Summer 2011

Discontinuous Galerkin methods for the radiative transfer equation and its approximations

Joseph A. Eichholz
University of Iowa

DISCONTINUOUS GALERKIN METHODS FOR THE RADIATIVE TRANSFER EQUATION AND ITS APPROXIMATIONS

by

Joseph A. Eichholz

An Abstract

Of a thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Applied Mathematical and Computational Sciences
in the Graduate College of
The University of Iowa

July 2011

Thesis Supervisor: Professor Weimin Han

ABSTRACT

Radiative transfer theory describes the interaction of radiation with scattering and absorbing media. It has applications in neutron transport, atmospheric physics, heat transfer, molecular imaging, and others. In steady state, the radiative transfer equation is an integro-differential equation of five independent variables. This high dimensionality and presence of integral term present a serious challenge when trying to solve the equation numerically. Over the past 50 years, several techniques for solving the radiative transfer equation have been introduced. These include, but are certainly not limited to, Monte Carlo methods, discrete-ordinate methods, spherical harmonics methods, spectral methods, finite difference methods, and finite element methods. Methods involving discrete ordinates have received particular attention in the literature due to their relatively high accuracy, flexibility, and relatively low computational cost.

In this thesis we present a discrete-ordinate discontinuous Galerkin method for solving the radiative transfer equation. In addition, we present a generalized Fokker-Planck equation that may be used to approximate the radiative transfer equation in certain circumstances. We provide well posedness results for this approximation, and introduce a discrete-ordinate discontinuous Galerkin method to approximate a solution. Theoretical error estimates are derived, and numerical examples demonstrating the efficacy of the methods are given.

Abstract Approved: _____

Thesis Supervisor

_____

Title and Department

_____

Date

DISCONTINUOUS GALERKIN METHODS FOR THE RADIATIVE TRANSFER EQUATION AND ITS APPROXIMATIONS

by

Joseph A. Eichholz

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Applied Mathematical and Computational Sciences
in the Graduate College of
The University of Iowa

July 2011

Thesis Supervisor: Professor Weimin Han

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Joseph A. Eichholz

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Applied Mathematical and Computational Sciences at the July 2011 graduation.

Thesis Committee: _____
Weimin Han, Thesis Supervisor

_____
Laurent Jay

_____
Yi Li

_____
Suely Oliveira

_____
David Stewart

To my parents, Kenneth and Carol Eichholz. This brief dedication is not enough room to express my gratitude that they are my parents.

ACKNOWLEDGEMENTS

I need to thank several people for their help making this thesis possible.

First, I would like to thank my advisor, Professor Weimin Han, for his excellent guidance, constant encouragement, and endless patience. His excellent mentorship will continue to shape my career long after I leave Iowa.

I would also like to thank my committee members, Professor Laurent Jay, Professor Yi Li, Professor Suely Oliveira, and Professor David Stewart for their time reviewing my work. In addition, many of my committee members were also excellent instructors who greatly influenced me during my early graduate career.

I thank Professor Kendall Atkinson for all his excellent advice.

I need to thank my friends and colleagues at Iowa. In particular, I would like to thank Scott Small and Stephen Welch for continuing to listen to all my ideas long after it became apparent that most of them don't lead anywhere.

Finally, I give thanks to my entire family for their endless encouragement. Their support has truly been a driving factor in my life.

ABSTRACT

Radiative transfer theory describes the interaction of radiation with scattering and absorbing media. It has applications in neutron transport, atmospheric physics, heat transfer, molecular imaging, and others. In steady state, the radiative transfer equation is an integro-differential equation of five independent variables. This high dimensionality and presence of integral term present a serious challenge when trying to solve the equation numerically. Over the past 50 years, several techniques for solving the radiative transfer equation have been introduced. These include, but are certainly not limited to, Monte Carlo methods, discrete-ordinate methods, spherical harmonics methods, spectral methods, finite difference methods, and finite element methods. Methods involving discrete ordinates have received particular attention in the literature due to their relatively high accuracy, flexibility, and relatively low computational cost.

In this thesis we present a discrete-ordinate discontinuous Galerkin method for solving the radiative transfer equation. In addition, we present a generalized Fokker-Planck equation that may be used to approximate the radiative transfer equation in certain circumstances. We provide well posedness results for this approximation, and introduce a discrete-ordinate discontinuous Galerkin method to approximate a solution. Theoretical error estimates are derived, and numerical examples demonstrating the efficacy of the methods are given.

TABLE OF CONTENTS

LIST OF TABLES vi

LIST OF FIGURES vii

CHAPTER

1 INTRODUCTION 1

1.1 Motivation 1
1.2 Sobolev spaces 5
1.3 Notation 9
1.4 The radiative transfer equation 12

2 DISCRETE-ORDINATE DISCONTINUOUS GALERKIN METHODS FOR THE RADIATIVE TRANSFER EQUATION 18

2.1 Introduction 18
2.2 Discrete-ordinate discontinuous Galerkin methods 19
2.3 Error analysis 25
2.4 Numerical results 39

3 GENERALIZED FOKKER-PLANCK EQUATION 57

3.1 The generalized Fokker-Planck equation 57
3.2 Well-posedness of the GFPE 59
3.3 An iteration method and its convergence 66
3.4 Discretizations 68
3.5 Numerical examples 73

4 CONCLUDING REMARKS AND FURTHER WORK 80

REFERENCES

LIST OF TABLES

2.1 Value of m for several choices of η and N 40

2.2 $||u - u_h||$ for Example 2.4.1, S_4 quadrature 42

2.3 $|||u - u_h|||$ for Example 2.4.1, S_4 quadrature 42

2.4 $||u - u_h||$ for Example 2.4.2, $\eta = .1$, S_4 quadrature 45

2.5 $|||u - u_h|||$ for Example 2.4.2, $\eta = .1$, S_4 quadrature 45

2.6 $||u - u_h||$ for Example 2.4.2, $\eta = .1$, S_{12} quadrature 45

2.7 $|||u - u_h|||$ for Example 2.4.2, $\eta = .1$, S_{12} quadrature 47

2.8 $||u - u_h||$ for Example 2.4.2, $\eta = .5$, S_4 quadrature 49

2.9 $|||u - u_h|||$ for Example 2.4.2, $\eta = .5$, S_4 quadrature 49

2.10 $||u - u_h||$ for Example 2.4.2, $\eta = .5$, S_{12} quadrature 49

2.11 $|||u - u_h|||$ for Example 2.4.2, $\eta = .5$, S_{12} quadrature 51

2.12 $||u - u_h||$ for Example 2.4.3, 18 angular nodes 53

2.13 $||u - u_h||$ for Example 2.4.3, 66 angular nodes 54

2.14 $||u - u_h||$ for Example 2.4.3, 258 angular nodes 55

3.1 $L^2(\Omega)$ error for Example 3.5.1 74

3.2 Example 3.5.2: L^2 error for several values of h and n_θ with $n_\psi = 6$ 76

3.3 Example 3.5.3: $||u - u_h||$ for different values of h and n_θ 77

LIST OF FIGURES

1.1 An illustration of $X_{\boldsymbol{\omega}}$ 12

1.2 A two dimensional example of $\partial X_{\boldsymbol{\omega},\pm}$ 13

2.1 A sample mesh 41

2.2 $||u - u_h||$ for Example 2.4.1, S_4 quadrature 43

2.3 $|||u - u_h|||$ for Example 2.4.1, S_4 quadrature 43

2.4 $||u - u_h||$ for Example 2.4.2, $\eta = .1$, S_4 quadrature 46

2.5 $|||u - u_h|||$ for Example 2.4.2, $\eta = .1$, S_4 quadrature 46

2.6 $||u - u_h||$ for Example 2.4.2, $\eta = .1$, S_{12} quadrature 47

2.7 $|||u - u_h|||$ for Example 2.4.2, $\eta = .1$, S_{12} quadrature 48

2.8 $||u - u_h||$ for Example 2.4.2, $\eta = .5$, S_4 quadrature 50

2.9 $|||u - u_h|||$ for Example 2.4.2, $\eta = .5$, S_4 quadrature 50

2.10 $||u - u_h||$ for Example 2.4.2, $\eta = .5$, S_{12} quadrature 51

2.11 $|||u - u_h|||$ for Example 2.4.2, $\eta = .5$, S_{12} quadrature 52

2.12 $||u - u_h||$ for Example 2.4.3, 18 angular nodes 53

2.13 $||u - u_h||$ for Example 2.4.3, 66 angular nodes 54

2.14 $||u - u_h||$ for Example 2.4.3, 258 angular nodes 55

3.1 Example 3.5.1: Numerical solution with $n_\theta = 128$ 75

3.2 Example 3.5.1: True solution 75

3.3 $||u - u_h||$ for Example 3.5.3, $n_\theta = 4$ 78

3.4 $||u - u_h||$ for Example 3.5.3, $n_\theta = 8$ 78

3.5 $||u - u_h||$ for Example 3.5.3, $n_\theta = 16$ 79

CHAPTER 1
INTRODUCTION

1.1 Motivation

Radiative transfer theory describes the interaction of radiation with scattering and absorbing media, which has wide applications in such areas as neutron transport, heat transfer, stellar atmospheres, optical molecular imaging, infrared and visible light in space and the atmosphere and so on. We refer to [3, 10, 30, 44, 47, 55, 56] and references therein for details about this subject.

The main focus of our research is the application of the radiative transfer equation in biomedical imaging; one particular potential application is in novel techniques for the detection of breast cancer. X-ray mammography is currently the most prevalent imaging modality for screening and diagnosis of breast cancers. The use of mammography results in a 25–30% decreased mortality rate in screened women ([46]). However, a multi-institutional trial funded by the American College of Radiology Imaging Network (ACRIN) suggested that about 30% of cancers were not detected by screening mammography, and 70–90% of biopsies performed based on suspicious mammograms were negative ([35, 50]). Some false negative and false positive diagnoses often led to missed cancers and inappropriate biopsies. The key factor that limits the success rate is the poor contrast between healthy and diseased tissues in the mammogram. Although x-ray CT of the breast can potentially improve diagnostic accuracy over mammography ([28, 29]), the state-of-the-art breast CT scanner

is still based on the attenuation mechanism. As a result, the use of breast CT requires an intravenous contrast medium and a high radiation dose, since elemental composition is almost uniform with little density variation in breast tissues. Still, it is rather difficult for breast CT to discern early-stage breast cancers.

