IMA Journal of Numerical Analysis (2018) **00**, 1–34
doi:10.1093/imanum/dry025

Numerical analysis of the energy-dependent radiative transfer equation

Kenneth Czuprynski*
Program in Applied Mathematical and Computational Sciences (AMCS), University of Iowa, Iowa City, IA 52242, USA
*Corresponding author: kenneth-czuprynski@uiowa.edu

Joseph Eichholz
Department of Mathematics, Rose-Hulman Institute of Technology, Terre Haute, IN 47803, USA
eichholz@rose-hulman.edu

and

Weimin Han
Department of Mathematics and Program in Applied Mathematical and Computational Sciences (AMCS), University of Iowa, Iowa City, IA 52242-1410, USA
weimin-han@uiowa.edu

[Received on 7 March 2017; revised on 9 October 2017]

The energy-dependent form of the radiative transfer equation (RTE) is important in a variety of applications. In a previous paper the well-posedness and an energy discretization for the energy-dependent RTE were investigated. In this paper the fully discrete scheme is introduced and analysed. Optimal-order error estimates and a well-posedness analysis of the discrete system are provided. The theoretical results are validated through numerical examples.

Keywords: radiative transfer equation; existence; uniqueness; numerical approximation; error estimate.

1. Introduction

The radiative transfer equation (RTE) provides a mathematical description of the transport of radiation through a medium where radiation is scattered, emitted and absorbed. This process arises in many physical disciplines (Chandrasekhar, 1960; Duderstadt & Hamilton, 1976; Duderstadt & Martin, 1979; Lewis & Miller, 1984; Kim & Moscoso, 2006; Azmy & Sartori, 2010; Modest, 2013). For example, the RTE is used to describe the transport of light through the human body in the context of biomedical imaging (Arridge, 1999) as well as charged particle transport in radiation therapy (Uilkema, 2012).

The monoenergetic form is the one predominantly studied in the literature on the RTE. Concerning the numerical analysis of the RTE there is little work specifically devoted to the energy-dependent case or to formulations permitting regions where absorption is negligible.

We begin by formulating the boundary value problem (BVP) and underlying assumptions. Consider a bounded domain $X \subset \mathbb{R}^3$ with boundary $\partial X \in C^1$. Let Ω denote the unit sphere in $\mathbb{R}^3$ and let $E = [e_{\min}, e_{\max}]$ be a closed bounded interval in $\mathbb{R}$ for the energy domain. We further define $U = X \times \Omega \times E$

and $\Gamma = \partial X \times \Omega \times E$. The inflow and outflow portions of Γ are given by

$$\Gamma_- = \{(x, \boldsymbol{\omega}, e) \in \Gamma \mid \boldsymbol{\omega} \cdot \boldsymbol{\nu}(x) < 0\} \text{ and } \Gamma_+ = \{(x, \boldsymbol{\omega}, e) \in \Gamma \mid \boldsymbol{\omega} \cdot \boldsymbol{\nu}(x) > 0\},$$

respectively; here $\boldsymbol{\nu}(x)$ denotes the unit outward normal at the spatial variable $x \in \partial X$. We consider the following BVPs of the RTE (cf. Duderstadt & Hamilton, 1976; Agoshkov, 1998):

$$\boldsymbol{\omega} \cdot \nabla_x u(x, \boldsymbol{\omega}, e) + \sigma_t(x, e)u(x, \boldsymbol{\omega}, e) - \mathcal{S}u(x, \boldsymbol{\omega}, e) = f(x, \boldsymbol{\omega}, e) \text{ in } U, \tag{1.1}$$

$$u(x, \boldsymbol{\omega}, e) = g(x, \boldsymbol{\omega}, e) \text{ on } \Gamma_-, \tag{1.2}$$

where ∇_x denotes the gradient with respect to the spatial variable $x \in X$ and

$$\mathcal{S}u(x, \boldsymbol{\omega}, e) = \int_E \int_\Omega \sigma_s(x, e')\mathcal{P}(x, \boldsymbol{\omega} \cdot \boldsymbol{\omega}', e, e')u(x, \boldsymbol{\omega}', e')\, \mathrm{d}\boldsymbol{\omega}'\, \mathrm{d}e'. \tag{1.3}$$

The function σ_t denotes the total cross section, $\sigma_t = \sigma_a + \sigma_s$, where σ_s and σ_a are the scattering and absorption cross sections, respectively. The functions f and g represent the source term and boundary data, respectively. The function $\mathcal{P}$ reflects the probability that particles scatter from direction $\boldsymbol{\omega}'$ with energy e' into direction $\boldsymbol{\omega}$ with energy e.

We let $I_0 := [-1, 1]$ be the range of the expression $\boldsymbol{\omega} \cdot \boldsymbol{\omega}'$. In the study of the BVPs (1.1) and (1.2) we make the following assumptions:

$$\sigma_t, \sigma_s \in L^\infty(X \times E),\ \sigma_t \geq \sigma_s \geq 0 \text{ and } \sigma_t \geq \sigma_s', \tag{1.4}$$

$$\mathcal{P} \in L^\infty\left(X \times I_0 \times E^2\right) \text{ is non-negative and } \mathcal{P}\left(x, \boldsymbol{\omega} \cdot \boldsymbol{\omega}', e, e'\right) = 0 \text{ for } e' < e, \tag{1.5}$$

$$\mathcal{P} \text{ is normalized such that } \int_E \int_\Omega \mathcal{P}\left(x, \boldsymbol{\omega} \cdot \boldsymbol{\omega}', e, e'\right) \mathrm{d}\boldsymbol{\omega}\, \mathrm{d}e = 1, \tag{1.6}$$

$$f \in L^2(X, C(\Omega \times E)), \quad g \in L^2(\Gamma_-) \text{ is continuous in } \boldsymbol{\omega} \text{ and } e, \tag{1.7}$$

where

$$\sigma_s'(x, e) := \int_E \int_\Omega \sigma_s(x, e')\mathcal{P}(x, \boldsymbol{\omega} \cdot \boldsymbol{\omega}', e, e')\, \mathrm{d}\boldsymbol{\omega}'\, \mathrm{d}e'.$$

Assumption (1.4) implies $\sigma_a = \sigma_t - \sigma_s \geq 0$, which permits the presence of regions with vanishing absorption in the formulation. The condition allowing a vanishing σ_a is not often considered in the literature due to stability issues in *a priori* estimates. We note that the second portion of (1.5) is a no-upscatter condition. It states that a particle cannot gain energy due to a collision. The redistribution function contains the inner product $\boldsymbol{\omega} \cdot \boldsymbol{\omega}'$ as an argument. Throughout this work it will be notationally simpler to consider $\mathcal{P}$ as a function of the arguments $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$ individually. Therefore, we replace $\mathcal{P}(x, \boldsymbol{\omega} \cdot \boldsymbol{\omega}', e, e')$ by $\mathcal{P}(x, \boldsymbol{\omega}, \boldsymbol{\omega}', e, e')$ but continue to enforce assumptions (1.5) and (1.6).

In a previous paper (Czuprynski & Han, 2017) the well-posedness of the BVP and an energy discretization of the energy-dependent RTE were presented. This paper introduces the angular discretization, spatial discretization, fully discrete scheme and numerical examples. At each level of discretization, optimal-order error estimates are derived and the stability of the fully discrete scheme is shown. For the angular discretization, we follow the Finite Element Method (FEM) idea presented in the study by Gao & Zhao (2013). However, unlike Gao & Zhao (2013), in this paper we take an isoparametric-type approach when dealing with the triangulation of the sphere and we prove that the angular convergence order is maintained under relaxed angular regularity assumptions, that is, instead of the $C^2(\Omega)$-type regularity assumptions used in Gao & Zhao (2013), we need only $H^2(\Omega)$-type regularity assumptions. Further, we remark that the proposed discretizations, up to minor modification, are revelant to commonly used techniques in current R&D codes (cf. Seubert, 2012; Baudron *et al.*, 2014). As a result, the following analysis provides theoretical foundations relevant to software used in real world application areas.

This paper is organized as follows. In Section 2 we briefly recall the energy discretization. In Section 3 we introduce the angular discretization and derive error estimates for the resulting semidiscrete scheme. In Section 4 we introduce the spatial discretization, prove the stability of the fully discrete scheme and derive error estimates for the numerical method. In Section 5 we present simulation results, illustrating numerical convergence orders that are in agreement with the theoretical results.

2. Energy discretization

The energy discretization begins by partitioning the energy domain, $E = [e_{\min}, e_{\max}]$, into N energy groups, $e_{\min} \equiv e_1 < e_2 < \cdots < e_N < e_{N+1} \equiv e_{\max}$. Following the convention in the literature we order the energy groups from highest to lowest (cf. Lamarsh, 1966; Duderstadt & Hamilton, 1976; Lewis & Miller, 1984). For $i = 1, \ldots, N$ define $E_i = [e_{N-i+1}, e_{N-i+2}]$, let $|E_i|$ denote the length of the interval and let $h_e = \max_{1\le i\le N} |E_i|$ denote the mesh size.

Denote $D = X \times \Omega$ and define $\partial D_\pm$ similarly to $\Gamma_\pm$. For $i = 1, \ldots, N$ we approximate the solution $u(x, \boldsymbol{\omega}, e)$ for $e \in E_i$ by $u_i(x, \boldsymbol{\omega})$ defined through the BVPs

$$\boldsymbol{\omega} \cdot \nabla_x u_i(x, \boldsymbol{\omega}) + \sigma_{t,\,i}(x) u_i(x, \boldsymbol{\omega}) - \sum_{j=1}^{i} |E_j| S_{i,j} u_j(x, \boldsymbol{\omega}) = f_i(x, \boldsymbol{\omega}) \text{ in } D, \tag{2.8}$$

$$u_i(x, \boldsymbol{\omega}) = g_i(x, \boldsymbol{\omega}) \text{ on } \partial D_-, \tag{2.9}$$

where

$$S_{i,j} u(x, \boldsymbol{\omega}) := \sigma_{s,j}(x) \int_\Omega \mathcal{P}_{i,j}\big(x, \boldsymbol{\omega}, \boldsymbol{\omega}'\big) u\big(x, \boldsymbol{\omega}'\big)\, \mathrm{d}\boldsymbol{\omega}', \quad 1 \le j \le i \le N.$$

The energy-group parameter functions $\sigma_{t,i}, \sigma_{s,i}, \sigma_{a,i}$ and data functions f_i, g_i are defined by the average value of their energy-dependent counterparts over each energy group. The energy group version of $\mathcal{P}$ is similarly defined and is given by

$$\mathcal{P}_{i,j}\big(x, \boldsymbol{\omega}, \boldsymbol{\omega}'\big) := ⨍_{E_i} ⨍_{E_j} \mathcal{P}\big(x, \boldsymbol{\omega}, \boldsymbol{\omega}', e, e'\big)\, \mathrm{d}e'\, \mathrm{d}e \tag{2.10}$$

for $1 \le j \le i \le N$. The symbol $⨍$ is used to denote integration in the average sense.

3. Angular discretization

The angular discretization is motivated by the finite-element-based approach presented in the studies by Gao & Zhao (2009, 2013). However, the treatment of the angular variable in this work differs in two ways: first, an isoparametric-type approach is taken when dealing with the triangulation of the sphere; second, the regularity assumptions in the angular variable are relaxed from $C^2(\Omega)$ to $H^2(\Omega)$ while maintaining the same convergence order.

We briefly define Sobolev spaces over Ω and refer the reader to Hebey (1991) for a more in-depth development. Let α be a multiindex. For $k \geq 0$ an integer and $1 \leq p < \infty$, the Sobolev space

$$W^{k,p}(\Omega) := \left\{ v \in L^p(\Omega) \mid D_S^\alpha v \in L^p(\Omega),\ |\alpha| \leq k \right\}$$

is a Banach space with norm $\|v\|_{W^{k,p}(\Omega)} = (\sum_{|\alpha|\leq k} \|D_S^\alpha v\|_{L^p(\Omega)}^p)^{\frac{1}{p}}$. When $p = 2$ it is a Hilbert space, and we use the canonical notation $H^k(\Omega) \equiv W^{k,2}(\Omega)$. Here D_S^α is the tangential αth-order partial derivative over the sphere.

3.1 *The scheme*

The scheme generates a set of discrete directions, denoted by Ω_{h_a}, over which the RTE is solved. To handle the scattering operator, a finite element representation is used. The set Ω_{h_a} is constructed by projecting a uniformly triangulated triangular plane onto the unit sphere in each octant of $\mathbb{R}^3$. The projection of the vertices contained in the plane gives rise to the set of discrete directions $\Omega_{h_a} = \{\boldsymbol{\omega}_l \in \Omega \mid 1 \leq l \leq L\}$. The set of nodes then partitions Ω into spherical triangles. Let $\mathcal{T}_{h_a}$ denote the partition. It then follows that $\Omega = \bigcup_{\Omega_K \in \mathcal{T}_{h_a}} \Omega_K$, where each Ω_K denotes a spherical element obtained from the projection.

Let $\mathbf{v}_1^K, \mathbf{v}_2^K$ and $\mathbf{v}_3^K$ denote the vertices of Ω_K and let K denote the corresponding planar triangle. In the study by Gao & Zhao (2013) this triangulation has been shown to be quite uniform through numerical experiment. Therefore, we assume that the underlying polyhedral approximation obtained from the set Ω_{h_a} is a shape-regular triangulation. We remark that although more uniform triangulations exist (e.g., Baumgardner & Frederickson, 1985) the analysis is more complicated. The mesh size of the triangulation $\mathcal{T}_{h_a}$ is defined by $h_a = \max_{\Omega_K \in \mathcal{T}_{h_a}} \operatorname{diam}(\Omega_K)$, which is a measure in the sense of geodesic distance. Throughout this work c will be used to denote a constant independent of u as well as the energy, angular and spatial mesh sizes.

In working directly with the curved elements, scattering cross sections and solutions with simple angular distributions are captured well. In this case h_a does not need to be very small to capture the behavior. For more complicated angular distributions multigrid can be used to obtain the needed accuracy, as is done in the studies by Gao & Zhao (2009, 2013), where it is also noted that because coarsening and refining the angular mesh are straightforward, the triangulation is particularly well suited for multigrid.

Consider the system of in-group RTEs (2.8) and (2.9), restricting $\boldsymbol{\omega}$ to Ω_{h_a}, we obtain the following system of hyperbolic equations in the spatial variable $x \in X$:

$$\boldsymbol{\omega}_l \cdot \nabla_x u_{i,l} + \sigma_{t,i} u_{i,l} = \sum_{j=1}^{i} |E_j| \sigma_{s,j} \int_\Omega \mathcal{P}_{i,j}\big(\boldsymbol{\omega}_l, \boldsymbol{\omega}'\big) u_j\big(\boldsymbol{\omega}'\big)\, \mathrm{d}\boldsymbol{\omega}' + f_{i,l} \text{ in } X, \tag{3.11}$$

$$u_{i,l} = g_{i,l} \text{ on } \partial X_-^l, \tag{3.12}$$

for $1 \leq l \leq L$, $1 \leq i \leq N$, where $\partial X_-^l = \{x \in \partial X \mid \boldsymbol{\omega}_l \cdot \nu(x) < 0\}$ is the inflow portion of the boundary of X with respect to the direction $\boldsymbol{\omega}_l$. In the above, $u_{i,l}(x)$ is an approximation of $u_i(x, \boldsymbol{\omega}_l)$, with $f_{i,l}(x) = f_i(x, \boldsymbol{\omega}_l)$ and $g_{i,l}(x) = g_i(x, \boldsymbol{\omega}_l)$ where $\boldsymbol{\omega}_l \in \Omega_{h_a}$.

