fluxes for the asymptotic-preserving method are all nonnegative, but the upwind dG1 approximation gives negative values for the scalar flux and the angular fluxes. We observe that the minimum value of the dG1 scalar flux is approximately -0.002 (Figure 3(b)), and the minimum value is -0.183 for the first angular flux (Figure 3(d)). The dG1 approximation is obviously more accurate than the asymptotic-preserving solution, but it experiences overshoots and undershoots at the interfaces between the two materials, whereas the positive asymptotic-preserving solution does not.



(a) $\overline{\psi}_h$, AP scheme (b) $\overline{\psi}_h$, dG1 scheme (c) $\psi_{h,1}$, AP scheme (d) $\psi_{h,1}$, dG1 scheme

FIG. 3. *Scalar $\overline{\psi}_h$ and angular flux $\psi_{h,1}$.*

6. Conclusions. We have introduced a (linear) positive-, asymptotic-preserving method for the approximation of the one-group radiation transport equation (see (4.6)). The approximation in space is discretization agnostic: the approximation can be done with continuous or discontinuous finite elements (or finite volumes). The method is first-order accurate in space. This type of accuracy is coherent with Godunov's theorem since the method is linear. The two key theoretical results of the paper are Theorem 4.4 and Theorem 4.8. We have illustrated the performance of the method with continuous finite elements. We have observed that the method converges with the rate $\mathcal{O}(h)$ in the L^2-norm on manufactured solutions. It converges with the rate $\mathcal{O}(h^2)$ in the L^2-norm in the diffusion limit. The method has also been observed to be nonnegative (in compliance with Theorem 4.8). It does not suffer from overshoots like the upwind dG1 approximation at the interfaces of optically thin and optically thick regions.

The present work is the first part of an ongoing project aimed at developing techniques that are high-order accurate, positivity-preserving, and asymptotic-preserving in the diffusion limit. To reach higher-order accuracy the technique must be made nonlinear. This could be done by using smoothness indicators like in [11, sect. 4.3], or by using limiting technique, or by enforcing positivity through inequality constraints (see, e.g., Hauck and McClarren [13, sect. 4]). Our progresses in this direction will be reported elsewhere.

REFERENCES

[1] M. Adams and E. Larsen, *Fast iterative methods for discrete-ordinates particle transport calculations*, Progr. Nucl. Energy, 40 (2002), pp. 3–159.

[2] M. L. Adams, *Discontinuous finite element transport solutions in thick diffusive problems*, Nucl. Sci. Engrg., 137 (2001), pp. 298–333.

[3] I. Babuška and M. Suri, *On locking and robustness in the finite element method*, SIAM J. Numer. Anal., 29 (1992), pp. 1261–1293.

[4] C. Buet and S. Cordier, *Asymptotic preserving scheme and numerical methods for radiative hydrodynamic models*, C.R. Math. Acad. Sci. Paris, 338 (2004), pp. 951–956.

[5] C. Buet and B. Després, *Asymptotic preserving and positive schemes for radiation hydrodynamics*, J. Comput. Phys., 215 (2006), pp. 717–740.

[6] C. Buet, B. Després, and E. Franck, *Design of asymptotic preserving finite volume schemes for the hyperbolic heat equation on unstructured meshes*, Numer. Math., 122 (2012), pp. 227–278.

[7] S. Chandrasekhar, *Radiative Transfer*, Oxford University Press, Oxford, UK, 1950.

[8] L. Gosse and G. Toscani, *An asymptotic-preserving well-balanced scheme for the hyperbolic heat equations*, C.R. Math. Acad. Sci. Paris, 334 (2002), pp. 337–342.

[9] J.-L. Guermond and G. Kanschat, *Asymptotic analysis of upwind discontinuous Galerkin approximation of the radiative transport equation in the diffusive limit*, SIAM J. Numer. Anal., 48 (2010), pp. 53–78.

[10] J.-L. Guermond and B. Popov, *Invariant domains and first-order continuous finite element approximation for hyperbolic systems*, SIAM J. Numer. Anal., 54 (2016), pp. 2466–2489.

[11] J.-L. Guermond and B. Popov, *Invariant domains and second-order continuous finite element approximation for scalar conservation equations*, SIAM J. Numer. Anal., 55 (2017), pp. 3120–3146.

[12] J.-L. Guermond, B. Popov, and I. Tomas, *Invariant domain preserving discretization-independent schemes and convex limiting for hyperbolic systems*, Comput. Methods Appl. Mech. Engrg., 347 (2019), pp. 143–175.

[13] C. Hauck and R. McClarren, *Positive P_N closures*, SIAM J. Sci. Comput., 32 (2010), pp. 2603–2626.

[14] S. Jin, *Efficient asymptotic-preserving (AP) schemes for some multiscale kinetic equations*, SIAM J. Sci. Comput., 21 (1999), pp. 441–454.

[15] S. Jin and C. D. Levermore, *Numerical schemes for hyperbolic conservation laws with stiff relaxation terms*, J. Comput. Phys., 126 (1996), pp. 449–467.

[16] E. W. Larsen, *On numerical solutions of transport problems in the diffusion limit*, Nucl. Sci. Engrg., 83 (1983), pp. 90–99.

[17] E. W. Larsen and J. E. Morel, *Asymptotic solutions of numerical transport problems in optically thick, diffusive regimes.* II., J. Comput. Phys., 83 (1989), pp. 212–236.

[18] E. W. Larsen, J. E. Morel, and W. F. Miller, Jr., *Asymptotic solutions of numerical transport problems in optically thick, diffusive regimes*, J. Comput. Phys., 69 (1987), pp. 283–324.

[19] P. Lesaint and P.-A. Raviart, *On a finite element method for solving the neutron transport equation*, In Mathematical Aspects of Finite Elements in Partial Differential Equations, C. de Boor, ed., Academic Press, New York, NY, 1974, pp. 89–123.

[20] Q. Li and L. Wang, *Implicit asymptotic preserving method for linear transport equations*, Commun. Comput. Phys., 22 (2017), pp. 157–181.

[21] F. Malvagi and G. C. Pomraning, *Initial and boundary conditions for diffusive linear transport problems*, J. Math. Phys., 32 (1991), pp. 805–820.

[22] C. G. Petra, O. Schenk, and M. Anitescu, *Real-time stochastic optimization of complex energy systems on high-performance computers*, Comput. Sci. Eng., 16 (2014), pp. 32–42.

[23] J. C. Ragusa, J.-L. Guermond, and G. Kanschat, *A robust S_N-DG-approximation for radiation transport in optically thick and diffusive regimes*, J. Comput. Phys., 231 (2012), pp. 1947–1962.

[24] W. Reed and T. Hill, *Triangular Mesh Methods for the Neutron Transport Equation*, Technical Report LA-UR-73-479, Los Alamos Scientific Laboratory, Los Alamos, NM, 1973.

[25] Y. Wang and J. Ragusa, *Diffusion synthetic acceleration for high-order discontinuous finite element S_n transport schemes and application to locally refined unstructured meshes*, Nucl. Sci. Engrg., 166 (2010), pp. 145–166.

[26] J. Xu and L. Zikatanov, *A monotone finite element scheme for convection-diffusion equations*, Math. Comp., 68 (1999), pp. 1429–1446.