The main components of the breast are (1) the adipose tissue consisting of large fatty cells and (2) the connective tissue containing fibrous collagen. Fibril ordering within the tumors is basically absent. Tissue remodeling is a crucial step during a malignant transformation of epithelial cells. Carcinoma invasion is accompanied by destruction and synthesis of fibrillary and non-fibrillary matrix proteins, creating characteristic desmoplastic stromal changes. The dense collagenous tumor core is firm in consistency, whereas the invasive tumor front is rich in looser fibrils. Invasive carcinoma cells migrate along collagen strands. These are attributed to structural degradation of invaded collagen or a low degree of ordering of the newly formed collagen among the adipocytes in the fatty tissue, where collagen is accompanied by invading cancer cells. Hence, a significant structural variation occurs with respect to supramolecular collagen architecture in tumor, resulting in a significant difference of x-ray scattering behaviors between tumor and normal tissues ([25, 51]). X-ray scattering is predominantly Rayleigh scattering ([41]). Every tissue component gives rise to a characteristic x-ray scattering pattern. Such scattering signals reveal critical information on ultrafine features of cellular and sub-cellular structures, and bear unique diagnostic values of molecules, cells and their clusters which are also known as basic functional units of 100–200 m in diameter ([9, 45]). Most importantly,

x-ray scattering properties of tumors are significantly different from that of healthy tissue. Hence, x-ray Rayleigh scattering imaging provides a new contrast mechanism and would well complement attenuation-based x-ray imaging ([45, 16, 49, 48]). Thus, imaging techniques incorporating scattering information can potentially provide significantly higher contrast than conventional imaging modalities.

In principal, given a model of light propagation dependent upon scattering, we may recover the scattering by matching measurements of light escaping the domain in experiment to predictions from the model with varying scattering parameters. In stationary one-velocity case, the photon (neutron) intensity, a function of three space variables and two angular variables, is governed mathematically by the radiative transfer equation (RTE) (cf. [3, 40]). Since the use of the RTE in molecular imaging will amount to solving an inverse problem with the RTE as the forward problem, fast and efficient methods of solving the RTE must be explored.

The RTE may be viewed as a hyperbolic type integro-differential equation. Due to this complication and the high dimension of the problem, it is a serious challenge to develop effective numerical solution methods; this topic has attracted much attention in the past five decades. The numerical methods used today include Monte Carlo methods, discrete ordinate methods, methods using spherical harmonics, finite difference methods, finite element methods and spectral methods and so on. We refer to [40] for an overview. Among all the methods just mentioned, the discrete ordinate methods (cf. [14, 40]) have received significant attention and development, owing to the good compromise between accuracy, flexibility and moderate computational

requirements. A number of discrete ordinate methods combined with Chebyshev spectral methods were introduced in [6, 23, 37]; in [15], a spatial multigrid algorithm was presented for isotropic neutron transport in x-y geometry, where the linear system to be solved is obtained using discrete ordinates in angle and corner balance finite differencing in space.

Since RTE is essentially a hyperbolic type system, it is natural to solve the problem by the discrete ordinate method combined with the discontinuous Galerkin (DG) discretization in space. We call the resulting methods discrete-ordinate discontinuous Galerkin (discrete-ordinate DG) methods. Here we briefly review some advances of DG methods related to our problem under consideration. In 1973, Reed and Hill introduced in [52] the first DG method for linear hyperbolic problems (the neutron transport equation without scattering term), and an error analysis of the method was established thoroughly in [39]. A comprehensive introduction of DG methods was given in [19] for computational fluid dynamics. A general framework was presented in [18] for constructing DG methods based on discrete stability identities, from which one can derive many efficient DG methods for a large number of partial differential equations (systems) in a unified way. The stabilization mechanism frequently used in DG methods was revealed in [12], and the theory is also applied to discuss DG methods for linear hyperbolic problems. In addition, a complete study about the latter problem was given in [13] and some optimal error estimates in certain discrete norm were obtained there too. In [42], a DG method was proposed for solving the 1–D spherical neutron transport equation without scattering term. DG

methods have also been applied successfully to solve elliptic problems (cf. [5]) and many other mathematical physical problems (cf. [20]).

Under certain circumstances, it is appropriate to approximate the RTE with an equation that is easier to solve numerically. Under the conditions present in many biological tissues one valid approximation is a generalized Fokker-Planck Equation (GFPE). In Chapter 3 we study various properties of one GFPE such as existence of a unique solution and positivity of the solution. The GFPE can be solved naturally with an iteration procedure, and we show convergence of the iteration procedure in Section 3.3. Although the convergence is shown only at the continuous level, it holds also in applying the iteration method to solve discretized systems of the GFPE. In Section 3.4, we describe numerical discretization schemes of the GFPE. The rest of the thesis is structured as follows. In Section 1.2 we introduce Sobolev spaces and review prerequisite material. In Section 1.3 we introduce notation common to the entire thesis, and in Section 1.4 we introduce the radiative transfer equation and function spaces related to its solution. Chapter 2 is devoted to a theoretical and numerical study of a discrete-ordinate discontinuous Galerkin method applied to the radiative transfer equation. Chapter 4 lists potential future projects.

1.2 Sobolev spaces

In the study of partial differential equations and their numerical approximations Sobolev spaces play a critical role. This section provides basic definitions and results concerning these spaces that will be used throughout the thesis.

Let X be a bounded domain in $\mathbb{R}^d$, and denote its boundary by ∂X. We will write a typical point in $\mathbb{R}^d$ as $\boldsymbol{x}$.

A function $f : A \to B$, is Lipschitz continuous if for all $x, y \in A$ f satisfies

$$||f(x) - f(y)||_B \leq C||x - y||_A.$$

Here, A and B are arbitrary normed spaces, and $||\cdot||_A$ and $||\cdot||_B$ are their respective norms. A domain X is said to be a Lipschitz domain if at each point $\boldsymbol{x} \in \partial X$ there exists $r > 0$ and a Lipschitz function $\gamma : \mathbb{R}^{d-1} \to \mathbb{R}$ such that, upon rotating and relabeling the coordinate axes if necessary,

$$X \cap B(\boldsymbol{x}, r) = \{y | \gamma(y_1, y_2, \ldots, y_{d-1}) < y_d\} \cap B(\boldsymbol{x}, r).$$

In other words, near $\boldsymbol{x}$, ∂X is the graph of a Lipschitz function. For this thesis, $d = 3$.

For $p \in [1, \infty)$ we define $L^p(X)$ as the space of all measurable functions v such that

$$||v||_{L^p(X)} = \left\{ \int_X |v|^p dx \right\}^{1/p} < \infty.$$

For any $p \in [1, \infty)$, $L^p(X)$ is a Banach space, and $L^2(X)$ is a Hilbert space with inner product

$$(v, w)_{L^2(X)} = \int_X vw dx.$$

We denote by $L^\infty(X)$ the space of all essentially bounded measurable functions and use the norm

$$||v||_{L^\infty(X)} = \text{ess sup}_{\boldsymbol{x} \in X} |v(\boldsymbol{x})|.$$

We denote the space of locally p-integrable functions as $L^p_{loc}(X)$. A function v

is locally p-integrable if $v \in L^p(X')$ for any proper subset X' of X. In the case $p = 1$ we say that v is locally integrable.

Let α be an n-tuple of non-negative integers, i.e.

$$\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$$

such that $\alpha_i \geq 0$ for $1 \leq i \leq n$. We denote by $|\alpha|$ the norm of α, $|\alpha| = \sum_{i=1}^{n} \alpha_i$. If v is an m-times differentiable function then for any α satisfying $|\alpha| \leq m$ we write

$$D^\alpha v = \frac{\partial^{|\alpha|} v}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \ldots \partial x_n^{\alpha_n}}.$$

We refer to this function as the α^{th} partial derivative of v.

Define $C^m(\overline{X})$ as the space of all functions which are continuous on $\overline{X}$ and such that all of their partial derivatives of order at most m are also continuous on $\overline{X}$. In the special case $m = 0$ we write $C(\overline{X})$. If we endow $C^m(\overline{X})$ with the norm

$$||v||_{C^m(\overline{X})} = \max_{|\alpha| \leq m} \max_{\boldsymbol{x} \in \overline{X}} |D^\alpha v(\boldsymbol{x})|$$

then $C^m(\overline{X})$ is a Banach space. We also define

$$C^\infty(\overline{X}) = \cap_{m=0}^{\infty} C^m(\overline{X}).$$

Finally, we define $C_0^\infty(\overline{X})$ as the space of all compactly supported infinitely differentiable functions.

Let v,w be locally integrable. Then w is called an α'^{th} weak derivative of v if

$$\int_X v D^\alpha \phi dx = (-1)^{|\alpha|} \int_X w \phi dx, \quad \forall \phi \in C_0^\infty(X).$$

Note that a weak derivative, if it exists, is defined only a.e. with respect to the usual Lebesgue measure.

Using this definition of weak derivatives, we may define Sobolev spaces. For a non-negative integer m and $p \in [1, \infty)$, the Sobolev space $W^{m,p}(X)$ consists of all the functions $v \in L^1_{loc}(X)$ such that for each α satisfying $|\alpha| \leq m$ the weak derivative $D^\alpha v$ exists and belongs to the space $L^p(X)$. It is well known that $W^{m,p}(X)$ is a Banach space under the norm

$$||v||_{W^{m,p}(X)} = \begin{cases} \left\{ \sum_{|\alpha| \leq m} ||D^\alpha v||^p_{L^p(X)} \right\}^{1/p}, & p \in [1, \infty), \\ \max_{|\alpha| \leq m} ||D^\alpha v||_{L^\infty(X)}, & p = \infty. \end{cases} \tag{1.1}$$

In the case $p = 2$ we write $H^m(X) \equiv W^{m,2}(X)$. In this case, $H^m(X)$ is a Hilbert space with inner product

$$(v, w)_{H^m(X)} = \sum_{|\alpha| \leq m} \int_X D^\alpha v D^\alpha w dx$$

Though functions in a Sobolev space are defined only a.e. in X, it is possible to define their boundary value through a trace operator. The following theorem can be found in the literature, e.g. ([24]).