For each $\Omega_K \in \mathcal{T}_{h_a}$ recall that K denotes the associated planar triangle in $\mathbb{R}^3$. The union over all K results in a polyhedral approximation to the sphere, denoted by $\Omega_\triangle$, where Ω and $\Omega_\triangle$ coincide at the nodes in Ω_{h_a}. Further, Ω_K and K may be related through a radial projection map $P : \Omega_\triangle \to \Omega$, defined by $P(\boldsymbol{\xi}) = \boldsymbol{\xi} / \|\boldsymbol{\xi}\|$, which is a smooth bijection with a piecewise smooth inverse. Figure 1 gives

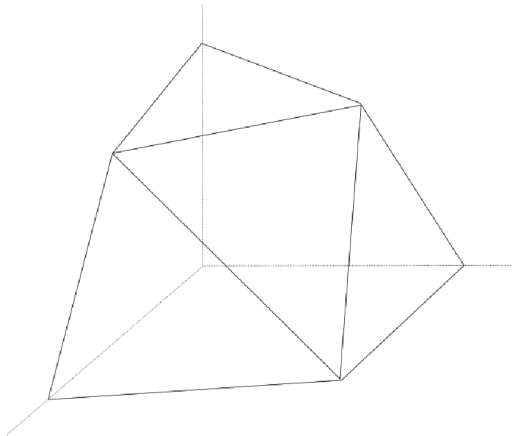

FIG. 1. Example $\Omega_\triangle$ in the first octant.

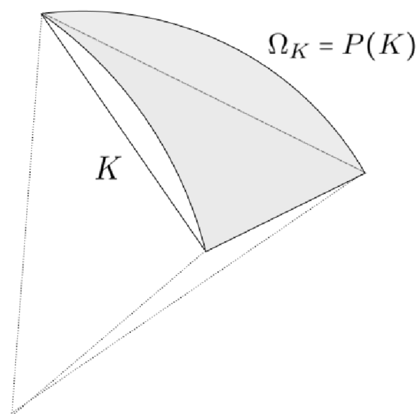

FIG. 2. Relationship between a triangle K on $\Omega_\triangle$ and spherical element Ω_K on Ω.

an example of $\Omega_\triangle$ in the first octant and Fig. 2 illustrates the relation between the planar triangles and the curved elements.

For each planar triangle K, let $\{\hat{\lambda}_m^K : K \to \mathbb{R} \mid m = 1, 2, 3\}$ denote the standard piecewise linear Lagrange shape functions. The corresponding shape functions over the curved elements, $\Omega_K \in \mathcal{T}_{h_a}$, are defined by $\{\lambda_m^K : \Omega_K \to \mathbb{R} \mid \lambda_m^K = \hat{\lambda}_m^K \circ P^{-1},\ m = 1, 2, 3\}$. From the elementwise-defined basis functions let $\{\phi_l : \Omega \to \mathbb{R} \mid 1 \le l \le L\}$ denote their global representation. Then the finite element interpolant of $v \in C(\Omega)$ is given by

$$\tilde{v}(\boldsymbol{\omega}) := \sum_{l=1}^{L} \phi_l(\boldsymbol{\omega}) v(\boldsymbol{\omega}_l). \tag{3.13}$$

This choice is a natural extension of piecewise Lagrange interpolation over planar elements. However, inherently this is an interpolation problem over the sphere with point locations Ω_{h_a} and function values

$\{u_{i,l} \mid l = 1, \ldots, L\}$. The integration over Ω is approximated by replacing the functions in the integrand by their corresponding approximations. This leads to

$$\int_{\Omega} \mathcal{P}_{i,j}\big(x, \boldsymbol{\omega}_l, \boldsymbol{\omega}'\big) u_i\big(x, \boldsymbol{\omega}'\big)\, \mathrm{d}\boldsymbol{\omega}' \approx \int_{\Omega} \tilde{\mathcal{P}}_{i,j}\big(x, \boldsymbol{\omega}_l, \boldsymbol{\omega}'\big) \tilde{u}_i\big(x, \boldsymbol{\omega}'\big)\, \mathrm{d}\boldsymbol{\omega}' = \sum_{k=1}^{L} w_{l,k}^{i,j}(x) u_{i,k}(x) \tag{3.14}$$

for $1 \le j \le i$, $1 \le i \le N$ and $1 \le l \le L$, where

$$w_{l,k}^{i,j}(x) = \int_{\Omega} \phi_k\big(\boldsymbol{\omega}'\big) \tilde{\mathcal{P}}_{i,j}\big(x, \boldsymbol{\omega}_l, \boldsymbol{\omega}'\big)\, \mathrm{d}\boldsymbol{\omega}'. \tag{3.15}$$

Replacing the integration in (3.11) with the above approximation (3.14), the full angular discretization of the problem is obtained; this results in L coupled equations per energy group, for a total of NL coupled hyperbolic problems in space. The system of BVPs obtained from the full discretizaton of the angular variable is given by

$$\boldsymbol{\omega}_l \cdot \nabla_x u_{i,l} + \sigma_{t,i} u_{i,l} = \sum_{j=1}^{i} |E_j| \sigma_{s,j} \sum_{k=1}^{L} w_{l,k}^{i,j} u_{j,k} + f_{i,l} \quad \text{in } X, \tag{3.16}$$

$$u_{i,l} = g_{i,l} \quad \text{on } \partial X_-^l, \tag{3.17}$$

for $1 \le i \le N$ and $1 \le l \le L$. For error analysis we will use the norm

$$\left\| \{v_{i,l}\}_{i=1,l=1}^{N,L} \right\| := \left(\sum_{i=1}^{N} \sum_{l=1}^{L} w_l |E_i| \left\| v_{i,l} \right\|_{L^2(X)}^2 \right)^{1/2}, \quad w_l := \int_{\Omega} \phi_l\big(\boldsymbol{\omega}'\big) \mathrm{d}\boldsymbol{\omega}', \quad 1 \le l \le L. \tag{3.18}$$

The semidiscrete norm is simply the $L^2(X \times \Omega \times E)$ norm of the function

$$v(\mathbf{x}, \boldsymbol{\omega}, e) = \sum_{i=1}^{N} \sum_{l=1}^{L} v_{i,l}(x) \phi_l(\boldsymbol{\omega}) \chi_{E_i}(e)$$

when a mass-lumping approximation of the angular integration is used.

3.2 *Some inequalities*

For the error analysis we follow the approach used in the study by Meir & Tuncer (2009), which applied radially projected finite elements to a second-order partial differential equation over the sphere. We introduce the notation

$$\hat{v}(\boldsymbol{\xi}) = v(P(\boldsymbol{\xi})), \quad \boldsymbol{\xi} \in \Omega_\triangle \tag{3.19}$$

when parameterizing v from any of the planar elements $K \subset \Omega_\triangle$. The main tool in the analysis is the following lemma, which relates the norm over planar elements to the norm over spherical elements in the triangulation. It is a restatement of Meir & Tuncer (2009, Lemma 3.3) with a minor modification and we refer the reader there for the proof. In the following, $\|\cdot\|$ refers to the norm defined in (3.18), and h_e and h_a denote the energy and angular mesh sizes, respectively.

LEMMA 3.1 There exist positive constants c_i, $1 \leq i \leq 5$ such that for each $\Omega_K \in \mathcal{T}_{h_a}$ and associated $K = P^{-1}(\Omega_K)$, we have

$$c_1 \left\|\hat{v}\right\|_{L^2(K)} \leq \|v\|_{L^2(\Omega_K)} \leq c_2 \left\|\hat{v}\right\|_{L^2(K)} \quad \forall\, v \in L^2(\Omega_K),$$
$$c_3 \left\|\hat{v}\right\|_{H^1(K)} \leq \|v\|_{H^1(\Omega_K)} \leq c_4 \left\|\hat{v}\right\|_{H^1(K)} \quad \forall\, v \in H^1(\Omega_K)$$

and

$$|\hat{v}|_{H^2(K)} \leq c_5 \|v\|_{H^2(\Omega_K)} \quad \forall\, v \in H^2(\Omega_K),$$

where $|\cdot|_{H^2(K)}$ denotes the seminorm of $H^2(K)$ and $\hat{v}$ is given by (3.19).

The next set of lemmas provides *a priori* error estimates for the finite element interpolation by employing Lemma 3.1 and the familiar reference-element-type technique. The first lemma is analogous to Meir & Tuncer (2009, Theorem 3.4).

LEMMA 3.2 Assume $v \in H^2(\Omega)$. Then

$$\|v - \tilde{v}\|_{L^2(\Omega)} \leq c h_a^2 \|v\|_{H^2(\Omega)},$$

where $\tilde{v}$ is the finite element interpolant given in (3.13).

Proof. Rewriting in terms of the planar elements and applying Lemma 3.1 yield,

$$\|v - \tilde{v}\|^2_{L^2(\Omega)} = \sum_{\Omega_K} \|v - \tilde{v}\|^2_{L^2(\Omega_K)} \leq c \sum_K \left\|\hat{v}_K - \hat{\tilde{v}}_K\right\|^2_{L^2(K)}.$$

Applying the standard finite element inequalities over the planar element then yields

$$\|v - \tilde{v}\|^2_{L^2(\Omega)} \leq c h_a^4 \sum_K |\hat{v}_K|^2_{H^2(K)}.$$

The proof concludes by using the seminorm inequality in Lemma 3.1. □

By integrating the previous result over X and E the following corollary may be obtained.

COROLLARY 3.3 Assume $v \in L^2(X \times E, H^2(\Omega))$. Then

$$\|v - \tilde{v}\|_{L^2(U)} \leq c h_a^2 \|v\|_{L^2(X \times E, H^2(\Omega))},$$

where $\tilde{v}$ is the finite element interpolant of the angular variable given in (3.13).

We have similar results for the interpolation of the function $\mathcal{P}$. The interpolation is done with respect to the second angular variable. That is,

$$\tilde{\mathcal{P}}\big(x, \boldsymbol{\omega}, \boldsymbol{\omega}', e, e'\big) = \sum_{l=1}^{L} \phi_l\big(\boldsymbol{\omega}'\big)\mathcal{P}\big(\mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\omega}_l, e, e'\big)$$

and $\tilde{\mathcal{P}}_{i,j}(x, \boldsymbol{\omega}, \boldsymbol{\omega}')$ is the analogous angular interpolation of $\mathcal{P}_{i,j}(x, \boldsymbol{\omega}, \boldsymbol{\omega}')$ defined in (2.10).

LEMMA 3.4 Assume $\mathcal{P}(x, \boldsymbol{\omega}, \cdot, e, e') \in H^2(\Omega)$. Then

$$\left\| \mathcal{P}_{i,j} - \tilde{\mathcal{P}}_{i,j} \right\|_{L^2(\Omega)} \leq c h_a^2 \left\| \mathcal{P}_{i,j} \right\|_{H^2(\Omega)},$$

where $\tilde{\mathcal{P}}_{i,j}$ is the finite element interpolant.

COROLLARY 3.5 Let $\boldsymbol{\omega} \in \Omega$. Assume $\mathcal{P}(\cdot, \boldsymbol{\omega}, \cdot, \cdot, \cdot) \in L^2(X \times E^2, H^2(\Omega))$. Then

$$\left\| \mathcal{P} - \tilde{\mathcal{P}} \right\|_{L^2\left(X \times \Omega \times E^2\right)} \leq c h_a^2 \left\| \mathcal{P} \right\|_{L^2\left(X \times E^2, H^2(\Omega)\right)},$$

where $\mathcal{P}$ denotes $\mathcal{P}(\cdot, \boldsymbol{\omega}, \cdot, \cdot, \cdot)$ and $\tilde{\mathcal{P}}$ denotes the finite element interpolant of $\mathcal{P}$ at $\boldsymbol{\omega}$.

The above two lemmas provide useful bounds on the error incurred by replacing the function by its finite element interpolant. The following two results build upon these lemmas. They will provide an estimate for the error due to the approximation in both energy and angle. The next two lemmas will act as intermediary results for obtaining bounds on the error incurred by approximating the scattering operator (1.3) by the approximation in (3.14). It will be useful to discuss the number of weak derivatives a function has in specific variables. Therefore, we introduce the spaces

$$H^{0,2,1}(X \times \Omega \times E) := \left\{ v \in L^2(X \times \Omega \times E) \mid \partial_e v, D_S^\alpha v \in L^2(X \times \Omega \times E),\ |\alpha| \leq 2 \right\},$$
$$H^{0,2,1}\left(X \times \Omega \times E^2\right) := \left\{ v \in L^2\left(X \times \Omega \times E^2\right) \mid \partial_e v, \partial_{e'} v, D_S^\alpha v \in L^2\left(X \times \Omega \times E^2\right),\ |\alpha| \leq 2 \right\}.$$

LEMMA 3.6 Assume $u \in H^{0,2,1}(X \times \Omega \times E)$. Then for $1 \leq j \leq N$,

$$\int_{E_j} \int_{E_j} \left\| u(\cdot, \cdot, e) - \tilde{u}\left(\cdot, \cdot, e'\right) \right\|_{L^2(X \times \Omega)}^2 \,\mathrm{d}e\,\mathrm{d}e'$$
$$\leq c|E_j| \left(h_a^4 \left\| u \right\|_{L^2\left(X \times E_j, H^2(\Omega)\right)}^2 + h_e^2 \left\| u \right\|_{L^2\left(X \times \Omega, H^1\left(E_j\right)\right)}^2 \right).$$

Proof. For $1 \leq j \leq N$ and $e, e' \in E_j$ we write

$$\left| u(x, \boldsymbol{\omega}, e) - \tilde{u}\left(x, \boldsymbol{\omega}, e'\right) \right|^2 \leq 2 \left(\left| u(x, \boldsymbol{\omega}, e) - u\left(x, \boldsymbol{\omega}, e'\right) \right|^2 + \left| u\left(x, \boldsymbol{\omega}, e'\right) - \tilde{u}\left(x, \boldsymbol{\omega}, e'\right) \right|^2 \right).$$

Using

$$u(x, \boldsymbol{\omega}, e) - u\left(x, \boldsymbol{\omega}, e'\right) = \int_{e'}^{e} \frac{\partial u}{\partial e}(x, \boldsymbol{\omega}, \xi)\,\mathrm{d}\xi$$

we can derive the bound

$$\left| u(x, \boldsymbol{\omega}, e) - u\left(x, \boldsymbol{\omega}, e'\right) \right|^2 \leq |E_j| \int_{E_j} \left| \frac{\partial u}{\partial e}(x, \boldsymbol{\omega}, \xi) \right|^2 \mathrm{d}\xi. \tag{3.20}$$

This results in

$$\left|u(x,\boldsymbol{\omega},e)-\tilde{u}\big(x,\boldsymbol{\omega},e'\big)\right|^2 \le c|E_j| \int_{E_j} \left|\frac{\partial u}{\partial e}(x,\boldsymbol{\omega},\xi)\right|^2 \,\mathrm{d}\xi + c\left|u\big(x,\boldsymbol{\omega},e'\big)-\tilde{u}\big(x,\boldsymbol{\omega},e'\big)\right|^2.$$

Integrating with respect to $\boldsymbol{\omega}$ over Ω, x over X and e and e' over E_j we have

$$\begin{aligned}\int_{E_j}\int_{E_j} \left\|u(\cdot,\cdot,e)-\tilde{u}\big(\cdot,\cdot,e'\big)\right\|^2_{L^2(X\times\Omega)} \,\mathrm{d}e\,\mathrm{d}e' \\ \le c\,|E_j| \left(|E_j|^2\,\|u\|^2_{L^2\left(X\times\Omega,H^1\left(E_j\right)\right)} + \|u-\tilde{u}\|^2_{L^2\left(X\times\Omega\times E_j\right)}\right).\end{aligned}$$

Applying Corollary 3.3 and bounding $|E_j|^2$ by h_e^2 yields the claim. □

LEMMA 3.7 Assume that $\mathcal{P}(\cdot,\boldsymbol{\omega},\cdot,\cdot,\cdot) \in H^{0,2,1}(X\times\Omega\times E^2)$ and is continuous in $\boldsymbol{\omega} : \Omega \to H^{0,2,1}(X\times\Omega\times E^2)$. Then

$$\left\|\mathcal{P}-\tilde{\mathcal{P}}_{i,j}\right\|^2_{L^2\left(X\times\Omega\times E_i\times E_j\right)} \le ch_a^4\,\|\mathcal{P}\|^2_{L^2\left(X\times E_i\times E_j,H^2(\Omega)\right)} + ch_e^2\,\|\mathcal{P}\|^2_{L^2\left(X\times\Omega,H^1\left(E_i\times E_j\right)\right)}$$

for $1\le j\le i$, $1\le i\le N$. Here $\mathcal{P}$ denotes $\mathcal{P}(\cdot,\boldsymbol{\omega},\cdot,\cdot)$ and $\tilde{\mathcal{P}}_{i,j}$ is the finite element interpolant at $\boldsymbol{\omega}$.