Theorem 1.2.1. *For $1 \leq p < \infty$, there exists a continuous linear operator $\gamma : W^{1,p}(X) \to L^p(\partial X)$ such that*

1. *$\gamma v = v|_{\partial X}$ if $v \in W^{1,p}(X) \cap C(\overline{X})$*

2. *For some constant $C > 0$, $||\gamma v||_{L^p(X)} \leq C||v||_{W^{1,p}(X)}$*

The operator γ is called the trace operator and γv is referred to as the generalized boundary value of v. To ease notation we use v for both $v \in W^{1,p}(X)$ and its trace, $\gamma v \in L^p(\partial X)$. The trace operator is neither an injection nor a surjection.

1.3 Notation

In this section we introduce some notation frequently used later on. For any two real quantities a and b, the symbol $a \lesssim b$ abbreviates $a \leq C\, b$, with C a positive constant independent of the finite element mesh size, which may take different values at different appearances.

We will use functions defined on the unit sphere in $\mathbb{R}^3$. Recall that the standard spherical coordinates are

$$x(\psi, \theta) = r \cos\psi \, \sin\theta, \quad y(\psi, \theta) = r \sin\psi \, \sin\theta, \quad z(\psi, \theta) = r \cos\theta$$

for $0 \leq \psi \leq 2\pi$, $0 \leq \theta \leq \pi$. Here, ψ represents the angle of rotation about the z axis, θ is the azimuth angle, and r is the distance from (x, y, z) to the origin.

We now briefly introduce an important set of functions on the sphere, the spherical harmonics. Spherical harmonics have been the subject of significant study, and we refer the reader to [43, 26] for more detail. Let $p(\boldsymbol{x})$ be a polynomial defined on $\mathbb{R}^3$ that is harmonic and also homogeneous of degree n, i.e.

$$\Delta p(\boldsymbol{x}) = 0,$$

and

$$p(\lambda \boldsymbol{x}) = \lambda^n p(\boldsymbol{x}).$$

The restriction of such a function to the unit sphere is called a spherical harmonic of degree n. The space of spherical harmonics of degree n, denoted as $Harm_n$, has dimension $2n+1$ and we denote any basis of $Harm_n$ as $\{Y_{n,m}\}$ for $1 \leq m \leq 2m+1$. The orthonormal basis is typically chosen as

$$Y_{n,1}(\phi,\theta) = c_n L_n(\cos\theta),$$

$$Y_{n,2m}(\phi,\theta) = c_{n,m} L_{n,m}(\cos\theta)\cos(m\phi),$$

$$Y_{n,2m+1}(\phi,\theta) = c_{n,m} L_{n,m}(\cos\theta)\sin(m\phi),$$

where $1 \leq m \leq n$. To complete this definition, define the Legendre polynomial of degree n as

$$L_n(t) = \frac{1}{2^n n!}\frac{d^n}{dt^n}\left[(t^2-1)^n\right],$$

and define the associated Legendre polynomials as

$$L_{n,m}(t) = (-1)^m(1-t^2)^{m/2}\frac{d^m}{dt^m}P_n(t).$$

The constants above are

$$c_n = \sqrt{\frac{2n+1}{4\pi}}$$
$$c_{n,m} = \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}}.$$

It can be shown that spherical harmonics of different orders are orthogonal in the sense of $L^2(\Omega)$ inner product. More precisely,

$$\int_\Omega Y_m Y_n d\sigma(\boldsymbol{\omega}) = 0$$

whenever $Y_m \in Harm_m$, $Y_n \in Harm_n$ and $m \neq n$. In fact, the set $\{Y_{n,m} | \, 0 \leq n \leq \infty, -n \leq m \leq n\}$ is an orthogonal basis for $L^2(\Omega)$, i.e.

$$f = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} (f, Y_{n,m})_{L^2(\Omega)} Y_{n,m}$$

for all $f \in L^2(\Omega)$.

Later, we will need to introduce Sobolev spaces on S^2. One possible way to define these spaces is to use local patch systems and proceed as in the previous section, only defining the Sobolev spaces in terms of the patches. An equivalent option (cf. [26]) is to define $H^s(\Omega)$ as the set of all functions f such that

$$||f||_{H^s(\Omega)} = \left[\sum_{n=0}^{\infty} \sum_{m=-n}^{n} (2n+1)^{2r} |(f, Y_{n,m})|^2 \right]^{1/2} < \infty.$$

The latter is the approach taken in this thesis.

We may express the Laplace operator Δ in spherical coordinates as

$$\Delta u = \frac{\partial^2 u}{\partial r^2} + \frac{2}{r} \frac{\partial u}{\partial r} + \frac{1}{r^2} \Delta^* u,$$

where Δ^* is defined as

$$\Delta^* u = \frac{1}{\sin^2 \theta} \frac{\partial^2 u}{\partial \psi^2} + \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \, \frac{\partial u}{\partial \theta} \right).$$

The operator Δ^* is called the spherical Laplacian or the Laplace-Beltrami operator. It is known that the spherical harmonics of degree n are eigenfunctions of the Laplace-Beltrami operator corresponding to the eigenvalues $-n(n+1)$. Similar to above, we may write the gradient operator as

$$\boldsymbol{\nabla} u = \left(\frac{\partial u}{\partial r}, \frac{1}{r \sin \theta} \frac{\partial u}{\partial \psi}, \frac{1}{r} \frac{\partial u}{\partial \theta} \right)^T.$$

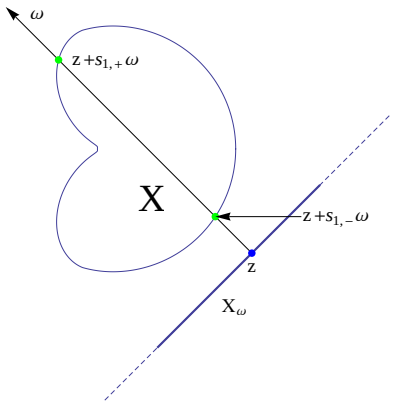

Figure 1.1: An illustration of $X_{\boldsymbol{\omega}}$

We define the spherical gradient, or surface gradient, $\boldsymbol{\nabla}^*$, as

$$\boldsymbol{\nabla}^* u = \left(\frac{1}{\sin\theta} \frac{\partial u}{\partial \psi}, \frac{\partial u}{\partial \theta} \right)^T.$$

1.4 The radiative transfer equation

Throughout this thesis, let X be a domain of the spatial variable, $X \subset \mathbb{R}^3$. Denote the boundary of X as ∂X and assume ∂X is Lipschitz. Note that this assumption implies that the unit outward normal $\boldsymbol{n}(\boldsymbol{x})$ exists a.e. on ∂X (cf. [8]). We will write Ω for S^2, the boundary of the unit ball in $\mathbb{R}^3$.

For each direction $\boldsymbol{\omega} \in \Omega$, we will define the inflow boundary of X with respect to $\boldsymbol{\omega}$. Define $X_{\boldsymbol{\omega}}$ as the orthogonal projection of X onto the plane perpendicular to $\boldsymbol{\omega}$ containing the origin. For any $\boldsymbol{z} \in X_{\boldsymbol{\omega}}$ define $X_{\boldsymbol{\omega},\boldsymbol{z}}$ as the intersection of X with the line $\{\boldsymbol{z} + s\boldsymbol{\omega} | s \in \mathbb{R}\}$. A two dimensional illustration is found in Figure 1.1. We assume that the domain X is such that for any $(\boldsymbol{\omega}, \boldsymbol{z})$ with $\boldsymbol{z} \in X_{\boldsymbol{\omega}}$, $X_{\boldsymbol{\omega},\boldsymbol{z}}$ is the union

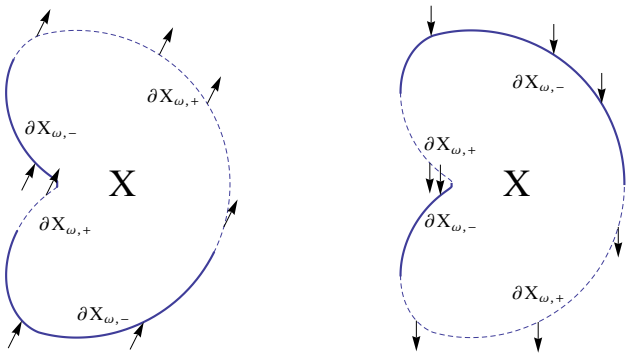

Figure 1.2: A two dimensional example of $\partial X_{\boldsymbol{\omega},\pm}$. The domain X is the interior of the region, and the direction $\boldsymbol{\omega}$ is represented by the arrows.

of a finite number of line segments:

$$X_{\boldsymbol{\omega},\boldsymbol{z}} = \cup_{i=1}^{N(\boldsymbol{\omega},\boldsymbol{z})} \{\boldsymbol{z} + s\boldsymbol{\omega} |\, s \in (s_{i,-}, s_{i,+})\}.$$

Here $s_{i,\pm} = s_{i,\pm}(\boldsymbol{\omega}, \boldsymbol{z})$ depend on $\boldsymbol{\omega}$ and $\boldsymbol{z}$, and $\boldsymbol{x}_{i,\pm} = \boldsymbol{z} + s_{i,\pm}\boldsymbol{\omega}$ are the intersection points of the line $\{\boldsymbol{z} + s\boldsymbol{\omega};\ s \in \mathbb{R}\}$ with ∂X. We further assume $\sup_{\boldsymbol{\omega},\boldsymbol{z}} N(\boldsymbol{\omega}, \boldsymbol{z}) < \infty$, known as a generalized convexity condition. With this notation in mind we introduce the following subsets of ∂X:

$$\partial X_{\boldsymbol{\omega},-} = \{\boldsymbol{z} + s_{i,-}\boldsymbol{\omega} |\, 1 \le i \le N(\boldsymbol{\omega}, \boldsymbol{z}),\ \boldsymbol{z} \in X_{\boldsymbol{\omega}}\},$$

$$\partial X_{\boldsymbol{\omega},+} = \{\boldsymbol{z} + s_{i,+}\boldsymbol{\omega} |\, 1 \le i \le N(\boldsymbol{\omega}, \boldsymbol{z}),\ \boldsymbol{z} \in X_{\boldsymbol{\omega}}\}.$$