Proof. Fix $\boldsymbol{\omega}\in\Omega$. Through addition and subtraction we have

$$\left\|\mathcal{P}-\tilde{\mathcal{P}}_{i,j}\right\|_{L^2\left(X\times\Omega\times E_i\times E_j\right)} \le \left\|\mathcal{P}-\mathcal{P}_{i,j}\right\|_{L^2\left(X\times\Omega\times E_i\times E_j\right)} + \left\|\mathcal{P}_{i,j}-\tilde{\mathcal{P}}_{i,j}\right\|_{L^2\left(X\times\Omega\times E_i\times E_j\right)}. \tag{3.21}$$

The first term on the right-hand side can be bounded as follows (Czuprynski & Han, 2017, Lemma 4.3):

$$\left\|\mathcal{P}-\mathcal{P}_{i,j}\right\|_{L^2\left(X\times\Omega\times E_i\times E_j\right)} \le ch_e^2\,\|\mathcal{P}\|_{L^2\left(X\times\Omega,H^1\left(E_i\times E_j\right)\right)}. \tag{3.22}$$

For the second term, using the Cauchy–Schwarz inequality we have

$$\begin{aligned}\left\|\mathcal{P}_{i,j}-\tilde{\mathcal{P}}_{i,j}\right\|_{L^2\left(X\times\Omega\times E_i\times E_j\right)} &\le c\left(|E_i||E_j| \fint_{E_i}\fint_{E_j} \left\|\mathcal{P}\big(e,e'\big)-\tilde{\mathcal{P}}\big(e,e'\big)\right\|^2_{L^2(X\times\Omega)} \,\mathrm{d}e'\,\mathrm{d}e\right)^{1/2} \\ &= c\left\|\mathcal{P}-\tilde{\mathcal{P}}\right\|^2_{L^2\left(X\times\Omega\times E_i\times E_j\right)},\end{aligned} \tag{3.23}$$

where for simplicity, $\mathcal{P}(e,e')$ and $\tilde{\mathcal{P}}(e,e')$ stand for $\mathcal{P}(\cdot,\cdot,e,e')$ and $\tilde{\mathcal{P}}(\cdot,\cdot,e,e')$, respectively. The proof concludes by applying Corollary 3.5 to (3.23) and combining the resulting inequality with inequality (3.22) as an upper bound for (3.21). □

3.3 *Error analysis*

The goal here is to establish an error estimate between the semidiscrete solution $\{u_{i,l}\}_{i=1,l=1}^{N,L}$ of (3.16) and (3.17) and the solution u of the original BVPs (1.1) and (1.2). For each energy group and angular direction we introduce the function $\varepsilon_{i,l} : X \times E_i \to \mathbb{R}$ defined by $\varepsilon_{i,l}(x, e_i) := u(x, \boldsymbol{\omega}_l, e_i) - u_{i,l}(x)$. We have

$$\begin{aligned}\boldsymbol{\omega}_l \cdot \nabla_x \varepsilon_{i,l}(x, e_i) + \sigma_{t,i}(x)\varepsilon_{i,l}(x, e_i) - \sum_{j=1}^{i} |E_j| \sigma_{s,j}(x) \sum_{k=1}^{L} w_{l,k}^{i,j}(x) \varepsilon_{j,k}\big(x, e_j\big) \\ = \Theta_{i,l}\big(x, \{e_j\}_{j=1}^{i}\big) + \Psi_{i,l}(x, e_i),\end{aligned} \tag{3.24}$$

where

$$\Theta_{i,l}\big(x, \{e_j\}_{j=1}^{i}\big) = (\mathcal{S}u)(x, \boldsymbol{\omega}_l, e_i) - \sum_{j=1}^{i} |E_j| \sigma_{s,j}(x) \sum_{k=1}^{L} w_{l,k}^{i,j}(x) u\big(x, \boldsymbol{\omega}_k, e_j\big)$$

and

$$\Psi_{i,l}(x, e_i) = [\sigma_{t,i}(x) - \sigma_t(x, e_i)]u(x, \boldsymbol{\omega}_l, e_i) + [f_{i,l}(x) - f(x, \boldsymbol{\omega}_l, e_i)]. \tag{3.25}$$

In the next three lemmas we establish estimates relating to the right-hand side of equality (3.24). We use $\{\Psi_{i,l}\}_{i=1,l=1}^{N,L}$ to denote the collection of $\Psi_{i,l}$ for $i = 1, \ldots, N$ and $l = 1, \ldots, L$.

LEMMA 3.8 Assume $u \in L^\infty(U)$, $\sigma_s \in L^2(X, H^1(E))$ and $f \in C(\Omega, L^2(X, H^1(E)))$. Then

$$\left\| \{\Psi_{i,l}\}_{i=1,l=1}^{N,L} \right\| \le c h_e \left[\|u\|_{L^\infty(U)} + \left(\sum_{l=1}^{L} w_l |f(\boldsymbol{\omega}_l)|^2_{L^2\left(X, H^1(E)\right)} \right)^{1/2} \right].$$

Proof. Beginning from equation (3.25), applying $\|\cdot\|_{L^2(X)}$ on both sides, we have

$$\left\|\Psi_{i,l}\right\|_{L^2(X)} \le \left\|(\sigma_{t,i} - \sigma_t(e_i)) u(\boldsymbol{\omega}_l, e_i)\right\|_{L^2(X)} + \left\| f_{i,l} - f(\boldsymbol{\omega}_l, e_i)\right\|_{L^2(X)}.$$

Squaring both sides,

$$\left\|\Psi_{i,l}\right\|^2_{L^2(X)} \le 2 \left(\left\|(\sigma_{t,i} - \sigma_t(e_i)) u(\boldsymbol{\omega}_l, e_i)\right\|^2_{L^2(X)} + \left\| f_{i,l} - f(\boldsymbol{\omega}_l, e_i)\right\|^2_{L^2(X)} \right).$$

Integrating over E_i and using the upper bound on u results in

$$\begin{aligned}\int_{E_i} \left\|\Psi_{i,l}\right\|^2_{L^2(X)} \,\mathrm{d}e \le 2 \bigg[\|u\|_{L^\infty(U)} \int_{E_i} \left\|\sigma_{t,i} - \sigma_t(e)\right\|^2_{L^2(X)} \,\mathrm{d}e \\ + \int_{E_i} \left\| f_{i,l} - f(\boldsymbol{\omega}_l, e)\right\|^2_{L^2(X)} \,\mathrm{d}e \bigg].\end{aligned}$$

Using the argument in (3.20) we can obtain

$$\int_{E_i} \left\|\sigma_{t,i} - \sigma_t(e)\right\|^2_{L^2(X)} \, \mathrm{d}e \le c h_e^2 |\sigma_s|^2_{L^2\left(X,H^1(E_i)\right)}.$$

Recall $|f_{i,l}(x) - f(x,\boldsymbol{\omega}_l,e)| = |f_i(x,\boldsymbol{\omega}_l) - f(x,\boldsymbol{\omega}_l,e)|$; then given $f(\cdot,\boldsymbol{\omega}_l,\cdot) \in L^2(X,H^1(E))$ for each $\boldsymbol{\omega}_l \in \Omega_{h_a}$ we have

$$\int_{E_i} \left\|f_{i,l} - f(\boldsymbol{\omega}_l,e)\right\|^2_{L^2(X)} \, \mathrm{d}e \le c\, h_e^2 |f(\boldsymbol{\omega}_l)|^2_{L^2\left(X,H^1(E_i)\right)}.$$

Combining the inequalities results in

$$\int_{E_i} \left\|\Psi_{i,l}\right\|^2_{L^2(X)} \, \mathrm{d}e \le c h_e^2 \left[\|u\|_{L^\infty(U)} \, |\sigma_s|^2_{L^2\left(X,H^1(E_i)\right)} + |f(\boldsymbol{\omega}_l)|^2_{L^2\left(X,H^1(E_i)\right)} \right].$$

The claim follows by building the definition of the norm (3.18). This is done by multiplying by w_l, summing over $1 \le l \le L$ and $1 \le i \le N$, absorbing $|\sigma_s|_{L^2(X,H^1(E))}$ into the constant and applying the square root to both sides. □

For convenience we introduce the quantity

$$\hat{\Theta}^2_{i,l}(x) := ⨍_{E_1} \ldots ⨍_{E_i} \Theta^2_{i,l}\left(x, \{e_j\}_{j=1}^i\right) \mathrm{d}e_i \ldots \mathrm{d}e_1. \tag{3.26}$$

Lemma 3.9 Assume $\sigma_s \in L^2(X,H^1(E))$, $\mathcal{P}(\cdot,\boldsymbol{\omega},\cdot,\cdot,\cdot) \in H^{0,2,1}(X \times \Omega \times E^2)$ and is continuous in $\boldsymbol{\omega} : \Omega \to H^{0,2,1}(X \times \Omega \times E^2)$. Then if $u \in H^{0,2,1}(X \times \Omega \times E)$ and is bounded a.e.,

$$\left\|\{\hat{\Theta}_{i,l}\}_{i=1,l=1}^{N,L}\right\| \le c \left[h_e \left(\|u\|_{L^\infty(U)} + \|u\|_{L^2\left(X\times\Omega,H^1(E)\right)} \right) + h_a^2 \left(\|u\|_{L^\infty(U)} + \|u\|_{L^2\left(X\times E,H^2(\Omega)\right)} \right) \right]$$

where $\hat{\Theta}_{i,l}$ is defined as in (3.26).

Proof. Similar to Czuprynski & Han (2017, Lemma 4.4) we have

$$\sum_{i=1}^N |E_i| \sum_{l=1}^L w_l \left\|\hat{\Theta}_{i,l}\right\|^2_{L^2(X)} \le \mathbb{I}_1 + \mathbb{I}_2 + \mathbb{I}_3, \tag{3.27}$$

where

$$\mathbb{I}_1 := c\, \|u\|^2_{L^\infty(U)} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{L} w_l \int_{E_i} \int_{E_j} \left\| \mathcal{P}\big(\boldsymbol{\omega}_l, e_i, e'\big) - \tilde{\mathcal{P}}_{i,j}(\boldsymbol{\omega}_l) \right\|^2_{L^2(X\times\Omega)} \,\mathrm{d}e'\,\mathrm{d}e_i,$$

$$\mathbb{I}_2 = c\, \|u\|^2_{L^\infty(U)} \sum_{j=1}^{N} \int_{E_j} \left\|\sigma_s\big(e'\big) - \sigma_{s,j}\right\|^2_{L^2\left(X\times\Omega^2\right)} \,\mathrm{d}e',$$

$$\mathbb{I}_3 = c \sum_{j=1}^{N} \fint_{E_j} \int_{E_j} \left\|u\big(e'\big) - \tilde{u}\big(e_j\big)\right\|^2_{L^2\left(X\times\Omega^2\right)} \,\mathrm{d}e'\,\mathrm{d}e_j.$$

In the following we use the assumptions to embed the norms of the functions $\mathcal{P}$ and σ_s into the constant. Starting with $\mathbb{I}_1$, applying Lemma 3.7 results in

$$\begin{aligned}\mathbb{I}_1 &\le c\, \|u\|^2 \sum_{l=1}^{L} w_l \left(h_a^4 \left\|\mathcal{P}(\boldsymbol{\omega}_l)\right\|^2_{L^2\left(X\times E^2, H^2(\Omega)\right)} + h_e^2 \left\|\mathcal{P}(\boldsymbol{\omega}_l)\right\|^2_{L^2\left(X\times\Omega, H^1\left(E^2\right)\right)} \right)\\ &\le c\, \|u\|^2_{L^\infty(U)} \big(h_e^2 + h_a^4\big).\end{aligned} \tag{3.28}$$

The estimate for $\mathbb{I}_2$ was shown in Lemma 3.8: we have

$$\mathbb{I}_2 \le c h_e^2 \, \|u\|^2_{L^\infty(U)}\,. \tag{3.29}$$

Finally, rewriting the right-hand side of $\mathbb{I}_3$ and applying Lemma 3.6 yields

$$\begin{aligned}\mathbb{I}_3 &\le c \sum_{j=1}^{N} \frac{1}{|E_j|} \int_{E_j} \int_{E_j} \left\|u\big(e'\big) - \tilde{u}\big(e_j\big)\right\|^2_{L^2\left(X\times\Omega^2\right)} \,\mathrm{d}e'\,\mathrm{d}e_j\\ &\le c \left(h_e^2 \, \|u\|^2_{L^2\left(X\times\Omega, H^1(E)\right)} + h_a^4 \, \|u\|^2_{L^2\left(X\times E, H^2(\Omega)\right)} \right).\end{aligned} \tag{3.30}$$

Combining the inequalities (3.27)–(3.30) we arrive at the stated result. □

We now integrate equation (3.24) along characteristics. Given any $x \in X$ and $\boldsymbol{\omega}_l \in \Omega_{h_a}$ there exists a point $x_- \in \partial X_-$ such that $x = x_- + t\boldsymbol{\omega}_l$ for some $0 \le t \le \tau(x, \boldsymbol{\omega}_l)$, where $\tau(x, \boldsymbol{\omega}_l) = \sup\{t \mid x_- + t\boldsymbol{\omega}_l \in X\}$. Replacing x in equation (3.24) by $x_- + t\boldsymbol{\omega}_l$ and integrating yields

$$\begin{aligned}\varepsilon_{i,l}(x_- + t\boldsymbol{\omega}_l) = \int_0^t e^{-\int_s^t \sigma_{t,i}(x_- + r\boldsymbol{\omega}_l)\,\mathrm{d}r} &\left(\sum_{j=1}^{i} |E_j| \sigma_{s,j}(x_- + s\boldsymbol{\omega}_l) \sum_{k=1}^{L} w^{i,j}_{l,k}(x_- + s\boldsymbol{\omega}_l) \varepsilon_{j,k}(x_- + s\boldsymbol{\omega}_l) \right.\\ &\left. + \Theta_{i,l}(x_- + s\boldsymbol{\omega}_l) + \Psi_{i,l}(x_- + s\boldsymbol{\omega}_l) \right) \mathrm{d}s.\end{aligned} \tag{3.31}$$

This will be the starting point for the error analysis. For $\{v_l\}_{l=1}^L \in (L^2(X))^L$ define

$$\left\|\{v_l\}_{l=1}^L\right\|_{h_a} := \left(\sum_{l=1}^L w_l \, \|v_l\|^2_{L^2(X)}\right)^{1/2}, \tag{3.32}$$

where w_l is defined as in (3.18). This is simply the semidiscrete norm defined in (3.18) without the energy dependence.

LEMMA 3.10 For h_e sufficiently small the relation

$$\left\|\{\varepsilon_{i,l}\}_{l=1}^L\right\|_{h_a} \leq c\left[\sum_{j=1}^{i-1} |E_j| \left\|\{\varepsilon_{j,k}\}_{l=1}^L\right\|_{h_a} + \left\|\{\Theta_{i,l}\}_{l=1}^L\right\|_{h_a} + \left\|\{\Psi_{i,l}\}_{l=1}^L\right\|_{h_a}\right]$$

holds, for $1 \leq i \leq N$, where $\|\cdot\|_{h_a}$ is given in (3.32).