The set $\partial X_{\boldsymbol{\omega},-}$ may be thought of as the collection of all the points $\boldsymbol{x}$ in the boundary of X such that $\boldsymbol{\omega}$ is pointing towards the interior of X from the point $\boldsymbol{x}$. Similarly, $X_{\boldsymbol{\omega},+}$ is the set of all points such that $\boldsymbol{\omega}$ is pointing away from X at the point $\boldsymbol{x}$. A two dimensional example may be seen in Figure 1.2. It can be shown that if $\boldsymbol{n}(\boldsymbol{x}_{i,-})$ exists with $\boldsymbol{x}_{i,-} = \boldsymbol{z} + s_{i,-}\boldsymbol{\omega}$, then $\boldsymbol{n}(\boldsymbol{x}_{i,-})\cdot\boldsymbol{\omega} \le 0$; if $\boldsymbol{x} \in \partial X$ and $\boldsymbol{n}(\boldsymbol{x})\cdot\boldsymbol{\omega} < 0$, then

$\boldsymbol{x} \in \partial X_{\boldsymbol{\omega},-}$. Likewise, if $\boldsymbol{n}(\boldsymbol{x}_{i,+})$ exists with $\boldsymbol{x}_{i,+} = \boldsymbol{z} + s_{i,+}\boldsymbol{\omega}$, then $\boldsymbol{n}(\boldsymbol{x}_{i,+})\cdot\boldsymbol{\omega} \geq 0$; if $\boldsymbol{x} \in \partial X$ and $\boldsymbol{n}(\boldsymbol{x})\cdot\boldsymbol{\omega} > 0$, then $\boldsymbol{x} \in \partial X_{\boldsymbol{\omega},+}$. Now, define the incoming boundary

$$\Gamma_- = \{(\boldsymbol{x},\boldsymbol{\omega})|\, \boldsymbol{x} \in \partial X_{\boldsymbol{\omega},-},\ \boldsymbol{\omega} \in \Omega\}$$

and the outgoing boundary

$$\Gamma_+ = \{(\boldsymbol{x},\boldsymbol{\omega})|\, \boldsymbol{x} \in \partial X_{\boldsymbol{\omega},+},\ \boldsymbol{\omega} \in \Omega\}.$$

Both are subsets of $\Gamma = \partial X \times \Omega$.

We will often integrate with respect to surface area of a domain. By $d\sigma(\boldsymbol{\omega})$ we will mean integration with respect to surface area on Ω and variable of integration $\boldsymbol{\omega}$. If we introduce the standard spherical coordinate system on S^2

$$\boldsymbol{\omega} = (\sin\theta\cos\psi,\ \sin\theta\sin\psi,\ \cos\theta)^T, \quad 0 \leq \theta \leq \pi,\ 0 \leq \psi \leq 2\pi, \tag{1.2}$$

then we may write

$$\int_\Omega f(\boldsymbol{\omega})d\sigma(\boldsymbol{\omega}) = \int_0^{2\pi}\int_0^{\pi} f(\psi,\theta)\sin(\theta)d\theta d\psi.$$

Define an integral operator S, applied to the function u as:

$$(Su)(\boldsymbol{x},\boldsymbol{\omega}) = \int_\Omega g(\boldsymbol{x},\boldsymbol{\omega}\cdot\hat{\boldsymbol{\omega}})\, u(\boldsymbol{x},\hat{\boldsymbol{\omega}})\, d\sigma(\hat{\boldsymbol{\omega}}). \tag{1.3}$$

We will require that g be a nonnegative normalized function:

$$\int_\Omega g(\boldsymbol{x},\boldsymbol{\omega}\cdot\hat{\boldsymbol{\omega}})\, d\sigma(\hat{\boldsymbol{\omega}}) = 1 \quad \forall\, \boldsymbol{x} \in X,\ \boldsymbol{\omega} \in \Omega. \tag{1.4}$$

g will be referred to as the phase function. In the applications of our interest the function g will be independent of $\boldsymbol{x}$. One well-known example is the Henyey-Greenstein

phase function (cf. [33])

$$g(t) = \frac{1-\eta^2}{4\pi(1+\eta^2-2\eta t)^{3/2}}, \quad t \in [-1,1], \tag{1.5}$$

where the parameter $\eta \in (-1,1)$. This particular phase function is popular in modeling scattering of light in a biological tissue.

We introduce a space of functions analogous to the Sobelev spaces defined previously. Let $C^{(1,0)}(X \times \Omega)$ be the space of functions that are continuously differentiable in the spatial variable $\boldsymbol{x}$ for every fixed $\boldsymbol{\omega}$ and such that the directional derivative in the direction $\boldsymbol{\omega}$ is continuous with respect to all its arguments. Let ϕ be an arbitrary function in $C^{(1,0)}(X \times \Omega)$. For a function $v \in L^2(X \times \Omega)$ we define the weak directional derivative in direction $\boldsymbol{\omega}$, w, such that for all $\psi \in C^{(1,0)}(X \times \Omega)$ with compact support in X:

$$\int_{X\times\Omega} w\phi dx d\sigma(\boldsymbol{\omega}) = -\int_{X\times\Omega} v\boldsymbol{\omega}\cdot\boldsymbol{\nabla}\phi dx d\sigma(\boldsymbol{\omega}). \tag{1.6}$$

If such a w exists, we will denote it as $\boldsymbol{\omega}\cdot\boldsymbol{\nabla}v$. We define the space $H_2^1(X\times\Omega)$ as:

$$H_2^1(X\times\Omega) = \{v | v \in L^2(X\times\Omega) \text{ and } \boldsymbol{\omega}\cdot\boldsymbol{\nabla}v \in L^2(X\times\Omega)\}. \tag{1.7}$$

The norm on $H_2^1(X\times\Omega)$ is:

$$||v||_{H_2^1(X\times\Omega)} = ||v||_{L^2(X\times\Omega)} + ||\boldsymbol{\omega}\cdot\boldsymbol{\nabla}v||_{L^2(X\times\Omega)}.$$

One useful property of the space $H_2^1(X\times\Omega)$ (cf. [3]) is:

Lemma 1.4.1. *$H_2^1(X\times\Omega)$ is complete, and $C^{(1,0)}(X\times\Omega)$ is dense in it.*

In Chapter 3, we also need function spaces $L^2(\Gamma_\pm)$ on $\Gamma_\pm$. They are defined to be spaces of functions v on $\Gamma_\pm$ such that

$$\|v\|_{L^2(\Gamma_\pm)} = \left[\int_\Omega \int_{X_{\boldsymbol{\omega}}} \sum_{i=1}^{N(\boldsymbol{\omega},\boldsymbol{z})} |v(\boldsymbol{z}+s_{i,\pm}\boldsymbol{\omega},\boldsymbol{\omega})|^2 dz d\sigma(\boldsymbol{\omega})\right]^{1/2} < \infty.$$

The inner products in $L^2(\Gamma_\pm)$ are

$$(u,v)_{L^2(\Gamma_\pm)} = \int_\Omega \int_{X_{\boldsymbol{\omega}}} \sum_{i=1}^{N(\boldsymbol{\omega},\boldsymbol{z})} u(\boldsymbol{z}+s_{i,\pm}\boldsymbol{\omega},\boldsymbol{\omega})\, v(\boldsymbol{z}+s_{i,\pm}\boldsymbol{\omega},\boldsymbol{\omega})\, dz d\sigma(\boldsymbol{\omega}).$$

We have the following statement for the trace of an $H^{1,2}(X\times\Omega)$ function ([3, Lemma 2.2]). If $v \in H^{1,2}(X\times\Omega)$ and $v|_{\Gamma_-} \in L^2(\Gamma_-)$, then $v|_{\Gamma_+} \in L^2(\Gamma_+)$ and for some constant c depending only on X,

$$\|v\|_{L^2(\Gamma_+)} \le c\left[\|v\|_{H^{1,2}(X\times\Omega)} + \|v\|_{L^2(\Gamma_-)}\right]. \tag{1.8}$$

The statement remains valid by switching Γ_+ and Γ_-.

The radiative transfer equation considered in Chapter 2 is: (cf. [3, 40])

$$\forall(\boldsymbol{x},\boldsymbol{\omega}) \in X\times\Omega,$$

$$\boldsymbol{\omega}\cdot\boldsymbol{\nabla} u(\boldsymbol{x},\boldsymbol{\omega}) + \mu_t(\boldsymbol{x})\, u(\boldsymbol{x},\boldsymbol{\omega}) = \mu_s(\boldsymbol{x})\,(Su)(\boldsymbol{x},\boldsymbol{\omega}) + f(\boldsymbol{x},\boldsymbol{\omega}), \tag{1.9}$$

$$u(\boldsymbol{x},\boldsymbol{\omega}) = 0, (\boldsymbol{x},\boldsymbol{\omega}) \in \Gamma_-. \tag{1.10}$$

In Chapter 3 we consider the general boundary condition

$$u(\boldsymbol{x},\boldsymbol{\omega}) = u_{\text{in}}(\boldsymbol{x},\boldsymbol{\omega}) \text{ on } \Gamma_- \tag{1.11}$$

with a given function $u_{\text{in}} \in L^2(\Gamma_-)$.

The unknown function is $u(\boldsymbol{x}, \boldsymbol{\omega})$, and μ_t, μ_s, and f are given functions. If we take μ_t and μ_s such that

$$\begin{aligned} &\mu_t, \mu_s \in L^\infty(X), \quad \mu_s \geq 0 \text{ a.e. in } X, \\ &\mu_t - \mu_s \geq c_0 \text{ in } X \text{ for some constant } c_0 > 0, \end{aligned} \tag{1.12}$$

and

$$f \in L^2(X \times \Omega) \text{ and is a continuous function with respect to } \boldsymbol{\omega} \in \Omega. \tag{1.13}$$

Then both the problem (1.9)–(1.10) and (1.9)–(1.11) have a unique solution $u \in H_2^1(X \times \Omega)$ [3].