Proof. From formula (3.31) we obtain

$$\begin{aligned}
&|\varepsilon_{i,l}(x_- + t\boldsymbol{\omega}_l)| \\
&\leq \int_0^t e^{-\int_s^t \sigma_{t,i}(x_- + r\boldsymbol{\omega}_l)\,\mathrm{d}r} \left(\sum_{j=1}^{i} |E_j| \sigma_{s,j}(x_- + s\boldsymbol{\omega}_l) \sum_{k=1}^{L} w_{l,k}^{i,j}(x_- + s\boldsymbol{\omega}_l) |\varepsilon_{j,k}(x_- + s\boldsymbol{\omega}_l)| \right. \\
&\quad \left. + |\Theta_{i,l}(x_- + s\boldsymbol{\omega}_l)| + |\Psi_{i,l}(x_- + s\boldsymbol{\omega}_l)| \right) \mathrm{d}s.
\end{aligned}$$

Using the bound $e^{-\int_s^t \sigma_{t,i}(x_- + r\boldsymbol{\omega}_l)\,\mathrm{d}r} \leq 1$, the bound

$$w_{l,k}^{i,j} \leq \|\mathcal{P}\|_{L^\infty\left(X\times\Omega^2\times E^2\right)} \int_\Omega \phi_k\big(\boldsymbol{\omega}'\big)\,\mathrm{d}\boldsymbol{\omega}' = \|\mathcal{P}\|_{L^\infty\left(X\times\Omega^2\times E^2\right)} \, w_k, \tag{3.33}$$

as well as the upper bound on σ_s we obtain

$$\begin{aligned}
|\varepsilon_{i,l}(x_- + t\boldsymbol{\omega}_l)| \leq \|\sigma_s\mathcal{P}\|_{L^\infty\left(X\times\Omega^2\times E^2\right)} \int_0^t &\left(\sum_{j=1}^{i} |E_j| \sum_{k=1}^{L} w_k |\varepsilon_{j,k}(x_- + s\boldsymbol{\omega}_l)| \right. \\
&\left. + |\Theta_{i,l}(x_- + s\boldsymbol{\omega}_l)| + |\Psi_{i,l}(x_- + s\boldsymbol{\omega}_l)| \right) \mathrm{d}s.
\end{aligned}$$

Extending the integration for $0 \le s \le \tau(\mathbf{x}_-, \boldsymbol{\omega}_l)$ and squaring both sides lead to

$$|\varepsilon_{i,l}(x_- + t\boldsymbol{\omega}_l)|^2 \le c \left[\int_0^\tau \left(\sum_{j=1}^{i} |E_j| \sum_{k=1}^{L} w_k |\varepsilon_{j,k}(x_- + s\boldsymbol{\omega}_l)| \right.\right.$$
$$\left.\left. + |\Theta_{i,l}(x_- + s\boldsymbol{\omega}_l)| + |\Psi_{i,l}(x_- + s\boldsymbol{\omega}_l)| \right) \mathrm{d}s \right]^2,$$

where $\tau = \tau(\mathbf{x}_-, \boldsymbol{\omega}_l)$ and the dependence on σ_s and $\mathcal{P}$ is now contained within the constant. Repeated application of the Cauchy–Schwarz inequality yields

$$|\varepsilon_{i,l}(x_- + t\boldsymbol{\omega}_l)|^2 \le c \left(\sum_{j=1}^{i} |E_j| \sum_{k=1}^{L} w_k \int_0^\tau |\varepsilon_{j,k}(x_- + s\boldsymbol{\omega}_l)|^2 \,\mathrm{d}s \right.$$
$$\left. + \int_0^\tau |\Theta_{i,l}(x_- + s\boldsymbol{\omega}_l)|^2 \,\mathrm{d}s + \int_0^\tau |\Psi_{i,l}(x_- + s\boldsymbol{\omega}_l)|^2 \,\mathrm{d}s \right).$$

Integrating with respect to $t \in [0, \tau]$ and $\mathbf{x}_- \in \partial X_-$ results in

$$\left\| \varepsilon_{i,l} \right\|_{L^2(X)}^2 \le c \left(\sum_{j=1}^{i} |E_j| \sum_{k=1}^{L} w_k \left\| \varepsilon_{j,k} \right\|_{L^2(X)}^2 + \left\| \Theta_{i,l} \right\|_{L^2(X)}^2 + \left\| \Psi_{i,l} \right\|_{L^2(X)}^2 \right),$$

where we have used the identity

$$\int_{\partial X_-} \int_0^{\tau(x_-, \boldsymbol{\omega})} v(x_- + t\boldsymbol{\omega}) \,\mathrm{d}t \,\mathrm{d}S(x_-) = \int_X v(x) \,\mathrm{d}x.$$

Multiplying by w_l, summing over $l = 1, \ldots, L$, renaming the indices on the right-hand side and separating the summation over the energy groups yields

$$\sum_{l=1}^{L} w_l \left\| \varepsilon_{i,l} \right\|_{L^2(X)}^2 \le c\, |E_i| \sum_{l=1}^{L} w_l \left\| \varepsilon_{i,l} \right\|_{L^2(X)}^2 + c \left(\sum_{j=1}^{i-1} |E_j| \sum_{l=1}^{L} w_l \left\| \varepsilon_{j,l} \right\|_{L^2(X)}^2 \right.$$
$$\left. + \sum_{l=1}^{L} w_l \left\| \Theta_{i,l} \right\|_{L^2(X)}^2 + \sum_{l=1}^{L} w_l \left\| \Psi_{i,l} \right\|_{L^2(X)}^2 \right).$$

For $|E_i|$ sufficiently small, i.e., $|E_i| < 1/c$, we have

$$\sum_{l=1}^{L} w_l \left\| \varepsilon_{i,l} \right\|_{L^2(X)}^2 \le c \left(\sum_{j=1}^{i-1} |E_j| \sum_{l=1}^{L} w_l \left\| \varepsilon_{j,l} \right\|_{L^2(X)}^2 + \sum_{l=1}^{L} w_l \left\| \Theta_{i,l} \right\|_{L^2(X)}^2 + \sum_{l=1}^{L} w_l \left\| \Psi_{i,l} \right\|_{L^2(X)}^2 \right).$$

The claim then follows from the definition of the norm $\|\cdot\|_{h_a}$ given in (3.32). $\square$

REMARK 3.11 The importance of the energy discretization introduced in Section 2 is apparent in the proof of Lemma 3.10. We are able to establish the inequality because the energy-group mesh size weights the contribution of the scattering term in the ith energy group. As a result, for sufficiently small h_e, a recurrence involving the error can be established.

The next result establishes an upper bound for the error between the semidiscrete solution and the solution of the original BVP.

THEOREM 3.12 Assume $f \in C(\Omega, L^2(X, H^1(E)))$ and that the conditions of Lemma 3.9 hold. Then for h_e sufficiently small the inequality

$$\left\|\{\varepsilon_{i,l}\}_{i=1,l=1}^{N,L}\right\| \le c\left[h_e\left(\|u\|_{L^\infty(U)} + \|u\|_{L^2\left(X\times\Omega, H^1(E)\right)} + \left(\sum_{l=1}^{L} w_l |f(\boldsymbol{\omega}_l)|^2_{L^2\left(X, H^1(E)\right)}\right)^{1/2}\right)\right.$$
$$\left. + h_a^2\left(\|u\|_{L^\infty(U)} + \|u\|_{L^2\left(X\times E, H^2(\Omega)\right)}\right)\right]$$

holds.

Proof. Apply Lemma 3.10 and the Cauchy–Schwarz inequality to obtain

$$\left\|\{\varepsilon_{i,l}\}_{l=1}^{L}\right\|_{h_a} \le c\left[\left(\sum_{j=1}^{i-1}|E_j|\left\|\{\varepsilon_{j,k}\}_{l=1}^{L}\right\|_{h_a}^2\right)^{1/2} + \left\|\{\Theta_{i,l}\}_{l=1}^{L}\right\|_{h_a} + \left\|\{\Psi_{i,l}\}_{l=1}^{L}\right\|_{h_a}\right].$$

Recalling the dependence on $\{e_j\}_{j=1}^{i}$, squaring both sides and using the Cauchy–Schwarz inequality leads to

$$\left\|\{\varepsilon_{i,l}(e_i)\}_{l=1}^{L}\right\|_{h_a}^2 \le c\left[\sum_{j=1}^{i-1}|E_j|\left\|\{\varepsilon_{j,l}(e_j)\}_{l=1}^{L}\right\|_{h_a}^2 + \left\|\left\{\Theta_{i,l}\left(\{e_k\}_{k=1}^{i}\right)\right\}_{l=1}^{L}\right\|_{h_a}^2 + \left\|\{\Psi_{i,l}(e_i)\}_{l=1}^{L}\right\|_{h_a}^2\right].$$

For $1 \le j \le i$ we integrate with respect to e_j over E_j in the average sense. This yields

$$\fint_{E_i}\left\|\{\varepsilon_{i,l}(e_i)\}_{l=1}^{L}\right\|_{h_a}^2 \mathrm{d}e_i \le c\left[\sum_{j=1}^{i-1}|E_j|\fint_{E_j}\left\|\{\varepsilon_{j,k}(e_j)\}_{l=1}^{L}\right\|_{h_a}^2 \mathrm{d}e_j\right.$$
$$\left. + \left\|\{\hat{\Theta}_{i,l}\}_{l=1}^{L}\right\|_{h_a}^2 + \fint_{E_i}\left\|\{\Psi_{i,l}(e_i)\}_{l=1}^{L}\right\|_{h_a}^2 \mathrm{d}e_i\right],$$

where we have used the definition of $\hat{\Theta}_{i,l}$ given by equation (3.26). Apply the discrete Gronwall inequality (see, e.g., Clark, 1987),

$$\fint_{E_i}\left\|\{\varepsilon_{i,l}(e_i)\}_{l=1}^{L}\right\|_{h_a}^2 \mathrm{d}e \le c\left(\gamma_i + \sum_{j=1}^{i-1}|E_j|\gamma_j\right),$$

where

$$\gamma_k = \left\| \{\hat{\Theta}_{k,l}\}_{l=1}^{L} \right\|_{h_a}^2 + \fint_{E_k} \left\| \{\Psi_{k,l}(e_k)\}_{l=1}^{L} \right\|_{h_a}^2 \, \mathrm{d}e_k,$$

for $k = 1, \ldots, i$. Extending the summation from $i-1$ to N, multiplying both sides by $|E_i|$ and summing over $1 \le i \le N$ results in

$$\sum_{i=1}^{N} \int_{E_i} \left\| \{\varepsilon_{i,l}(e_i)\}_{l=1}^{L} \right\|_{h_a}^2 \, \mathrm{d}e_i \le c \sum_{i=1}^{N} \left(|E_i| \left\| \{\hat{\Theta}_{i,l}\}_{l=1}^{L} \right\|_{h_a}^2 + \int_{E_i} \left\| \{\Psi_{i,l}(e_i)\}_{l=1}^{L} \right\|_{h_a}^2 \, \mathrm{d}e_i \right).$$

The claim follows from the definition of the norm $\|\cdot\|$ and Lemmas 3.8 and 3.9. □

4. Spatial discretization

Finally, the spatial variable in (3.16) and (3.17) is dealt with using a discontinuous Galerkin (DG) method. DG methods have recently been used in the discretization of the spatial variable of the monoenergetic form of the RTE under the assumption of nonvanishing absorption (Han *et al.*, 2010; Gao & Zhao, 2013). A more general treatment of DG methods as applied to first-order hyperbolic BVPs is found in the study by Brezzi *et al.* (2004). Comprehensive coverage of DG methods can be found in a number of books, e.g., Di Pietro & Ern (2012). Our DG method is derived through a standard approach; however, for the convenience of the reader we will sketch the main steps of the derivation.

4.1 *DG method*

For ease of presentation assume X is a polyhedron. Let $\{\mathcal{T}_h\}_{h>0}$ be a regular family of triangulations of the spatial domain X and from this family consider the triangulation $\mathcal{T}_h$. For each $K \in \mathcal{T}_h$ let h_K denote the diameter of K and let $h := \max_{K\in\mathcal{T}_h} h_K$. Given $K \in \mathcal{T}_h$, ∂K denotes the faces of element K and with each face we associate an outward-facing unit normal vector $\boldsymbol{\nu}_K$. On a face on the boundary ∂X, $\boldsymbol{\nu}_K$ is the unit outward normal vector. Let $\mathcal{E}_h$ denote the set of interior faces of the triangulation. For each face $\rho \in \mathcal{E}_h$ let K^+ and K^- denote adjacent elements sharing ρ; we then define $\boldsymbol{\nu}_\rho$ to be the normal vector pointing from K^- toward K^+. When dealing with a function v over an edge ρ let $v^+ := v|_{K^+}$ and $v^- := v|_{K^-}$ denote the restriction of v over the corresponding adjacent elements. We denote a jump by $[\![v]\!] := v^+ - v^-$ and define the space

$$V_h = \left\{ v \in L^2(X) \mid v|_K \in P_r(K) \; \forall K \in \mathcal{T}_h \right\}, \tag{4.34}$$

where $r \in \mathbb{N}$ and $P_r(K)$ denotes polynomials on K of a degree less than or equal to r.

For $v \in V_h$ multiply equation (3.16) by v and integrate over X to obtain

$$\sum_{K\in\mathcal{T}_h} \int_K \boldsymbol{\omega}_l \cdot \nabla_x u_{i,l} v_K \, \mathrm{d}x + \sum_{K\in\mathcal{T}_h} \int_K \sigma_{t,i} u_{i,l} v_K \, \mathrm{d}x$$
$$= \sum_{K\in\mathcal{T}_h} \int_K v_K \sum_{j=1}^{i} |E_j| \sigma_{s,j} \sum_{k=1}^{L} w_{l,k}^{i,j} u_{j,k} \, \mathrm{d}x + \sum_{K\in\mathcal{T}_h} \int_K f_{i,l} v_K \, \mathrm{d}x, \tag{4.35}$$

where $v_K = v|_K \in P_r(K)$. We perform integration by parts on the first term of (4.35) to yield

$$\sum_{K\in\mathcal{T}_h}\int_{\partial K}\boldsymbol{\omega}_l\cdot\boldsymbol{\nu}_K\hat{u}_{i,l}v_K\,\mathrm{d}x-\sum_{K\in\mathcal{T}_h}\int_K u_{i,l}(\boldsymbol{\omega}_l\cdot\nabla_x v_K)\,\mathrm{d}x+\sum_{K\in\mathcal{T}_h}\int_K\sigma_{t,i}u_{i,l}v_K\,\mathrm{d}x$$
$$-\sum_{K\in\mathcal{T}_h}\int_K v_K\sum_{j=1}^{i}|E_j|\sigma_{s,j}\sum_{k=1}^{L}w_{l,k}^{i,j}u_{j,k}\,\mathrm{d}x=\sum_{K\in\mathcal{T}_h}\int_K f_{i,l}v_K\,\mathrm{d}x,$$

where $\hat{u}_{i,l}$ is the numerical trace

$$\hat{u}_{i,l}(x)=\begin{cases}g_{i,l}(x) & \text{if } x\in\partial X_-^l,\\ \lim_{\varepsilon\to 0^+}u_{i,l}(x-\varepsilon\boldsymbol{\omega}_l) & \text{otherwise,}\end{cases}$$

∂X_-^l and ∂X_+^l being the inflow and outflow portions, respectively, of ∂X with respect to the direction $\boldsymbol{\omega}_l$, $1\le l\le L$. Applying the definition of the trace along the inflow boundary and breaking up the summation over the energy groups leads to

$$\sum_{\rho\in\mathcal{E}_h}\int_\rho\boldsymbol{\omega}_l\cdot\boldsymbol{\nu}_\rho\hat{u}_{i,l}[\![v_l]\!]\,\mathrm{d}x-\sum_{K\in\mathcal{T}_h}\int_K u_{i,l}(\boldsymbol{\omega}_l\cdot\nabla_x v_K)\,\mathrm{d}x+\int_{\partial X_+^l}\boldsymbol{\omega}_l\cdot\boldsymbol{\nu}\hat{u}_{i,l}v_K\,\mathrm{d}s$$
$$+\sum_{K\in\mathcal{T}_h}\int_K\sigma_{t,i}u_{i,l}v_K\,\mathrm{d}x-\sum_{K\in\mathcal{T}_h}\int_K v_K|E_i|\sigma_{s,i}\sum_{k=1}^{L}w_{l,k}^{i,i}u_{i,k}\,\mathrm{d}x$$
$$=\sum_{K\in\mathcal{T}_h}\int_K v_K\sum_{j=1}^{i-1}|E_j|\sigma_{s,j}\sum_{k=1}^{L}w_{l,k}^{i,j}u_{j,k}\,\mathrm{d}x+\sum_{K\in\mathcal{T}_h}\int_K f_{i,l}v_K\,\mathrm{d}x-\int_{\partial X_-^l}\boldsymbol{\omega}_l\cdot\boldsymbol{\nu}g_{i,l}\,\mathrm{d}s.$$

Denote an element $\{v_l\}_{l=1}^L\in(V_h)^L$ by $\mathbf{v}=\{v_l\}_{l=1}^L$. For $\mathbf{u},\mathbf{v}\in(V_h)^L$ define the following three bilinear forms:

$$\mathbf{a}(\mathbf{u},\mathbf{v})=\sum_{l=1}^{L}w_l\left[\sum_{\rho\in\mathcal{E}_h}\int_\rho\boldsymbol{\omega}_l\cdot\boldsymbol{\nu}_\rho\hat{u}_l[\![v_l]\!]\,\mathrm{d}x-\sum_{K\in\mathcal{T}_h}\int_K u_l(\boldsymbol{\omega}_l\cdot\nabla_x v_{l,K})\,\mathrm{d}x+\int_{\partial X_+^l}\boldsymbol{\omega}_l\cdot\boldsymbol{\nu}u_lv_l\,\mathrm{d}s\right],$$

$$\mathbf{b}_i(\mathbf{u},\mathbf{v})=\sum_{l=1}^{L}w_l\left[\sum_{K\in\mathcal{T}_h}\int_K\sigma_{t,i}u_lv_{l,K}\,\mathrm{d}x-\sum_{K\in\mathcal{T}_h}\int_K v_{l,K}|E_i|\sigma_{s,i}\sum_{k=1}^{L}w_{l,k}^{i,i}u_k\,\mathrm{d}x\right],$$

$$\mathbf{c}_{i,j}(\mathbf{u},\mathbf{v})=\sum_{l=1}^{L}w_l\sum_{K\in\mathcal{T}_h}\int_K v_{l,K}|E_j|\sigma_{s,j}\sum_{k=1}^{L}w_{l,k}^{i,j}u_k\,\mathrm{d}x,$$

for $1\le j\le i$ with $1\le i\le N$, where $v_{l,K}=v_l|_K$. For $1\le i\le N$ define the linear functional

$$\mathbf{l}_i(\mathbf{v}) = \sum_{l=1}^{L} w_l \left[\sum_{K \in \mathcal{T}_h} \int_K f_{i,l} v_{l,K} \,\mathrm{d}x - \int_{\partial X_-^l} \boldsymbol{\omega}_l \cdot \boldsymbol{\nu} g_{i,l} \,\mathrm{d}s \right].$$

We then obtain a coupled system of equations and the following problem:

$$\begin{cases} \text{find } \mathbf{u}_i^h \in (V_h)^L \text{ s.t.} \\ \mathbf{a}\big(\mathbf{u}_i^h, \mathbf{v}\big) + \mathbf{b}_i\big(\mathbf{u}_i^h, \mathbf{v}\big) = \sum_{j=1}^{i-1} \mathbf{c}_{i,j}\big(\mathbf{u}_j^h, \mathbf{v}\big) + \mathbf{l}_i(\mathbf{v}) \quad \forall\, \mathbf{v} \in (V_h)^L, \end{cases} \tag{4.36}$$

for $1 \le i \le N$. Here $\mathbf{u}_i^h = \{u_{i,l}^h\}_{l=1}^L \in (V_h)^L$ denotes the solution of the ith fully discrete energy-group equation in (4.36) for $1 \le i \le N$.

REMARK 4.1 Due to the no upscatter assumption (1.5) the fully discrete scheme can be seen as solving a sequence of DG schemes. The bilinear form $c_{i,j}(\cdot,\cdot)$ is then used to couple solutions over each energy group. As a result, the stability of the method can be approached by considering the method over each individual energy group.

4.2 *Stability*

The stability and error estimates of the method are derived using the norm

$$|||\mathbf{v}|||_h := \left[\sum_{l=1}^{L} w_l \left(\|v_l\|_{L^2(X)}^2 + \sum_{\rho \in \mathcal{E}_h} \left\| |\boldsymbol{\omega}_l \cdot \boldsymbol{\nu}_\rho|^{1/2} [\![v_l]\!] \right\|_{L^2(\rho)}^2 + \int_{\partial X_+^l} \boldsymbol{\omega}_l \cdot \boldsymbol{\nu} v_l^2 \,\mathrm{d}s \right) \right]^{1/2}, \tag{4.37}$$

defined on $(H^1(X) + V_h)^L$. We first establish the coercivity of the left-hand side of (4.36).

LEMMA 4.2 Assume there exists a constant $m_{\sigma_t} > 0$ such that $\sigma_t \ge m_{\sigma_t} > 0$. Then for h_e sufficiently small,

$$\mathbf{a}(\mathbf{v},\mathbf{v}) + \mathbf{b}_i(\mathbf{v},\mathbf{v}) \ge c|||\mathbf{v}|||_h^2 \quad \forall\, \mathbf{v} \in (V_h)^L,$$

for $1 \le i \le N$.

Proof. For $\mathbf{v} \in (V_h)^L$,

$$\mathbf{a}(\mathbf{v},\mathbf{v}) = \sum_{l=1}^{L} w_l \left[\sum_{\rho \in \mathcal{E}_h} \int_\rho \boldsymbol{\omega}_l \cdot \boldsymbol{\nu}_\rho \hat{v}_l [\![v_l]\!] \,\mathrm{d}x - \sum_{K \in \mathcal{T}_h} \int_K v_{l,K} (\boldsymbol{\omega}_l \cdot \nabla_x v_{l,K}) \,\mathrm{d}x + \int_{\partial X_+^l} \boldsymbol{\omega}_l \cdot \boldsymbol{\nu} (v_l)^2 \,\mathrm{d}s \right]. \tag{4.38}$$

For the second term in (4.38), application of the divergence theorem and splitting the summation over the interior and physical boundaries leads to

$$\begin{aligned}\sum_{K\in\mathcal{T}_h}\int_K v_{l,K}(\boldsymbol{\omega}_l\cdot\nabla_x v_{l,K})\,\mathrm{d}x &= \frac{1}{2}\sum_{K\in\mathcal{T}_h}\int_K \boldsymbol{\omega}_l\cdot\nabla_x\big(v_{l,K}^2\big)\,\mathrm{d}x\\ &= \frac{1}{2}\sum_{K\in\mathcal{T}_h}\int_{\partial K}(\boldsymbol{\omega}_l\cdot\boldsymbol{\nu}_K)v_{l,K}^2\,\mathrm{d}x\\ &= \frac{1}{2}\sum_{\rho\in\mathcal{E}_h}\int_\rho \boldsymbol{\omega}_l\cdot\boldsymbol{\nu}_\rho \left[\!\left[v_l^2\right]\!\right]\,\mathrm{d}x + \frac{1}{2}\int_{\partial X^l}(\boldsymbol{\omega}_l\cdot\boldsymbol{\nu})v_l^2\,\mathrm{d}x.\end{aligned}$$

Over the inflow boundary, $\boldsymbol{\omega}_l\cdot\boldsymbol{\nu}_K < 0$; therefore, omitting the term results in

$$\sum_{K\in T_h}\int_K v_{l,K}(\boldsymbol{\omega}_l\cdot\nabla_x v_{l,K})\,\mathrm{d}x \le \frac{1}{2}\sum_{\rho\in\mathcal{E}_h}\int_\rho \boldsymbol{\omega}_l\cdot\boldsymbol{\nu}_\rho \left[\!\left[v_l^2\right]\!\right]\,\mathrm{d}x + \frac{1}{2}\int_{\partial X_+^l}(\boldsymbol{\omega}_l\cdot\boldsymbol{\nu})v_l^2\,\mathrm{d}x.$$

Returning to (4.38) we now have

$$\begin{aligned}\mathbf{a}(\mathbf{v},\mathbf{v}) &\ge \sum_{l=1}^{L} w_l\left[\sum_{\rho\in\mathcal{E}_h}\int_\rho \boldsymbol{\omega}_l\cdot\boldsymbol{\nu}_\rho\hat{v}_l[\![v_l]\!]\,\mathrm{d}x - \frac{1}{2}\sum_{\rho\in\mathcal{E}_h}\int_\rho \boldsymbol{\omega}_l\cdot\boldsymbol{\nu}_\rho\left[\!\left[v_l^2\right]\!\right]\,\mathrm{d}x + \frac{1}{2}\int_{\partial X_+^l}(\boldsymbol{\omega}_l\cdot\boldsymbol{\nu})v_l^2\,\mathrm{d}s\right]\\ &= \sum_{l=1}^{L} w_l\left[\sum_{\rho\in\mathcal{E}_h}\int_\rho \boldsymbol{\omega}_l\cdot\boldsymbol{\nu}_\rho\left(\hat{v}_l[\![v_l]\!] - \frac{1}{2}\left[\!\left[v_l^2\right]\!\right]\right)\mathrm{d}x + \frac{1}{2}\int_{\partial X_+^l}(\boldsymbol{\omega}_l\cdot\boldsymbol{\nu})v_l^2\,\mathrm{d}s\right]. \qquad (4.39)\end{aligned}$$

When $\boldsymbol{\omega}_l\cdot\boldsymbol{\nu}_\rho \ge 0$ the integration over the interior edge becomes

$$\begin{aligned}\int_\rho |\boldsymbol{\omega}_l\cdot\boldsymbol{\nu}_\rho|\left(\hat{v}_l[\![v_l]\!] - \frac{1}{2}\left[\!\left[v_l^2\right]\!\right]\right)\mathrm{d}x &= \int_\rho |\boldsymbol{\omega}_l\cdot\boldsymbol{\nu}_\rho|\left(v_l^+[\![v_l]\!] - \frac{1}{2}\left[\!\left[v_l^2\right]\!\right]\right)\mathrm{d}x\\ &= \frac{1}{2}\int_\rho |\boldsymbol{\omega}_l\cdot\boldsymbol{\nu}_\rho|[\![v_l]\!]^2\,\mathrm{d}x.\end{aligned}$$

When $\boldsymbol{\omega}_l\cdot\boldsymbol{\nu}_\rho < 0$,

$$\begin{aligned}\int_\rho \boldsymbol{\omega}_l\cdot\boldsymbol{\nu}_\rho\left(\hat{v}_l[\![v_l]\!] - \frac{1}{2}\left[\!\left[v_l^2\right]\!\right]\right)\mathrm{d}x &= -\int_\rho |\boldsymbol{\omega}_l\cdot\boldsymbol{\nu}_\rho|\left(v_l^-[\![v_l]\!] - \frac{1}{2}\left[\!\left[v_l^2\right]\!\right]\right)\mathrm{d}x\\ &= \frac{1}{2}\int_\rho |\boldsymbol{\omega}_l\cdot\boldsymbol{\nu}_\rho|[\![v_l]\!]^2\,\mathrm{d}x.\end{aligned}$$

Therefore,

$$\sum_{\rho\in\mathcal{E}_h}\int_{\rho}\boldsymbol{\omega}_l\cdot\boldsymbol{\nu}_\rho\left(\hat{v}_l[\![v_l]\!]-\frac{1}{2}\left[\!\!\left[v_l^2\right]\!\!\right]\right)\mathrm{d}x=\frac{1}{2}\sum_{\rho\in\mathcal{E}_h}\int_{\rho}|\boldsymbol{\omega}_l\cdot\boldsymbol{\nu}_\rho|[\![v_l]\!]^2\,\mathrm{d}x,$$

which, upon substitution into inequality (4.39), yields

$$\mathbf{a}^h(\mathbf{v},\mathbf{v})\geq\frac{1}{2}\sum_{l=1}^{L}\boldsymbol{\omega}_l\left[\sum_{\rho\in\mathcal{E}_h}\int_{\rho}|\boldsymbol{\omega}_l\cdot\boldsymbol{\nu}_\rho|[\![v_l]\!]^2\,\mathrm{d}x+\int_{\partial X_+^l}|\boldsymbol{\omega}_l\cdot\boldsymbol{\nu}|v_l^2\,\mathrm{d}s\right].\tag{4.40}$$

We next consider the bilinear form

$$\mathbf{b}_i(\mathbf{v},\mathbf{v})=\sum_{l=1}^{L}w_l\left[\sum_{K\in\mathcal{T}_h}\int_K\sigma_{t,i}v_{l,K}^2\,\mathrm{d}x-\sum_{K\in\mathcal{T}_h}\int_K v_{l,K}|E_i|\sigma_{s,i}\sum_{k=1}^{L}w_{l,k}^{i,i}v_{k,K}\,\mathrm{d}x\right]\tag{4.41}$$

for $1\leq i\leq N$. It is straightforward to show

$$\int_K|E_i|\sigma_{s,i}\sum_{k=1}^{L}w_{l,k}^{i,i}v_{k,K}v_{l,K}\,\mathrm{d}x\leq\frac{|E_i|}{2}\int_K\sigma_{s,i}\sum_{k=1}^{L}w_{l,k}^{i,i}\left(v_{l,K}^2+v_{k,K}^2\right)\mathrm{d}x.$$

Then use the bound on the weights $w_{l,k}^{i,i}$ given in (3.33) and substitute back into (4.41):

$$\mathbf{b}_i(\mathbf{v},\mathbf{v})\geq\sum_{l=1}^{L}w_l\left[\sum_{K\in\mathcal{T}_h}\int_K\sigma_{t,i}v_{l,K}^2\,\mathrm{d}x-\sum_{K\in\mathcal{T}_h}\frac{|E_i|}{2}c\int_K\sigma_{s,i}\sum_{k=1}^{L}w_k\left(v_{l,K}^2+v_{k,K}^2\right)\mathrm{d}x\right].$$

After renaming indices and simplifying, the following bound is obtained:

$$\mathbf{b}_i(\mathbf{v},\mathbf{v})\geq\sum_{l=1}^{L}w_l\sum_{K\in\mathcal{T}_h}\int_K\left[\sigma_{t,i}-|E_i|c\sigma_{s,i}\right]v_{l,K}^2\,\mathrm{d}x.$$

Define $\mu=\min_{1\leq i\leq N}\{\sigma_{t,i}-|E_i|c\sigma_{s,i}\}$. Given $\sigma_t\geq m_{\sigma_t}>0$ it follows that for h_e sufficiently small, $\mu>0$ and therefore

$$\mathbf{b}_i(\mathbf{v},\mathbf{v})\geq\mu\sum_{l=1}^{L}w_l\sum_{K\in\mathcal{T}_h}\int_K v_{l,K}^2\,\mathrm{d}x.\tag{4.42}$$

Combining inequalities (4.40) and (4.42) yields the claim. □

Then by applying the Lax–Milgram lemma recursively we have the following existence and uniqueness result.

COROLLARY 4.3 Under the assumptions of Lemma 4.2, system (4.36) has a unique solution $\mathbf{u}_h\in(V_h)^{NL}$.