CHAPTER 2
DISCRETE-ORDINATE DISCONTINUOUS GALERKIN METHODS FOR THE RADIATIVE TRANSFER EQUATION

2.1 Introduction

The discrete-ordinate discontinuous Galerkin (DG) methods for the radiative transfer equation (RTE) considered in this thesis are developed in two steps. In the first step, the integration term in the RTE is replaced by a quadrature rule, and we require that the resulting equations hold only in the directions corresponding to the numerical integration nodes. This results in a linear first-order hyperbolic system of PDE. To approximate the solution to this system of PDE we follow [12, 13, 19] and use DG methods. As in [12, 13], we also consider the DG methods for which a stabilization term is added for penalizing the jump of the approximate solution across interior faces of the triangulation, in order to get numerical solutions with more physical meaning. To derive error estimates we first consider the error between the numerical solution and the exact solution to the discrete-ordinate system. We give a lemma for error estimates of general DG methods, which is similar to the second Strang lemma in error analysis of nonconforming element methods (cf. [17]). We then prove stability estimates for the methods. Combining these estimates and estimates involving the L^2-orthogonal projection operator gives an error estimate between the numerical solution and the solution to the discrete ordinate system. We estimate the error between the solution of the discrete-ordinate system and the solution to the RTE using an error formula for numerical quadrature on the unit sphere (cf. [34]). These

results together give error estimates between the exact solution and our numerical solutions in the L^2-norm. Results from several numerical examples are provided to illustrate computational performance of the methods proposed here.

2.2 Discrete-ordinate discontinuous Galerkin methods

In the literature, discrete-ordinate methods typically refer to methods in which the integration term in the RTE is replaced by a quadrature rule, and the resulting equation is required to hold only at the quadrature nodes rather than all directions $\boldsymbol{\omega}$. We follow the same procedure in this thesis.

We write the numerical quadrature to be used in the form:

$$\int_{\Omega} F(\boldsymbol{\omega})\, d\sigma(\boldsymbol{\omega}) \approx \sum_{l=1}^{L} w_l F(\boldsymbol{\omega}_l), \quad w_l > 0,\ \boldsymbol{\omega}_l \in \Omega,\ 1 \le l \le L, \tag{2.1}$$

where F is a continuous function over the unit sphere Ω. Any quadrature rule may be used, for example the product numerical integration formulas (cf. [7, 54]),

$$\int_{\Omega} F(\boldsymbol{\omega})\, d\sigma(\boldsymbol{\omega}) = \int_0^{2\pi} \int_0^{\pi} \bar{F}(\theta, \psi) \sin\theta \, d\theta \, d\psi \approx \frac{\pi}{m} \sum_{j=1}^{2m} \sum_{i=1}^{m} w_i \bar{F}(\theta_i, \psi_j), \tag{2.2}$$

where ψ, and θ refer to the usual spherical coordinates (1.3), and $\bar{F}$ stands for the representation of F in that system. Here, $\{\theta_i\}$ are chosen so that $\{\cos(\theta_i)\}$ and $\{w_i\}$ are the Gauss-Legendre nodes and weights on $[-1, 1]$. The points $\{\psi_j\}$ are evenly spaced on $[0, 2\pi]$ with a spacing π/m; typically,

$$\psi_j = j\pi/m \ \text{ or } \ (j - 1/2)\pi/m.$$

The above integration method integrates exactly any polynomial $F(\boldsymbol{x})$ of a total degree no more than $2m - 1$.

A second typical choice of nodes and weights come from symmetry considerations. For example, the S_N quadrature rules have nodes which are invariant under rotation about all three axes. Such a requirement is common in situations in which there is no natural preferred orientation of the domain X in $\mathbb{R}^3$. We refer to [14, 40] for details of these methods.

With a choice of quadrature rule in mind the integral operator S (1.3) can be approximated by a discretized operator S_d given by

$$S_d u(\boldsymbol{x}, \boldsymbol{\omega}) = \sum_{l=1}^{L} w_l g(\boldsymbol{x}, \boldsymbol{\omega}\cdot\boldsymbol{\omega}_l)\, u(\boldsymbol{x}, \boldsymbol{\omega}_l). \tag{2.3}$$

Recalling condition (1.4) we define

$$m(\boldsymbol{x}) = \max_{1\le i\le L} \sum_{l=1}^{L} w_l g(\boldsymbol{x}, \boldsymbol{\omega}_l\cdot\boldsymbol{\omega}_i). \tag{2.4}$$

If the quadrature rule is exact for g then $m(\boldsymbol{x}) \equiv 1$. Thus, if the quadrature is high order we expect $m(\boldsymbol{x}) \approx 1$ in X. More precisely, we cite [34] to state the following theorem.

Theorem 2.2.1. *Let g be a function defined on S^2, and let $\{w_i\}_{i=1}^N$, $\{\boldsymbol{\omega}_i\}_{i=1}^N$ be a set of nodes and weights such that $w_i > 0$ for $1 \le i \le N$. If*

$$\int_\Omega p(\boldsymbol{\omega}) d\sigma(\boldsymbol{\omega}) = \sum_{i=1}^{N} w_i p(\boldsymbol{\omega}_i)$$

for all polynomials p of degree no more than n, and if $g \in H^s(\Omega)$, then

$$\left| \sum_{i=1}^{N} w_i g(\boldsymbol{\omega}_i) - \int_\Omega g(\boldsymbol{\omega}) d\sigma(\boldsymbol{\omega}) \right| \le c_s n^{-s} ||g||_{H^s(\Omega)}.$$

where c_s is a positive constant depending only on s. The result is valid for $s > 1$.

In the case of the Henyey-Greenstein phase function g this implies

$$\left|1 - \sum_{l=1}^{L} w_l g(\boldsymbol{\omega}_i \cdot \boldsymbol{\omega}_l)\right| \leq c_s n^{-s} ||g(\boldsymbol{\omega}_i \cdot \boldsymbol{\omega})||_{s,\Omega} \quad \forall\, s > 1.$$

With this in mind we see that the assumption

$$\mu_t - m\,\mu_s \geq c_0' \text{ in } X \text{ for some constant } c_0' > 0 \tag{2.5}$$

is not too restrictive for the Henyey-Greenstein phase function. Numerical results illustrating this point are included in Section 2.4.

Replacing Su in (1.9) with the quadrature rule (2.3), and requiring that the resulting equation hold only along the node points $\{\boldsymbol{\omega}_l\}$ results in the system of equations:

$$\boldsymbol{\omega}_l \cdot \nabla u^l + \mu_t u^l = \mu_s \sum_{i=1}^{L} w_i g(\cdot, \boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i)\, u^i + f_l, \quad 1 \leq l \leq L. \tag{2.6}$$

Here we use the notation $f_l = f(\boldsymbol{x}, \boldsymbol{\omega}_l)$. Then $u^l(\boldsymbol{x})$ is an approximation of $u(\boldsymbol{x}, \boldsymbol{\omega}_l)$ satisfying the boundary condition $u^l = 0$ on ∂X_-^l. For the rest of the thesis we use the simplified notation $\partial X_\pm^l = \partial X_{\boldsymbol{\omega}_l, \pm}$.

We follow [12, 13, 19] to discretize (2.6). For ease of presentation, we assume X is a polyhedron. Let $\{\mathcal{T}_h\}_{h>0}$ be a regular family of triangulations of X into tetrahedral elements (cf. [8, 11, 17]); $h_K = \text{diam}(K)$ and $h = \max_{K \in \mathcal{T}_h} h_K$. Denote by $\boldsymbol{n}_K$ the unit outward normal to ∂K for $K \in \mathcal{T}_h$. Let $\mathcal{E}_h^i$ be the set of all interior faces of $\mathcal{T}_h$. Moreover, with each $e \in \mathcal{E}_h^i$ we associate a unit normal direction $\boldsymbol{n}_e$. Associated with the triangulation $\mathcal{T}_h$, define a finite element space by

$$V_h = \{v \in L^2(\Omega) |\, v|_K \in P_k(K)\ \forall\, K \in \mathcal{T}_h\},$$

where k is a nonnegative integer and $P_k(K)$ denotes the set of all polynomials on K of a total degree no more than k. Define $\boldsymbol{V}_h = (V_h)^L$, and write a generic element in $\boldsymbol{V}_h$ as $\boldsymbol{v}_h = \{v_h^l\}_{l=1}^L$ or simply $\boldsymbol{v}_h = \{v_h^l\}$.