4.3 *Error analysis*

Next, a bound between $\mathbf{u}_h = \{u^h_{i,l}\}_{i=1,l=1}^{N,L}$ and the solution, $\{u_{i,l}\}_{i=1,l=1}^{N,L}$, of the angular semidiscretizations (3.16) and (3.17) will be established. The norm to be used is

$$\left\| \{u_{i,l}\}_{i=1,l=1}^{N,L} - \mathbf{u}_h \right\|_h := \left[\sum_{i=1}^{N} |E_i| \left|\left|\left| \left\{ u_{i,l} - u^h_{i,l} \right\}_{l=1}^{L} \right|\right|\right|_h^2 \right]^{1/2}. \tag{4.43}$$

Let $\pi_h : L^2(X) \to V_h$ denote the L^2-projector onto V_h. By definition, for any $u \in L^2(X)$,

$$\int_X (u - \pi_h u) v \, \mathrm{d}x = 0 \tag{4.44}$$

for all $v \in V_h$. Let $u \in H^{r+1}(X)$. Then the error between u and its projection onto V_h, over any element $K \in \mathcal{T}_h$, has the bound

$$\|u - \pi_h u\|_{H^m(K)} \le c h^{r+1-m} \|u\|_{H^{r+1}(K)}, \tag{4.45}$$

where r is the polynomial degree used in definition (4.34) of V_h. The following trace inequality holds: (cf. Brezzi *et al.*, 2004):

$$\|u - \pi_h u\|_{L^2(\rho)} \le c h^{r+\frac{1}{2}} \|u\|_{H^{r+1}(K)} \tag{4.46}$$

for all $\rho \in \mathcal{E}_h$. For $\{u_{i,l}\}_{i=1,l=1}^{N,L}$ satisfying the angular semidiscretizations (3.16) and (3.17), and $\{u^h_{i,l}\}_{i=1,l=1}^{N,L}$ satisfying the fully discrete problem (4.36), define

$$\boldsymbol{\delta}_i = \{\delta_{i,l}\}_{l=1}^{L} = \{u_{i,l} - \pi_h u_{i,l}\}_{l=1}^{L} \quad \text{and} \quad \boldsymbol{\eta}_i = \{\eta_{i,l}\}_{l=1}^{L} = \left\{ u^h_{i,l} - \pi_h u_{i,l} \right\}_{l=1}^{L}.$$

Examining the right-hand sides of (4.45) and (4.46) we impose the regularity assumption

$$u_{i,l} \in H^{r+1}(X) \tag{4.47}$$

for $1 \le i \le N$, $1 \le l \le L$.

LEMMA 4.4 Assume (4.47). Then

$$\left| \mathbf{a}\left(\boldsymbol{\delta}_i, \boldsymbol{\eta}_i\right) \right| \le c h^{r+\frac{1}{2}} \left(\sum_{l=1}^{L} w_l \left\| u_{i,l} \right\|_{H^{r+1}(X)}^2 \right)^{1/2} \left|\left|\left| \boldsymbol{\eta}_i \right|\right|\right|_h, \quad 1 \le i \le N.$$

Proof. By definition,

$$\mathbf{a}\left(\boldsymbol{\delta}_i, \boldsymbol{\eta}_i\right) = \sum_{l=1}^{L} w_l \left[\sum_{\rho \in \mathcal{E}_h} \int_\rho \boldsymbol{\omega}_l \cdot \boldsymbol{\nu}_\rho \widehat{\delta_{i,l}} [\![\eta_{i,l}]\!] \mathrm{d}x \right.$$
$$\left. - \sum_{K \in \mathcal{T}_h} \int_K \delta_{i,l} (\boldsymbol{\omega}_l \cdot \nabla_x \eta_{i,l}) \,\mathrm{d}x + \int_{\partial X_+^l} \boldsymbol{\omega}_l \cdot \boldsymbol{\nu} \delta_{i,l} \eta_{i,l} \,\mathrm{d}s \right].$$

Consider the second term. Note that $\eta_{i,l} = u_{i,l}^h - \pi_h u_{i,l} \in V_h$, implying that $\boldsymbol{\omega}_l \cdot \nabla_x \eta_{i,l} \in V_h$. Then from (4.44) we have

$$\sum_{K \in \mathcal{T}_h} \int_K \delta_{i,l} (\boldsymbol{\omega}_l \cdot \nabla_x \eta_{i,l}) \,\mathrm{d}x = \int_X \left(u_{i,l} - \pi_h u_{i,l}\right) (\boldsymbol{\omega}_l \cdot \nabla_x \eta_{i,l}) \,\mathrm{d}x = 0.$$

Therefore,

$$\mathbf{a}\left(\boldsymbol{\delta}_i, \boldsymbol{\eta}_i\right) = \sum_{l=1}^{L} w_l \left[\sum_{\rho \in \mathcal{E}_h} \int_\rho \boldsymbol{\omega}_l \cdot \boldsymbol{\nu}_\rho \widehat{\delta_{i,l}} [\![\eta_{i,l}]\!] \,\mathrm{d}x + \int_{\partial X_+^l} \boldsymbol{\omega}_l \cdot \boldsymbol{\nu} \delta_{i,l} \eta_{i,l} \,\mathrm{d}s \right].$$

Next consider the integration over the interior edges. Applying the triangle and Cauchy–Schwarz inequalities,

$$\sum_{\rho \in \mathcal{E}_h} \int_\rho \boldsymbol{\omega}_l \cdot \boldsymbol{\nu}_\rho \widehat{\delta_{i,l}} [\![\eta_{i,l}]\!] \,\mathrm{d}x \leq \sum_{\rho \in \mathcal{E}_h} \int_\rho |\boldsymbol{\omega}_l \cdot \boldsymbol{\nu}_\rho| |\delta_{i,l}| |[\![\eta_{i,l}]\!]| \,\mathrm{d}x$$
$$\leq \sum_{\rho \in \mathcal{E}_h} \left(\int_\rho |\boldsymbol{\omega}_l \cdot \boldsymbol{\nu}_\rho| \left(\delta_{i,l}\right)^2 \,\mathrm{d}x \right)^{1/2} \left(\int_\rho |\boldsymbol{\omega}_l \cdot \boldsymbol{\nu}_\rho| [\![\eta_{i,l}]\!]^2 \,\mathrm{d}x \right)^{1/2}.$$

Using that $|\boldsymbol{\omega}_l \cdot \boldsymbol{\nu}_\rho| \leq 1$, applying the Cauchy–Schwarz inequality and applying the trace inequality (4.46),

$$\sum_{\rho \in \mathcal{E}_h} \int_\rho \boldsymbol{\omega}_l \cdot \boldsymbol{\nu}_\rho \widehat{\delta_{i,l}} [\![\eta_{i,l}]\!] \,\mathrm{d}\mathbf{x} \leq \sum_{\rho \in \mathcal{E}_h} \left\| \delta_{i,l} \right\|_{L^2(\rho)} \left\| |\boldsymbol{\omega}_l \cdot \boldsymbol{\nu}_\rho|^{1/2} [\![\eta_{i,l}]\!] \right\|_{L^2(\rho)}$$
$$\leq c h^{r+\frac{1}{2}} \left(\sum_{K \in \mathcal{T}_h} \left\| u_{i,l} \right\|_{H^{r+1}(K)}^2 \right)^{1/2} \left(\sum_{\rho \in \mathcal{E}_h} \left\| |\boldsymbol{\omega}_l \cdot \boldsymbol{\nu}_\rho|^{1/2} [\![\eta_{i,l}]\!] \right\|_{L^2(\rho)}^2 \right)^{1/2}$$
$$\leq c h^{r+\frac{1}{2}} \left\| u_{i,l} \right\|_{H^{r+1}(X)} \left(\sum_{\rho \in \mathcal{E}_h} \left\| |\boldsymbol{\omega}_l \cdot \boldsymbol{\nu}_\rho|^{1/2} [\![\eta_{i,l}]\!] \right\|_{L^2(\rho)}^2 \right)^{1/2}.$$

Multiplying by w_l, summing over $1 \le l \le L$, applying the Cauchy–Schwarz inequality and using the definition of $|||\cdot|||_h$ given in (4.37) we obtain

$$\begin{aligned}
&\sum_{l=1}^{L} w_l \sum_{\rho\in\mathcal{E}_h} \int_\rho \boldsymbol{\omega}_l \cdot \boldsymbol{\nu}_\rho \widehat{\delta_{i,l}} [\![\eta_{i,l}]\!] \,\mathrm{d}x \\
&\quad \le ch^{r+\frac{1}{2}} \left(\sum_{l=1}^{L} w_l \left\| u_{i,l} \right\|^2_{H^{r+1}(X)} \right)^{1/2} \left(\sum_{l=1}^{L} w_l \sum_{\rho\in\mathcal{E}_h} \left\| |\boldsymbol{\omega}_l \cdot \boldsymbol{\nu}_\rho|^{1/2} [\![\eta_{i,l}]\!] \right\|^2_{L^2(\rho)} \right)^{1/2} \\
&\quad \le ch^{r+\frac{1}{2}} \left(\sum_{l=1}^{L} w_l \left\| u_{i,l} \right\|^2_{H^{r+1}(X)} \right)^{1/2} |||\boldsymbol{\eta}_i|||_h. \qquad (4.48)
\end{aligned}$$

For the integral term over the boundary we apply the Cauchy–Schwarz inequality to obtain

$$\begin{aligned}
\sum_{l=1}^{L} w_l \int_{\partial X_+^l} \boldsymbol{\omega}_l \cdot \boldsymbol{\nu} \delta_{i,l} \eta_{i,l} \,\mathrm{d}s &\le \left(\sum_{l=1}^{L} w_l \int_{\partial X_+^l} \boldsymbol{\omega}_l \cdot \boldsymbol{\nu} \left(\delta_{i,l}\right)^2 \,\mathrm{d}s \right)^{1/2} \left(\sum_{l=1}^{L} w_l \int_{\partial X_+^l} \boldsymbol{\omega}_l \cdot \boldsymbol{\nu} \left(\eta_{i,l}\right)^2 \,\mathrm{d}s \right)^{1/2} \\
&\le \left(\sum_{l=1}^{L} w_l \int_{\partial X_+^l} \boldsymbol{\omega}_l \cdot \boldsymbol{\nu} \left(\delta_{i,l}\right)^2 \,\mathrm{d}s \right)^{1/2} |||\boldsymbol{\eta}_i|||_h.
\end{aligned}$$

Noting $|\boldsymbol{\omega}_l \cdot \boldsymbol{\nu}| \le 1$ and using the trace inequality (4.46),

$$\int_{\partial X_+^l} \boldsymbol{\omega}_l \cdot \boldsymbol{\nu} \left(\delta_{i,l}\right)^2 \,\mathrm{d}s \le \sum_{\rho\in\partial X_+^l} \left\| \delta_{i,l} \right\|^2_{L^2(\rho)} \le ch^{2r+1} \sum_{K\in\mathcal{T}_h} \left\| u_{i,l} \right\|^2_{H^{r+1}(K)} = ch^{2r+1} \left\| u_{i,l} \right\|^2_{H^{r+1}(X)}.$$

This then leads to

$$\sum_{l=1}^{L} w_l \int_{\partial X_+^l} \boldsymbol{\omega}_l \cdot \boldsymbol{\nu} \delta_{i,l} \eta_{i,l} \,\mathrm{d}s \le ch^{r+\frac{1}{2}} \left(\sum_{l=1}^{L} w_l \left\| u_{i,l} \right\|^2_{H^{r+1}(X)} \right)^{1/2} |||\boldsymbol{\eta}_i|||_h. \qquad (4.49)$$

Combining inequalities (4.48) and (4.49) yields the desired inequality. □

LEMMA 4.5 Assume (4.47). Then

$$\left| \mathbf{b}_i \left(\boldsymbol{\delta}_i, \boldsymbol{\eta}_i\right) \right| \le ch^{r+1} |E_i| \left(\sum_{l=1}^{L} w_l \left\| u_{i,l} \right\|^2_{H^{r+1}(X)} \right)^{1/2} |||\boldsymbol{\eta}_i|||_h, \quad 1 \le i \le N.$$

Proof. By definition

$$\mathbf{b}_i\left(\boldsymbol{\delta}_i, \boldsymbol{\eta}_i\right) = \sum_{l=1}^{L} w_l \left[\sum_{K\in\mathcal{T}_h} \int_K \sigma_{t,i}\delta_{i,l}\eta_{i,l}\,\mathrm{d}x - \sum_{K\in\mathcal{T}_h} \int_K \eta_{i,l}|E_i|\sigma_{s,i} \sum_{k=1}^{L} w_{l,k}^{i,i}\delta_{i,k}\,\mathrm{d}x \right]. \tag{4.50}$$

Begin by considering the second term. Using the bound on $w_{l,k}^{i,i}$ given in (3.33) we obtain

$$\sum_{K\in\mathcal{T}_h} \int_K \eta_{i,l}\sigma_{s,i} \sum_{k=1}^{L} w_{l,k}^{i,i}\delta_{i,k}\,\mathrm{d}x \le c \sum_{K\in\mathcal{T}_h} \int_K |\eta_{i,l}| \sum_{k=1}^{L} w_k|\delta_{i,k}|\,\mathrm{d}x.$$

By repeated application of the Cauchy–Schwarz inequality,

$$\sum_{K\in\mathcal{T}_h} \int_K \eta_{i,l}\sigma_{s,i} \sum_{k=1}^{L} w_{l,k}^{i,i}\delta_{i,k}\,\mathrm{d}x \le c \left(\sum_{K\in\mathcal{T}_h} \left\|\eta_{i,l}\right\|_{L^2(K)}^2 \right)^{1/2} \left(\sum_{k=1}^{L} w_k \sum_{K\in\mathcal{T}_h} \left\|\delta_{i,k}\right\|_{L^2(K)}^2 \right)^{1/2}.$$

Applying inequality (4.45),

$$\sum_{K\in\mathcal{T}_h} \int_K \eta_{i,l}\sigma_{s,i} \sum_{k=1}^{L} w_{l,k}^{i,i}\delta_{i,k}\,\mathrm{d}x \le ch^{r+1} \left\|\eta_{i,l}\right\|_{L^2(X)} \left(\sum_{k=1}^{L} w_k \left\|u_{i,l}\right\|_{H^{r+1}(X)}^2 \right)^{1/2}.$$

Multiplying by w_l, summing over $1 \le l \le L$, applying the Cauchy–Schwarz inequality and renaming indices,

$$\begin{aligned}
&\sum_{l=1}^{L} w_l \sum_{K\in\mathcal{T}_h} \int_K \eta_{i,l}\sigma_{s,i} \sum_{k=1}^{L} w_{l,k}^{i,i}\delta_{i,k}\,\mathrm{d}x \\
&\quad \le ch^{r+1} \sum_{l=1}^{L} w_l \left\|\eta_{i,l}\right\|_{L^2(X)} \left(\sum_{k=1}^{L} w_k \left\|u_{i,l}\right\|_{H^{r+1}(X)}^2 \right)^{1/2} \\
&\quad \le ch^{r+1} \left(\sum_{l=1}^{L} w_l \left\|\eta_{i,l}\right\|_{L^2(X)}^2 \right)^{1/2} \left(\sum_{l=1}^{L} w_l \left\|u_{i,l}\right\|_{H^{r+1}(X)}^2 \right)^{1/2}.
\end{aligned}$$

Using the definition of $|||\cdot|||_h$,

$$\sum_{l=1}^{L} w_l \sum_{K\in\mathcal{T}_h} \int_K \eta_{i,l}|E_i|\sigma_{s,i} \sum_{k=1}^{L} w_{l,k}^{i,i}\delta_{i,k}\,\mathrm{d}x \le c|E_i|h^{r+1} \left(\sum_{l=1}^{L} w_l \left\|u_{i,l}\right\|_{H^{r+1}(X)}^2 \right)^{1/2} |||\boldsymbol{\eta}_i|||_h. \tag{4.51}$$

The remaining term in (4.50) can be dealt with using similar techniques:

$$\sum_{l=1}^{L} w_l \sum_{K\in\mathcal{T}_h} \int_K \sigma_{t,i}\delta_{i,l}\eta_{i,l}\,\mathrm{d}x \le c|E_i|h^{r+1}\left(\sum_{l=1}^{L} w_l \left\|u_{i,l}\right\|^2_{H^{r+1}(X)}\right)^{1/2} \left|\!\left|\!\left|\boldsymbol{\eta}_i\right|\!\right|\!\right|_h. \tag{4.52}$$

Combining (4.51) and (4.52) produces the claim. □

Next a Galerkin orthogonality type result is established. Consider

$$\mathbf{a}\big(\mathbf{u}_i^h,\mathbf{v}\big)+\mathbf{b}_i\big(\mathbf{u}_i^h,\mathbf{v}\big)=\sum_{j=1}^{i-1}\mathbf{c}_{i,j}\big(\mathbf{u}_j^h,\mathbf{v}\big)+\mathbf{l}_i(\mathbf{v}) \quad \forall\,\mathbf{v}\in(V_h)^L, \tag{4.53}$$

$$\mathbf{a}(\mathbf{u}_i,\mathbf{v})+\mathbf{b}_i(\mathbf{u}_i,\mathbf{v})=\sum_{j=1}^{i-1}\mathbf{c}_{i,j}\big(\mathbf{u}_j,\mathbf{v}\big)+\mathbf{l}_i(\mathbf{v}) \quad \forall\,\mathbf{v}\in(V_h)^L, \tag{4.54}$$

for $1\le i\le N$. Subtract equation (4.53) from (4.54):

$$\mathbf{a}\big(\mathbf{u}_i-\mathbf{u}_i^h,\mathbf{v}\big)+\mathbf{b}_i\big(\mathbf{u}_i-\mathbf{u}_i^h,\mathbf{v}\big)-\sum_{j=1}^{i-1}\mathbf{c}_{i,j}\big(\mathbf{u}_j-\mathbf{u}_j^h,\mathbf{v}\big)=0 \quad \forall\,\mathbf{v}\in(V_h)^L.$$

Observing that

$$\left|\mathbf{c}_{i,j}\big(\mathbf{u}_j-\mathbf{u}_j^h,\mathbf{v}\big)\right|\le\left|\mathbf{c}_{i,j}\big(\boldsymbol{\delta}_j,\mathbf{v}\big)\right|+\left|\mathbf{c}_{i,j}\big(\boldsymbol{\eta}_j,\mathbf{v}\big)\right|$$

we obtain

$$0\le\mathbf{a}\big(\mathbf{u}_i-\mathbf{u}_i^h,\mathbf{v}\big)+\mathbf{b}_i\big(\mathbf{u}_i-\mathbf{u}_i^h,\mathbf{v}\big)+\sum_{j=1}^{i-1}\left(\left|\mathbf{c}_{i,j}\big(\boldsymbol{\delta}_j,\mathbf{v}\big)\right|+\left|\mathbf{c}_{i,j}(\boldsymbol{\eta}_j,\mathbf{v})\right|\right) \tag{4.55}$$

for all $\mathbf{v}\in(V_h)^L$. Applying Lemma 4.2 with $\mathbf{v}=\boldsymbol{\eta}_i$,

$$c\left|\!\left|\!\left|\boldsymbol{\eta}_i\right|\!\right|\!\right|_h^2\le\mathbf{a}\left(\boldsymbol{\eta}_i,\boldsymbol{\eta}_i\right)+\mathbf{b}_i\left(\boldsymbol{\eta}_i,\boldsymbol{\eta}_i\right). \tag{4.56}$$

Adding (4.55) and (4.56), again with $\mathbf{v}=\boldsymbol{\eta}_i$, then yields

$$c\left|\!\left|\!\left|\boldsymbol{\eta}_i\right|\!\right|\!\right|_h^2\le\mathbf{a}\left(\boldsymbol{\delta}_i,\boldsymbol{\eta}_i\right)+\mathbf{b}_i\left(\boldsymbol{\delta}_i,\boldsymbol{\eta}_i\right)+\sum_{j=1}^{i-1}\left(\left|\mathbf{c}_{i,j}\left(\boldsymbol{\delta}_j,\boldsymbol{\eta}_i\right)\right|+\left|\mathbf{c}_{i,j}\big(\boldsymbol{\eta}_j,\boldsymbol{\eta}_i\big)\right|\right).$$

Applying Lemmas 4.4 and 4.5 results in

$$\left|\!\left|\!\left|\boldsymbol{\eta}_i\right|\!\right|\!\right|_h^2\le ch^{r+\frac{1}{2}}|E_i|\left(\sum_{l=1}^{L} w_l\left\|u_{i,l}\right\|^2_{H^{r+1}(X)}\right)^{1/2}\left|\!\left|\!\left|\boldsymbol{\eta}_i\right|\!\right|\!\right|_h+c\sum_{j=1}^{i-1}|E_j|\left(\left|\!\left|\!\left|\boldsymbol{\eta}_j\right|\!\right|\!\right|_h+\left|\!\left|\!\left|\boldsymbol{\delta}_j\right|\!\right|\!\right|_h\right)\left|\!\left|\!\left|\boldsymbol{\eta}_i\right|\!\right|\!\right|_h,$$

from which we obtain

$$\left|\left|\left|\boldsymbol{\eta}_i\right|\right|\right|_h \leq ch^{r+\frac{1}{2}} |E_i| \left(\sum_{l=1}^{L} w_l \left\|u_{i,l}\right\|^2_{H^{r+1}(X)}\right)^{1/2} + c\sum_{j=1}^{i-1} |E_j| \left(\left|\left|\left|\boldsymbol{\eta}_j\right|\right|\right|_h + \left|\left|\left|\boldsymbol{\delta}_j\right|\right|\right|_h\right).$$

Consider $\left|\left|\left|\boldsymbol{\delta}_j\right|\right|\right|_h$. From the definition of $|||\cdot|||_h$, application of inequalities (4.45) and (4.46) yields

$$\left|\left|\left|\boldsymbol{\delta}_j\right|\right|\right|_h \leq ch^{r+\frac{1}{2}} \left(\sum_{l=1}^{L} w_l \left\|u_{j,l}\right\|^2_{H^{r+1}(X)}\right)^{1/2}. \tag{4.57}$$

We then have

$$\left|\left|\left|\boldsymbol{\eta}_i\right|\right|\right|_h \leq ch^{r+\frac{1}{2}} \sum_{k=1}^{i} |E_k| \left(\sum_{l=1}^{L} w_l \left\|u_{k,l}\right\|^2_{H^{r+1}(X)}\right)^{1/2} + c\sum_{j=1}^{i-1} |E_j| \left|\left|\left|\boldsymbol{\eta}_j\right|\right|\right|_h.$$

Applying the discrete Gronwall inequality and extending the summation from i to N yields

$$\left|\left|\left|\boldsymbol{\eta}_i\right|\right|\right|_h \leq ch^{r+\frac{1}{2}} \sum_{k=1}^{N} |E_k| \left(\sum_{l=1}^{L} w_l \left\|u_{k,l}\right\|^2_{H^{r+1}(X)}\right)^{1/2}.$$

Squaring both sides, multiplying by $|E_i|$, summing over $1 \leq i \leq N$ and taking the square root yields

$$\left(\sum_{i=1}^{N} |E_i| \left|\left|\left|\boldsymbol{\eta}_i\right|\right|\right|^2_h\right)^{1/2} \leq ch^{r+\frac{1}{2}} \left(\sum_{i=1}^{N} |E_i| \sum_{l=1}^{L} w_l \left\|u_{i,l}\right\|^2_{H^{r+1}(X)}\right)^{1/2}. \tag{4.58}$$

The same procedure applied to inequality (4.57) results in

$$\left(\sum_{i=1}^{N} |E_i| |||\boldsymbol{\delta}_i|||^2_h\right)^{1/2} \leq ch^{r+\frac{1}{2}} \left(\sum_{i=1}^{N} |E_i| \sum_{l=1}^{L} w_l \left\|u_{i,l}\right\|^2_{H^{r+1}(X)}\right)^{1/2}. \tag{4.59}$$

With inequalities (4.58) and (4.59), an estimate of the error between solutions of the angular semidiscretization and the fully discrete system is available.

THEOREM 4.6 (Fully discrete to semidiscrete). Assume that $\sigma_t \geq m_{\sigma_t} > 0$ and that (4.47) holds. Let $\{u_{i,l}\}_{i=1,l=1}^{N,L}$ satisfy (3.16) and (3.17) and $\mathbf{u}_h = \{u^h_{i,l}\}_{i=1,l=1}^{N,L}$ satisfy (4.36). Then for h_e sufficiently small,

$$\left\|\{u_{i,l}\}_{i=1,l=1}^{N,L} - \mathbf{u}_h\right\|_h \leq ch^{r+\frac{1}{2}} \left(\sum_{i=1}^{N} |E_i| \sum_{l=1}^{L} w_l \left\|u_{i,l}\right\|^2_{H^{r+1}(X)}\right)^{1/2},$$

where the norm $\|\cdot\|_h$ is defined in (4.43).

Proof. Using the definitions of $\boldsymbol{\delta}_i$ and $\boldsymbol{\eta}_i$ followed by the triangle inequality produces

$$\left(\sum_{i=1}^{N} |E_i| \left|\left|\left| \{u_{i,l}\}_{l=1}^{L} - \{u_{i,l}^h\}_{l=1}^{L} \right|\right|\right|_h^2 \right)^{1/2} \leq \left(\sum_{i=1}^{N} |E_i| |||\boldsymbol{\delta}_i|||_h \right)^{1/2} + \left(\sum_{i=1}^{N} |E_i| |||\boldsymbol{\eta}_i|||_h \right)^{1/2}.$$

The claim then follows from inequalities (4.58) and (4.59). □

With the results obtained in this section and Section 3 an error bound between the original BVPs (1.1) and (1.2) and the fully discrete method (4.36) can be established.

THEOREM 4.7 Assume the conditions of Theorems 3.12 and 4.6 hold. Then

$$\left\| \{u(\cdot, \boldsymbol{\omega}_l, e_i)\}_{i=1,l=1}^{N,L} - \mathbf{u}_h \right\| \leq C \left(h_e + h_a^2 + h^{r+\frac{1}{2}} \right),$$

where the dependence of the constant C on u and f is given in (4.60) below.

Proof. Application of the triangle inequality and using the definition of the norm $\|\cdot\|_h$ yields

$$\left\| \{u(\cdot, \boldsymbol{\omega}_l, e_i)\}_{i=1,l=1}^{N,L} - \mathbf{u}_h \right\| \leq \left\| \{u(\cdot, \boldsymbol{\omega}_l, e_i)\}_{i=1,l=1}^{N,L} - \{u_{i,l}\}_{i=1,l=1}^{N,L} \right\| + \left\| \{u_{i,l}\}_{i=1,l=1}^{N,L} - \mathbf{u}_h \right\|_h.$$

Then applying Theorems 3.12 and 4.6 leads to

$$\begin{aligned} &\left\| \{u(\cdot, \boldsymbol{\omega}_l, e_i)\}_{i=1,l=1}^{N,L} - \mathbf{u}_h \right\| \\ &\leq c \left[h_e \left(\|u\|_{L^\infty(U)} + \|u\|_{L^2\left(X \times \Omega, H^1(E)\right)} + \left(\sum_{l=1}^{L} w_l |f(\boldsymbol{\omega}_l)|_{L^2\left(X, H^1(E)\right)}^2 \right)^{1/2} \right) \right. \\ &\quad \left. + h_a^2 \left(\|u\|_{L^\infty(U)} + \|u\|_{L^2\left(X \times E, H^2(\Omega)\right)} \right) + h^{r+\frac{1}{2}} \left(\sum_{i=1}^{N} |E_i| \sum_{l=1}^{L} w_l \left\| u_{i,l} \right\|_{H^{r+1}(X)}^2 \right)^{1/2} \right]. \end{aligned} \tag{4.60}$$

Given the regularity assumptions on u and f the result follows. □

5. Numerical examples

In this section we present numerical examples of the method proposed in the previous sections. We begin by presenting an analytic scattering kernel describing the energy change due to a collision. We will denote this by $\mathcal{P}_e(e_i \to e_f)$, where energies e_i and e_f denote the energy before and after a collision, respectively. A common example of such a function in the context of neutron transport (cf. Lamarsh, 1966; Duderstadt & Hamilton, 1976; Lewis & Miller, 1984) is given by

$$\mathcal{P}_e(e_i \to e_f) := \begin{cases} \frac{1}{(1-\alpha)e_i}, & e_i\alpha \le e_f \le e_i, \\ 0 & \text{otherwise,} \end{cases}$$

where $0 \le \alpha < 1$. If we assume the scattering is isotropic the function $\mathcal{P}$ can be writte

$$\mathcal{P}(x, \boldsymbol{\omega}, \boldsymbol{\omega}', e, e') := \frac{1}{4\pi}\mathcal{P}_e(e' \to e) \tag{5.61}$$

and we note that $\mathcal{P}$ satisfies assumptions (1.5) and (1.6). In each example we take the scattering distribution $\mathcal{P}$ to be given by (5.61) with $\alpha = 0$.

In all four numerical examples below, the spatial and energy domains are $X = [0, 1]^3$ and $E = [1, 2]$, respectively. Further, the angular discretization will be fixed and chosen such that it is sufficiently fine to not interfere in the convergence orders in the spatial and energy mesh sizes; the angular discretization projects four triangles onto the sphere in each octant, the number of directions is then $|\Omega_{h_a}| = 18$. The resulting linear systems are solved using Gauss–Seidel iteration. In all of the numerical experiments the norm $\|\cdot\|$ as defined in (3.18) is used, with $\mathbf{u}_h = \{u^h_{i,l}\}_{i=1,l=1}^{N,L}$ denoting the solution of the fully discrete problem and u the solution of the original BVP. For a given level of discretization h let h^+ denote the adjacent finer level of discretization. Then the numerical convergence order is computed by $\mathrm{Order}(h) = \log_{\frac{h}{h^+}}(\frac{\mathrm{Error}(h)}{\mathrm{Error}(h^+)})$. Here $\mathrm{Error}(h)$ denotes the error computed with respect to the norm $\|\cdot\|$ defined in (3.18) and it is understood that the remaining discretization parameters are taken sufficiently small. Piecewise linear approximation is used in the DG scheme; this corresponds to $r = 1$ in definition (4.34).

EXAMPLE 5.1. In this example $\sigma_s(x, e) = 2$ and $\sigma_a(x, e) = 1$. The solution is given by $u(x, \boldsymbol{\omega}, e) = \sin(x_1\pi)\sin(x_2\pi)\sin(x_3\pi)e$, with f and g being determined by substitution of u into the BVP. Consulting Table 1 and Figs 3 and 4 it appears that convergence with respect to the spatial discretization approaches quadratic while linear convergence with respect to the energy mesh is observed.

TABLE 1. *The error and convergence order for Example 1 in both h_e and h. (a) Fixed energy mesh of h_e=1/1024 with various mesh sizes h. (b) The spatial variable in the DG scheme uses cubic interpolation with $h = \sqrt{2}/5$*

(a)			(b)		
h	$\|u-\mathbf{u}_h\|$	Order(h)	h_e	$\|u-\mathbf{u}_h\|$	Order(h_e)
$\frac{\sqrt{2}}{2}$	1.9961e−01	1.8313	$\frac{1}{4}$	9.0796e−02	1.0002
$\frac{\sqrt{2}}{4}$	5.6092e−02	1.9038	$\frac{1}{8}$	4.5392e−02	9.9917e−01
$\frac{\sqrt{2}}{6}$	2.5921e−02	1.9095	$\frac{1}{16}$	2.2709e−02	9.9740e−01
$\frac{\sqrt{2}}{8}$	1.4965e−02	1.9545	$\frac{1}{32}$	1.1375e−02	9.9264e−01
$\frac{\sqrt{2}}{10}$	9.6753e−03	–	$\frac{1}{64}$	5.7166e−03	9.7796e−01
			$\frac{1}{128}$	2.9023e−03	–

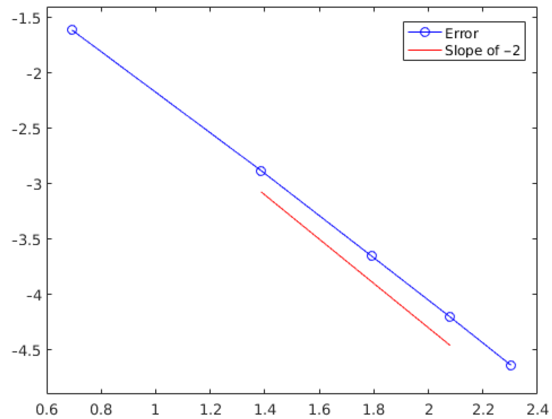

FIG. 3. Log-log plot of the spatial error for Example 1.