We shall call the numerical solution $\boldsymbol{u}_h = \{u_h^l\}$ and require that $\boldsymbol{u}_h \in \boldsymbol{V}_h$. To define $\boldsymbol{u}_h$, we begin by finding an appropriate weak formulation of (2.6). We multiply (2.6) by a function $v_K \in P_k(K)$ and integrate over K to obtain

$$\begin{aligned}
&\forall K \in \mathcal{T}_h, 1 \le l \le L, \\
&\int_K \boldsymbol{\omega}_l \cdot \boldsymbol{\nabla} u^l \, v_K d\boldsymbol{x} + \int_K \mu_t u^l \, v_K d\boldsymbol{x} \\
&\qquad = \int_K \mu_s \sum_{i=1}^{L} w_i g(\cdot, \boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i) \, u^i v_K d\boldsymbol{x} + \int_K f_l v_K d\boldsymbol{x}.
\end{aligned}$$

Integration by parts gives

$$\begin{aligned}
&\forall K \in \mathcal{T}_h, 1 \le l \le L, \\
&\int_{\partial K} \boldsymbol{\omega}_l \cdot \boldsymbol{n}_K u^l \, v_K d\sigma(\boldsymbol{x}) - \int_K u^l \, \boldsymbol{\omega}_l \cdot \boldsymbol{\nabla} v_K \, d\boldsymbol{x} + \int_K \mu_t u^l v_K d\boldsymbol{x} \\
&\qquad = \int_K \mu_s \sum_{i=1}^{L} w_i g(\cdot, \boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i) \, u^i v_K d\boldsymbol{x} + \int_K f_l v_K d\boldsymbol{x}.
\end{aligned} \tag{2.7}$$

We now define $\boldsymbol{u}_h$ by requiring

$$\begin{aligned}
&\forall K \in \mathcal{T}_h, \forall v_K \in P_k(K), \ 1 \le l \le L, \\
&\int_{\partial K} \boldsymbol{\omega}_l \cdot \boldsymbol{n}_K \, \hat{u}_h^l \, v_K d\sigma(\boldsymbol{x}) - \int_K u_h^l \, \boldsymbol{\omega}_l \cdot \boldsymbol{\nabla} v_K \, d\boldsymbol{x} + \int_K \mu_t u_h^l \, v_K d\boldsymbol{x} \\
&\qquad = \int_K \mu_s \sum_{i=1}^{L} w_i g(\cdot, \boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i) \, u_h^i v_K d\boldsymbol{x} + \int_K f_l v_K d\boldsymbol{x}.
\end{aligned} \tag{2.8}$$

The quantity $\hat{u}_h^l$ is defined as

$$\hat{u}_h^l(\boldsymbol{x}) = \begin{cases} 0, & \text{if } (\boldsymbol{x}, \boldsymbol{\omega}_l) \in \Gamma_-, \\ \lim_{\varepsilon \to 0+} u_h^l(\boldsymbol{x} - \varepsilon \boldsymbol{\omega}_l), & \text{otherwise} \end{cases} \tag{2.9}$$

and is called the numerical trace. This choice of numerical trace reflects the natural flow of information along characteristic lines. Choice of numerical trace is an important topic when developing discontinuous methods. In this thesis we consider only the above "upwinding" numerical trace and refer the reader to [18] for more detail and alternative choices.

The global formulation of the discrete method (2.8) can be expressed as

$$\begin{aligned} &\forall \boldsymbol{v}_h = \{v_h^l\} \in \boldsymbol{V}_h, \\ &\sum_{l=1}^{L} w_l \sum_{K \in \mathcal{T}_h} \left\{ \int_{\partial K} \boldsymbol{\omega}_l \cdot \boldsymbol{n}_K \hat{u}_h^l \, v_{h,K}^l d\sigma(\boldsymbol{x}) - \int_K u_h^l \, \boldsymbol{\omega}_l \cdot \boldsymbol{\nabla} v_{h,K}^l \, d\boldsymbol{x} + \int_K \mu_t u_h^l \, v_{h,K}^l d\boldsymbol{x} \right\} \\ &\qquad - \sum_{l=1}^{L} w_l \int_X \mu_s \sum_{i=1}^{L} w_i g(\cdot, \boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i) \, u_h^i v_h^l d\boldsymbol{x} = \sum_{l=1}^{L} w_l \int_X f_l v_h^l d\boldsymbol{x} \end{aligned} \tag{2.10}$$

where $v_{h,K}^l = v_h^l|_K$ for all $v_h^l \in V_h$. Define the bilinear form $a^{(1)} : \boldsymbol{V}_h \times \boldsymbol{V}_h \to \mathbb{R}$ as

$$\begin{aligned} a_h^{(1)}(\boldsymbol{u}_h, \boldsymbol{v}_h) = &\sum_{l=1}^{L} w_l \sum_{K \in \mathcal{T}_h} \left\{ \int_{\partial K} \boldsymbol{\omega}_l \cdot \boldsymbol{n}_K \hat{\boldsymbol{u}}_h^l v_{h,K}^l d\sigma(\boldsymbol{x}) - \int_K u_h^l \boldsymbol{\omega}_l \cdot \boldsymbol{\nabla} v_{h,K}^l d\boldsymbol{x} \right. \\ &\left. + \int_K \mu_t u_h^l v_{h,K}^l d\boldsymbol{x} \right\} - \sum_{l=1}^{L} w_l \int_X \mu_s \sum_{i=1}^{L} w_i g(\cdot, \boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i) u_h^i v_h^l d\boldsymbol{x}, \end{aligned} \tag{2.11}$$

and define the linear operator $f : \boldsymbol{V}_h \to \mathbb{R}$ by

$$f(\boldsymbol{v}_h) = \sum_{l=1}^{L} w_l \int_X f_l v_h^l d\boldsymbol{x}. \tag{2.12}$$

We may then restate the discrete-ordinate DG problem (2.10) as:

Find $\boldsymbol{u}_h \in \boldsymbol{V}_h$ such that

$$a_h^{(1)}(\boldsymbol{u}_h, \boldsymbol{v}_h) = f(\boldsymbol{v}_h) \quad \forall\, \boldsymbol{v}_h \in \boldsymbol{V}_h. \tag{2.13}$$

When using discontinuous Galerkin methods, the resulting numerical solution may be discontinuous across element interfaces. This is undesirable if the solution is to have physical significance. We may follow [12, 13], when solving hyperbolic problems and add some stabilization terms in the DG scheme to penalize the jump of the solution across interior faces of the triangulation. To do this, we introduce some new notation. For an interior face $e \in \mathcal{E}_h^i$, let K^+ and K^- be two adjacent tetrahedrons sharing e, with the unit direction $\boldsymbol{n}_e$ pointing from K^- to K^+. For a scalar-valued function v, write $v^+ = v|_{K^+}$ and $v^- = v|_{K^-}$. Then define the jump of v on e by

$$[v] = v^+ - v^-.$$

Now, we can define the following discrete-ordinate DG method with stabilization:

Find $\boldsymbol{u}_h \in \boldsymbol{V}_h$ such that

$$a_h^{(2)}(\boldsymbol{u}_h, \boldsymbol{v}_h) = f(\boldsymbol{v}_h) \quad \forall\, \boldsymbol{v}_h \in \boldsymbol{V}_h, \tag{2.14}$$

where

$$a_h^{(2)}(\boldsymbol{u}_h, \boldsymbol{v}_h) = a_h^{(1)}(\boldsymbol{u}_h, \boldsymbol{v}_h) + c_p \sum_{l=1}^{L} w_l \sum_{e \in \mathcal{E}_h^i} \int_e [u_h^l]\, [v_h^l]\, d\sigma(\boldsymbol{x}), \tag{2.15}$$

with $a_h^{(1)}(\cdot, \cdot)$ given by (2.11) and $c_p > 0$ a penalty parameter which can be chosen as required.

2.3 Error analysis

We begin by demonstrating the stability and unique solvability of the systems of equations (2.13) and (2.14). Before we proceed, we introduce one lemma which we frequently need.

Lemma 2.3.1. *For all* $\boldsymbol{v} = \{v^l\}$, $\boldsymbol{w} = \{w^l\} \in (L^2(\Omega))^L$,

$$\sum_{l=1}^{L} w_l \int_X \mu_s \sum_{i=1}^{L} w_i g(\cdot, \boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i)\, v^i w^l dx \\ \leq \left[\sum_{l=1}^{L} w_l \int_X m\, \mu_s (v^l)^2 dx\right]^{1/2} \left[\sum_{l=1}^{L} w_l \int_X m\, \mu_s (w^l)^2 dx\right]^{1/2}.$$

Proof. Interchanging the order of summation,

$$\sum_{l=1}^{L} w_l \int_X \mu_s \sum_{i=1}^{L} w_i g(\cdot, \boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i)\, v^i w^l dx \\ = \sum_{i=1}^{L} w_i \sum_{l=1}^{L} w_l \int_X \mu_s g(\cdot, \boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i)\, v^i w^l dx.$$

Using the Cauchy-Schwarz inequality, we find

$$\sum_{l=1}^{L} w_l \int_X \mu_s \sum_{i=1}^{L} w_i g(\cdot, \boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i)\, v^i w^l dx \\ \leq \sum_{i=1}^{L} w_i \left[\sum_{l=1}^{L} w_l \int_X \mu_s g(\cdot, \boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i)\, (v^i)^2 dx\right]^{1/2} \\ \cdot \left[\sum_{l=1}^{L} w_l \int_X \mu_s g(\cdot, \boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i)\, (w^l)^2 dx\right]^{1/2}. \tag{2.16}$$

By the definition (2.4), it is clear that

$$\sum_{l=1}^{L} w_l \int_X \mu_s g(\cdot, \boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i)\, (v^i)^2 dx \leq \int_X m\, \mu_s (v^i)^2 dx.$$

Combining (2.16) and the Cauchy-Schwarz inequality we arrive at

$$\sum_{l=1}^{L} w_l \int_X \mu_s \sum_{i=1}^{L} w_i g(\cdot, \boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i)\, v^i w^l dx$$
$$\leq \sum_{i=1}^{L} w_i \left[\int_X m\, \mu_s (v^i)^2 dx \right]^{1/2} \left[\sum_{l=1}^{L} w_l \int_X \mu_s g(\cdot, \boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i)\, (w^l)^2 dx \right]^{1/2}.$$

Using Cauchy-Schwarz again yields

$$\sum_{l=1}^{L} w_l \int_X \mu_s \sum_{i=1}^{L} w_i g(\cdot, \boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i)\, v^i w^l dx$$
$$\leq \left[\sum_{i=1}^{L} w_i \int_X m\, \mu_s (v^i)^2 dx \right]^{1/2} \left[\sum_{l,i=1}^{L} w_i w_l \int_X \mu_s g(\cdot, \boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i)\, (w^l)^2 dx \right]^{1/2}$$
$$\leq \left[\sum_{l=1}^{L} w_l \int_X m\, \mu_s (v^l)^2 dx \right]^{1/2} \left[\sum_{l=1}^{L} w_l \int_X m\, \mu_s (w^l)^2 dx \right]^{1/2},$$

as desired. □

Now define norms $|||\cdot|||_h^{(1)}$ and $|||\cdot|||_h^{(2)}$ over $\boldsymbol{V}_h$ in order to demonstrate the stability of the schemes (2.13) and (2.14):

$$|||\boldsymbol{v}|||_h^{(1)} = \Bigg[\sum_{l=1}^{L} w_l \Bigg(||v^l||_{0,X}^2 + \sum_{e \in \mathcal{E}_h^i} \int_e |\boldsymbol{\omega}_l \cdot \boldsymbol{n}_e| |[v^l]|^2 d\sigma(\boldsymbol{x})$$
$$+ \int_{\partial X_+^l} \boldsymbol{\omega}_l \cdot \boldsymbol{n}\, |v^l|^2 d\sigma(\boldsymbol{x}) \Bigg) \Bigg]^{1/2}, \tag{2.17}$$
$$|||\boldsymbol{v}|||_h^{(2)} = \Bigg[\sum_{l=1}^{L} w_l \Bigg(||v^l||_{0,X}^2 + c_p \sum_{e \in \mathcal{E}_h^i} \int_e |[v^l]|^2 d\sigma(\boldsymbol{x})$$
$$+ \int_{\partial X_+^l} \boldsymbol{\omega}_l \cdot \boldsymbol{n}\, |v^l|^2 d\sigma(\boldsymbol{x}) \Bigg) \Bigg]^{1/2}. \tag{2.18}$$

With these definitions in hand we have the following stability estimates.