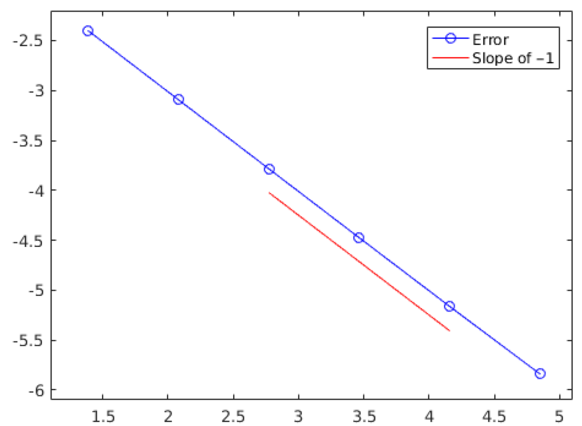

FIG. 4. Log-log plot of the energy error for Example 1.

EXAMPLE 5.2. Choose $u(x, \boldsymbol{\omega}, e) = \sin(x_1)\sin(x_2)\sin(x_3)$ with cross sections $\sigma_s(x, e) = 1$ and $\sigma_a(x, e) = 0.5$, where f and g are determined through the BVP. The problem remains energy dependent due to the scattering term. To determine the convergence order of the energy discretization, the spatial discretization uses cubic interpolation with $h = \sqrt{2}/5$. Probably due to the minimal energy dependence, use of this spatial discretization significantly reduced the overall error and a finer resolution in the energy domain is needed for determining the convergence order in energy. This can be seen in Table 2.

In this example it is clear from Fig. 5 and Table 2(a) that quadratic convergence with respect to the spatial mesh is obtained, while Fig. 6 and Table 2(b) again indicate linear convergence with respect to the energy mesh.

EXAMPLE 5.3. In this example the cross sections are allowed to depend on the energy variable. We use $\sigma_t(x, e) = \ln(e)$ and $\sigma_s(x, e) = 0.5\sigma_t(x, e)$ and choose the true solution $u(x, \boldsymbol{\omega}, e) = \ln(e)\exp(\omega_1 + \sin(\pi x_1)\sin(\pi x_2)\sin(\pi x_3))$. The source and boundary data are determined through the BVP.

TABLE 2. *The error and convergence order for Example 2 in both h_e and h. (a) Fixed energy mesh of h_e=1/256 with various mesh sizes h. (b) The spatial variable in the DG scheme uses cubic interpolation with $h = \sqrt{2}/5$*

(a)

h	$\|u - \mathbf{u}_h\|$	Order(h)
$\frac{\sqrt{2}}{2}$	4.0374e−02	2.0442
$\frac{\sqrt{2}}{4}$	9.7890e−03	2.0340
$\frac{\sqrt{2}}{6}$	4.2911e−03	1.8962
$\frac{\sqrt{2}}{8}$	2.4869e−03	1.9751
$\frac{\sqrt{2}}{10}$	1.6005e−03	–

(b)

h_e	$\|u - \mathbf{u}_h\|$	Order(h_e)
$\frac{1}{100}$	4.0750e−05	1.0267
$\frac{1}{200}$	2.0002e−05	1.0354
$\frac{1}{300}$	1.3144e−05	1.0340
$\frac{1}{400}$	9.7623e−06	1.0222
$\frac{1}{500}$	7.7713e−06	9.9973e−01
$\frac{1}{600}$	6.4764e−06	–

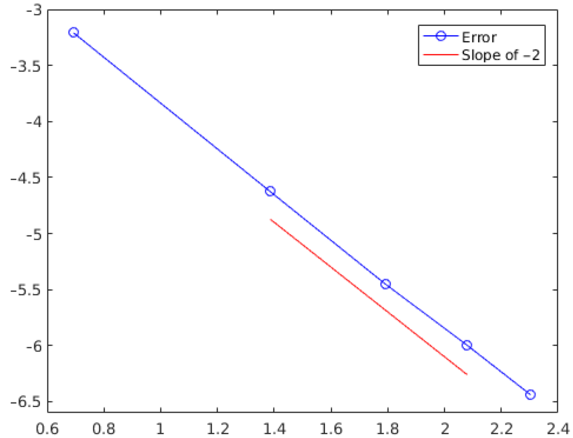

FIG. 5. Log-log plot of the spatial error for Example 2.

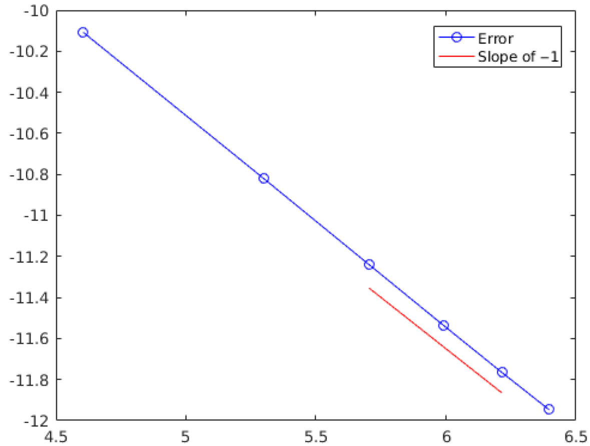

FIG. 6. Log-log plot of the energy error for Example 2.

TABLE 3. *The error and convergence order for Example 3 in both h_e and h. (a) Fixed energy mesh of h_e=1/1024 with various mesh sizes h. (b) The spatial variable in the DG scheme uses cubic interpolation with $h = \sqrt{2}/5$*

(a)

h	$\|u - \mathbf{u}_h\|$	Order(h)
$\frac{\sqrt{2}}{2}$	1.1533e−01	1.6923
$\frac{\sqrt{2}}{4}$	3.5686e−02	1.9397
$\frac{\sqrt{2}}{6}$	1.6253e−02	1.9398
$\frac{\sqrt{2}}{8}$	9.3020e−03	1.9071
$\frac{\sqrt{2}}{10}$	6.0780e−03	–

(b)

h_e	$\|u - \mathbf{u}_h\|$	Order(h_e)
$\frac{1}{4}$	3.3669e−01	9.9765e−01
$\frac{1}{8}$	1.6862e−01	9.9942e−01
$\frac{1}{16}$	8.4344e−02	9.9979e−01
$\frac{1}{32}$	4.2178e−02	9.9979e−01
$\frac{1}{64}$	2.1092e−02	9.9959e−01
$\frac{1}{128}$	1.0549e−02	–

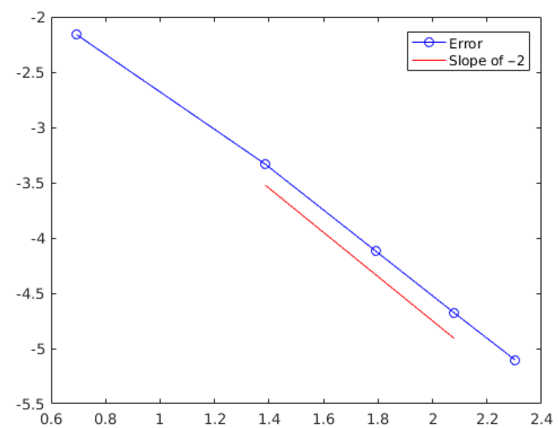

FIG. 7. Log-log plot of the spatial error for Example 3.

Consulting Table 3 and Figs 7 and 8 we find behavior similar to Example 1: quadratic convergence with respect to the spatial mesh and linear convergence with respect to the energy mesh.

EXAMPLE 5.4. In this example the cross sections depend on the energy variable and $\sigma_a = 0$ when $e = 1.5$. We use $\sigma_t(x, e) = (e - 1.5)^2 + 1$ and $\sigma_s(x, e) = 1$ and choose the solution $u(x, \boldsymbol{\omega}, e) = \ln(e)\sin(x_1)\sin(x_2)\sin(x_3)$. The source and boundary data are determined through the BVP.

Consulting Table 4 and Figs 9 and 10 we find good agreement with the previous examples. In all four numerical examples first-order convergence with respect to the energy mesh size h_e is observed. This agrees with the theoretical results presented in the preceding sections. Additionally, the theoretical results in Section 4 showed a convergence order of $h^{1.5}$. In the numerical examples, however, second-order convergence in the spatial mesh size h was observed. This is not completely unexpected; numerical methods applying DG methods to the monoenergetic RTE also observe this behavior (cf. Han *et al.*, 2010; Gao & Zhao, 2013). Therefore, our results agree with previously used DG methods.

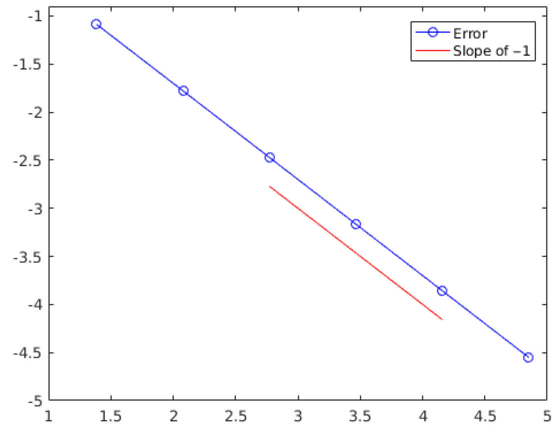

FIG. 8. Log-log plot of the energy error for Example 3.

TABLE 4. *The error and convergence order for Example 4 in both h_e and h. (a) Fixed energy mesh of h_e=1/1024 with various mesh sizes h. (b) The spatial variable in the DG scheme uses cubic interpolation with $h = \sqrt{2}/5$*

(a)			(b)		
h	$\|u - \mathbf{u}_h\|$	Order(h)	h_e	$\|u - \mathbf{u}_h\|$	Order(h_e)
$\frac{\sqrt{2}}{2}$	1.7696e−02	2.0359	$\frac{1}{4}$	2.5725e−02	9.9793e−01
$\frac{\sqrt{2}}{4}$	4.3154e−03	2.0087	$\frac{1}{8}$	1.2881e−02	9.9933e−01
$\frac{\sqrt{2}}{6}$	1.9112e−03	1.9741	$\frac{1}{16}$	6.4435e−03	9.9949e−01
$\frac{\sqrt{2}}{8}$	1.0831e−03	1.9244	$\frac{1}{32}$	3.2229e−03	9.9870e−01
$\frac{\sqrt{2}}{10}$	7.0497e−04	–	$\frac{1}{64}$	1.6129e−03	9.9614e−01
			$\frac{1}{128}$	8.0861e−04	–

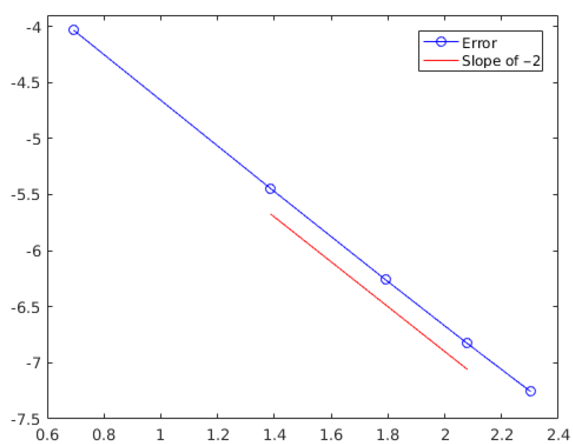

FIG. 9. Log-log plot of the spatial error for Example 4.

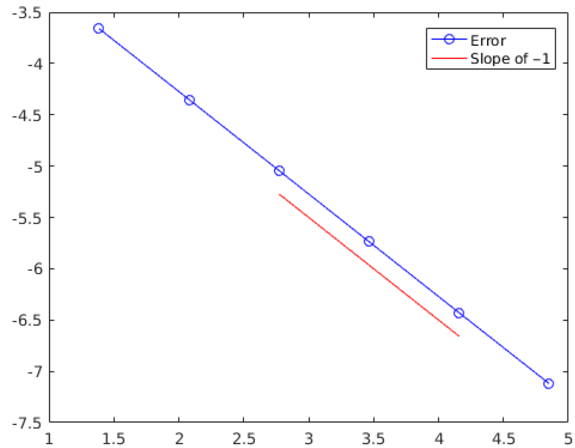

FIG. 10. Log-log plot of the energy error for Example 4.

Acknowledgements

The authors thank the anonymous referees for their valuable comments.

REFERENCES

AGOSHKOV, V. (1998) *Boundary Value Problems for Transport Equations*. Boston: Birkhäuser.

ARRIDGE, S. R. (1999) Optical tomography in medical imaging. *Inverse Probl.*, **15**, R41–R93.

AZMY, Y. & SARTORI, E. (2010) *Nuclear Computational Science: A Century in Review*. Netherlands: Springer.

BAUDRON, A. M., LAUTARD, J. J., MADAY, Y. & MULA, O. (2014) MINARET: towards a parallel 3D time-dependent neutron transport solver. *SNA + MC 2013 - Joint International Conference on Supercomputing in Nuclear Applications + Monte Carlo*, 04103. https://doi.org/10.1051/snamc/201404103

BAUMGARDNER, J. R. & FREDERICKSON, P. O. (1985) Icosahedral discretization of the two-sphere. *SIAM J. Numer. Anal.*, **22**, 1107–1115.

BREZZI, F., MARINI, L. D. & SÜLI, E. (2004) Discontinuous Galerkin methods for first-order hyperbolic problems. *Math. Models Methods Appl. Sci.*, **14**, 1893–1903.

CHANDRASEKHAR, S. (1960) *Radiative Transfer*. New York: Dover Publications.

CLARK, D. S. (1987) Short proof of a discrete Gronwall inequality. *Discrete Appl. Math.*, **16**, 279–281.

CZUPRYNSKI, K. & HAN, W. (2017) Energy dependent radiative transfer equation and energy discretization. *J. Comput. Appl. Math.*, **323**, 147–158.

DI PIETRO, D. A. & ERN, A. (2012) *Mathematical Aspects of Discontinuous Galerkin Methods*. Berlin: Springer.

DUDERSTADT, J. & HAMILTON, L. (1976) *Nuclear Reactor Analysis*. New York: John Wiley.

DUDERSTADT, J. & MARTIN, W. (1979) *Transport Theory*. New York: John Wiley.

GAO, H. & ZHAO, H. (2009) A fast forward solver of radiative transfer equation. *Transport Theory Statist. Phys.*, **38**, 149–192.

GAO, H. & ZHAO, H. (2013) Analysis of a numerical solver for radiative transport equation. *Math. Comp.*, **82**, 153–172.

HAN, W., HUANG, J. & EICHHOLZ, J. (2010) Discrete-ordinate discontinuous Galerkin methods for solving the radiative transfer equation. *SIAM J. Sci. Comput.*, **32**, 477–497.

HEBEY, E. (1991) *Sobolev Spaces on Riemannian Manifolds*. Lecture Notes in Mathematics, vol. 1635. Berlin: Springer.

KIM, A. D. & MOSCOSO, M. (2006) Radiative transport theory for optical molecular imaging. *Inverse Probl.*, **22**, 23–42.

LAMARSH, J. R. (1966) *Introduction to Nuclear Reactor Theory*. Reading, MA: Addison-Wesley.

LEWIS, E. E. & MILLER, W. F. (1984) *Computational Methods of Neutron Transport*. New York: John Wiley.

MEIR, A. J. & TUNCER, N. (2009) Radially projected finite elements. *SIAM J. Sci. Comput.*, **31**, 2368–2385.

MODEST, M. F. (2013) *Radiative Heat Transfer*, 3rd edn. Oxford, UK: Academic Press.

SEUBERT, A., SUREDA, A., BADER, J., LAPINS, J., BUCK, M. & LAURIEN, E. (2012) The 3-D time-dependent transport code TORT-TD and its coupling with the 3D thermal-hydraulic code ATTICA3D for HTGR applications. *Nucl. Eng. Des.*, **251**, 173–180.

UILKEMA, S. B. (2012) Proton therapy planning using the S_N method with the Fokker–Planck approximation. *Master's Thesis*, Delft: Delft University of Technology.