Lemma 2.3.2 (Stability)**.** *Under the assumption* (2.5)*,*

$$\left(|||\boldsymbol{v}_h|||_h^{(1)}\right)^2 \lesssim a_h^{(1)}(\boldsymbol{v}_h, \boldsymbol{v}_h) \quad \forall\, \boldsymbol{v}_h \in \boldsymbol{V}_h, \tag{2.19}$$

$$\left(|||\boldsymbol{v}_h|||_h^{(2)}\right)^2 \lesssim a_h^{(2)}(\boldsymbol{v}_h, \boldsymbol{v}_h) \quad \forall\, \boldsymbol{v}_h \in \boldsymbol{V}_h. \tag{2.20}$$

Proof. Recall definition (2.11). For any $\boldsymbol{v}_h = \{v_h^l\} \in \boldsymbol{V}_h$ write

$$a_h^{(1)}(\boldsymbol{v}_h, \boldsymbol{v}_h) = \mathrm{I}_1 + \mathrm{I}_2, \tag{2.21}$$

where

$$\mathrm{I}_1 = \sum_{l=1}^{L} w_l \sum_{K\in\mathcal{T}_h} \left\{ \int_{\partial K} \boldsymbol{\omega}_l\cdot\boldsymbol{n}_K\, \hat{v}_h^l\, v_{h,K}^l d\sigma(\boldsymbol{x}) - \int_K v_h^l\, \boldsymbol{\omega}_l\cdot\boldsymbol{\nabla} v_{h,K}^l\, dx \right\},$$
$$\mathrm{I}_2 = \sum_{l=1}^{L} w_l \int_X \mu_t (v_h^l)^2 dx - \sum_{l=1}^{L} w_l \int_X \mu_s \sum_{i=1}^{L} w_i g(\cdot, \boldsymbol{\omega}_l\cdot\boldsymbol{\omega}_i)\, v_h^i v_h^l dx.$$

Then

$$\mathrm{I}_2 = \sum_{l=1}^{L} w_l \int_X \mu_t (v_h^l)^2 dx - \sum_{l=1}^{L} w_l \int_X \mu_s \sum_{i=1}^{L} w_i g(\cdot, \boldsymbol{\omega}_l\cdot\boldsymbol{\omega}_i)\, v_h^i v_h^l dx,$$

and by Lemma 2.3.1

$$\begin{aligned}\mathrm{I}_2 &\geq \sum_{l=1}^{L} w_l \int_X \mu_t (v_h^l)^2 dx - \sum_{l=1}^{L} w_l \int_X m\mu_s (v_h^l)^2 dx \\ &= \sum_{l=1}^{L} w_l \int_X (\mu_t - m\mu_s)(v_h^l)^2 dx.\end{aligned}$$

Assumption (2.5) gives

$$\mathrm{I}_2 \geq c_0' \sum_{l=1}^{L} w_l \int_X (v_h^l)^2 dx. \tag{2.22}$$

To bound I_1 we first perform an integration by parts to see that

$$\begin{aligned}
&\int_{\partial K} \boldsymbol{\omega}_l \cdot \boldsymbol{n}_K \, \hat{v}_h^l \, v_{h,K}^l d\sigma(\boldsymbol{x}) - \int_K v_h^l \, \boldsymbol{\omega}_l \cdot \boldsymbol{\nabla} v_{h,K}^l \, dx \\
&\quad = \int_{\partial K} \boldsymbol{\omega}_l \cdot \boldsymbol{n}_K \hat{v}_h^l v_{h,K}^l \, d\sigma(\boldsymbol{x}) - \frac{1}{2} \int_K \boldsymbol{\omega}_l \cdot \boldsymbol{\nabla} (v_h^l)^2 dx \\
&\quad = \int_{\partial K} \boldsymbol{\omega}_l \cdot \boldsymbol{n}_K \hat{v}_h^l v_{h,K}^l \, d\sigma(\boldsymbol{x}) - \frac{1}{2} \int_{\partial K} \boldsymbol{\omega}_l \cdot \boldsymbol{n}_K (v_h^l)^2 d\sigma(\boldsymbol{x}) \\
&\quad = \frac{1}{2} \int_{\partial K} \boldsymbol{\omega}_l \cdot \boldsymbol{n}_K \left[(\hat{v}_h^l)^2 - (v_h^l - \hat{v}_h^l)^2 \right] d\sigma(\boldsymbol{x}).
\end{aligned}$$

Multiplying the above equality by w_l and summing over all $K \in \mathcal{T}_h$ and $1 \leq l \leq L$ we see that

$$\mathrm{I}_1 = \frac{1}{2} \sum_{l=1}^{L} w_l \sum_{K \in \mathcal{T}_h} \int_{\partial K} \boldsymbol{\omega}_l \cdot \boldsymbol{n}_K (\hat{v}_h^l)^2 d\sigma(\boldsymbol{x}) - \frac{1}{2} \sum_{l=1}^{L} w_l \sum_{K \in \mathcal{T}_h} \int_{\partial K} \boldsymbol{\omega}_l \cdot \boldsymbol{n}_K (v_h^l - \hat{v}_h^l)^2 d\sigma(\boldsymbol{x}).$$

By the definition of $\hat{v}_h^l$, (2.9), if $\boldsymbol{\omega}_l \cdot \boldsymbol{n}_K \geq 0$, then $\hat{v}_h^l(\boldsymbol{x}) = v_h^l(\boldsymbol{x})$. Thus,

$$\mathrm{I}_1 = \frac{1}{2} \sum_{l=1}^{L} w_l \sum_{K \in \mathcal{T}_h} \int_{\partial K} \boldsymbol{\omega}_l \cdot \boldsymbol{n}_K (\hat{v}_h^l)^2 d\sigma(\boldsymbol{x}) + \frac{1}{2} \sum_{l=1}^{L} w_l \sum_{K \in \mathcal{T}_h} \int_{\partial K} |\boldsymbol{\omega}_l \cdot \boldsymbol{n}_K| (v_h^l - \hat{v}_h^l)^2 d\sigma(\boldsymbol{x}).$$

In the above equality, each edge $e \in \mathcal{E}_h^i$ is summed over exactly twice, corresponding to the two mesh elements that share it. These two elements have opposite outward pointing normals, and their contributions will cancel each other in the sum. Further, for $e \in X_-^l$, $\hat{v}_h^l|_e = 0$. Thus,

$$\mathrm{I}_1 = \frac{1}{2} \sum_{l=1}^{L} w_l \int_{\partial X_+^l} \boldsymbol{\omega}_l \cdot \boldsymbol{n} \, (v_h^l)^2 d\sigma(\boldsymbol{x}) + \frac{1}{2} \sum_{l=1}^{L} w_l \sum_{K \in \mathcal{T}_h} \int_{\partial K} |\boldsymbol{\omega}_l \cdot \boldsymbol{n}_K| (v_h^l - \hat{v}_h^l)^2 d\sigma(\boldsymbol{x}).$$

Since $v_h^l - \hat{v}_h^l$ is nonzero on only one of the two elements that share the edge e, we find

$$\mathrm{I}_1 \geq \frac{1}{2} \sum_{l=1}^{L} w_l \int_{\partial X_+^l} \boldsymbol{\omega}_l \cdot \boldsymbol{n} \, (v_h^l)^2 d\sigma(\boldsymbol{x}) + \frac{1}{2} \sum_{l=1}^{L} w_l \sum_{e \in \mathcal{E}_h^i} \int_e |\boldsymbol{\omega}_l \cdot \boldsymbol{n}_e| |[v^l]|^2 d\sigma(\boldsymbol{x}). \tag{2.23}$$

Therefore, (2.19) follows directly from (2.21)–(2.23).

To prove (2.20) we use the result (2.19).

$$\begin{aligned}
&\sum_{l=1}^{L} w_l \left(||v^l||_{0,X}^2 + \int_{\partial X_+^l} \boldsymbol{\omega}_l \cdot \boldsymbol{n} \, |v^l|^2 d\sigma(\boldsymbol{x}) \right) \\
&\quad \le \sum_{l=1}^{L} w_l \left(||v^l||_{0,X}^2 + \sum_{e \in \mathcal{E}_h^i} \int_e |\boldsymbol{\omega}_l \cdot \boldsymbol{n}_e| |[v^l]|^2 d\sigma(\boldsymbol{x}) + \int_{\partial X_+^l} \boldsymbol{\omega}_l \cdot \boldsymbol{n} \, |v^l|^2 d\sigma(\boldsymbol{x}) \right) \\
&\quad \lesssim a_h^{(1)}(\boldsymbol{v}_h, \boldsymbol{v}_h).
\end{aligned}$$

Adding the quantity $c_p \sum_{l=1}^{L} w_l \sum_{e \in \mathcal{E}_h^i} \int_e |[v^l]|^2 d\sigma(\boldsymbol{x})$ to the leftmost and rightmost terms in the above inequality gives

$$\begin{aligned}
&\sum_{l=1}^{L} w_l \left(||v^l||_{0,X}^2 + c_p \sum_{e \in \mathcal{E}_h^i} \int_e |[v^l]|^2 d\sigma(\boldsymbol{x}) + \int_{\partial X_+^l} \boldsymbol{\omega}_l \cdot \boldsymbol{n} \, |v^l|^2 d\sigma(\boldsymbol{x}) \right) \\
&\qquad \lesssim a_h^{(1)}(\boldsymbol{v}_h, \boldsymbol{v}_h) + c_p \sum_{l=1}^{L} w_l \sum_{e \in \mathcal{E}_h^i} \int_e |[v^l]|^2 d\sigma(\boldsymbol{x}),
\end{aligned}$$

which proves (2.20) by definition. □

The discrete problems (2.13) and (2.14) amount to linear systems of finite dimension. Thus the preceding stability result yields the following theorem.

Theorem 2.3.3. *Under the assumption* (2.5)*, both the discrete-ordinate DG methods* (2.13) *and* (2.14) *have a unique solution.*

In order to give error estimates for the methods (2.13) and (2.14) we first introduce some new notation and give two preliminary lemmas. For all $K \in \mathcal{T}_h$, let P_K be the $L^2(K)$-orthogonal projection operator from $L^2(K)$ onto $P_k(K)$. Then by the scaling argument and the trace theorem we can obtain the following result (cf. [8, 11, 17]).

Lemma 2.3.4. *For all $v \in H^{k+1}(K)$ with $k \geq 1$ and $K \in \mathcal{T}_h$,*

$$||v - P_K v||_{0,K} + h_K^{1/2}||v - P_K v||_{0,\partial K} \lesssim h_K^{k+1}||v||_{k+1,K}.$$

Proof. Let $\hat{K}$ be a particular tetrahedron, for example one with unit length edges, and let T_K be an affine function taking $\hat{K}$ to K. Define $\hat{v}(\hat{\boldsymbol{x}})$ on $\hat{K}$ as

$$\hat{v}(\hat{\boldsymbol{x}}) = v(\boldsymbol{x})$$

where

$$T_K(\hat{\boldsymbol{x}}) = \boldsymbol{x}.$$

Then

$$\hat{v} \in H^{k+1}(\hat{K}) \iff v \in H^{k+1}(K).$$

Define $e_K = v - P_K v$. The standard error estimates for polynomial orthogonal projection (c.f. [8]) give

$$||e_K||_{p,K} \lesssim h_K^{k+1-p}|v|_{k+1,K}. \tag{2.24}$$

By the trace theorem (cf. [8]),

$$||\hat{v}||_{0,\partial\hat{K}} \lesssim ||\hat{v}||_{1,K}.$$

Using the definitions of the norms we find

$$\left\{\int_{\partial\hat{K}} |\hat{v}(\hat{\boldsymbol{x}})|^2 d\sigma(\boldsymbol{x})\right\}^{1/2} \lesssim \left\{\int_{\hat{K}} |\hat{v}(\hat{\boldsymbol{x}})|^2 + |\boldsymbol{\nabla}\hat{v}(\hat{\boldsymbol{x}})|^2 d\boldsymbol{x}\right\}^{1/2}.$$

Scaling to the the element K yields

$$\left\{h_K^{-2}\int_{\partial K} |v(\boldsymbol{x})|^2 d\sigma(\boldsymbol{x})\right\}^{1/2} \lesssim \left\{h_K^{-3}\int_K |v(\boldsymbol{x})|^2 + h_K^2|\boldsymbol{\nabla}v(\boldsymbol{x})|^2 d\boldsymbol{x}\right\}^{1/2}.$$

Thus,

$$h_K^{-1}||v||_{0,\partial K} \lesssim h_K^{-3/2}||v||_{0,K} + h_K^{-1/2}|v|_{1,K}.$$

Finally,

$$h_K^{1/2}||v||_{0,\partial K} \lesssim ||v||_{0,K} + h_K|v|_{1,K}.$$

Applying this estimate to e_K we find

$$h_K^{1/2}||e_K||_{0,\partial K} \lesssim ||e_K||_{0,K} + h_K|e_K|_{1,K}.$$

If we now use (2.24) on both right-hand terms we find

$$h_K^{1/2}||e_K||_{0,\partial K} \lesssim h_K^{k+1}||v||_{k+1,K} + h_K^{k+1}||v||_{k+1,K}.$$

Combining this with the original (2.24) again yields

$$||e_K||_{0,K} + h^{1/2}||e_K||_{0,\partial K} \lesssim h_K^{k+1}||v||_{k+1,K},$$

as desired. □

We now give a useful lemma for bounding the error of the discrete ordinate DG method.

Lemma 2.3.5. *Let $\{W_h\}_{h>0}$ be a family of finite dimensional spaces equipped with the norms $\{||\cdot||_h\}_{h>0}$. Let $b_h(\cdot,\cdot)$ be a uniformly coercive bilinear form over W_h, i.e., there exists a positive constant α independent of h such that*

$$\alpha||w_h||_h^2 \leq b_h(w_h, w_h) \quad \forall\, w_h \in W_h. \tag{2.25}$$

Assume that v is a function such that $b_h(v, w_h)$ is well defined for all $w_h \in W_h$, and v_h is a function in W_h satisfying

$$b_h(v - v_h, w_h) = 0 \quad \forall\, w_h \in W_h. \tag{2.26}$$

Then

$$||v - v_h||_h \leq \inf_{w_h \in W_h} \left\{ ||v - w_h||_h + \frac{1}{\alpha} \sup_{\bar{w}_h \in W_h} \frac{b_h(v - w_h, \bar{w}_h)}{||\bar{w}_h||_h} \right\}. \tag{2.27}$$

Proof. For any $w_h \in W_h$, it follows from the triangle inequality that

$$||v - v_h||_h \leq ||v - w_h||_h + ||w_h - v_h||_h. \tag{2.28}$$

From (2.25) and (2.26) we find

$$\begin{aligned} \alpha||w_h - v_h||_h^2 &\leq b_h(w_h - v_h, w_h - v_h) \\ &= b_h(w_h - v + v - v_h, w_h - v_h) \\ &= b_h(w_h - v, w_h - v_h) + b_h(v - v_h, w_h - v_h) \\ &= b_h(w_h - v, w_h - v_h). \end{aligned} \tag{2.29}$$

Thus for $w_h \neq v_h$,

$$||w_h - v_h||_h \leq \frac{1}{\alpha} \frac{b_h(w_h - v, w_h - v_h)}{||w_h - v_h||_h} \leq \frac{1}{\alpha} \sup_{\bar{w}_h \in W_h} \frac{b_h(v - w_h, \bar{w}_h)}{||\bar{w}_h||_h}. \tag{2.30}$$

By (2.28) we have

$$||v - v_h||_h \leq ||v - w_h||_h + ||w_h - v_h||_h.$$

Combining with (2.30) yields

$$||v - v_h||_h \leq ||v - w_h||_h + \frac{1}{\alpha} \sup_{\bar{w}_h \in W_h} \frac{b_h(v - w_h, \bar{w}_h)}{||\bar{w}_h||_h}.$$

Since w_h is an arbitrary element of W_h we have

$$||v - v_h||_h \leq \inf_{w_h \in W_h} \left\{ ||v - w_h||_h + \frac{1}{\alpha} \sup_{\bar{w}_h \in W_h} \frac{b_h(v - w_h, \bar{w}_h)}{||\bar{w}_h||_h} \right\}.$$

as desired. □

Now, we are ready to give error estimates for methods (2.13) and (2.14).

Theorem 2.3.6. *Let* $\{u^l\}$ *be the solution to* (2.6) *and assume it has regularity* $k+1$ *for* $1 \le l \le L$. *Then under the assumption* (2.5)*, the discrete-ordinate DG method* (2.13) *admits the following error estimate:*

$$|||\{u^l\} - \boldsymbol{u}_h|||_h^{(1)} \lesssim h^{k+1/2}\bigg(\sum_{l=1}^{L} w_l ||u^l||_{k+1,X}^2\bigg)^{1/2}.$$

Proof. By (2.7) and the definitions (2.11)–(2.12), we have

$$a_h^{(1)}(\{u^l\}, \boldsymbol{v}_h) = f(\boldsymbol{v}_h) \quad \forall\, \boldsymbol{v}_h \in V_h.$$

Subtract (2.13) from the above equality to obtain the Galerkin orthogonality

$$a_h^{(1)}(\{u^l\} - \boldsymbol{u}_h, \boldsymbol{v}_h) = 0 \quad \forall\, \boldsymbol{v}_h \in V_h.$$

Therefore, it follows from Lemma 2.3.5 and the stability estimate (2.19) that

$$|||\{u^l\} - \boldsymbol{u}_h|||_h^{(1)} \lesssim |||\{u^l\} - \{P_h u^l\}|||_h^{(1)} + \sup_{\boldsymbol{v}_h \in \boldsymbol{V}_h} \frac{a_h^{(1)}(\{u^l\} - \{P_h u^l\}, \boldsymbol{v}_h)}{|||\boldsymbol{v}_h|||_h^{(1)}}, \tag{2.31}$$

where P_h denotes the L^2-orthogonal operator onto V_h in an elementwise way, i.e., for $v \in L^2(X)$, $P_h v|_K = P_K v$. By the definition of $|||\cdot|||_h^{(1)}$, we know

$$\begin{aligned}
\big(|||\{u^l\} - \{P_h u^l\}|||_h^{(1)}\big)^2 \lesssim & \sum_{l=1}^{L} w_l ||u^l - P_h u^l||_{0,X}^2 + \sum_{l=1}^{L} w_l \sum_{e \in \mathcal{E}_h^i} \int_e |[u^l - P_h u^l]|^2 d\sigma(\boldsymbol{x}) \\
& + \sum_{l=1}^{L} w_l \int_{\partial X_+^l} |u^l - P_h u^l|^2 d\sigma(\boldsymbol{x}) \\
\lesssim & \sum_{l=1}^{L} w_l \sum_{K \in \mathcal{T}} ||u^l - P_h u^l||_{0,K}^2 + \sum_{l=1}^{L} w_l \sum_{K \in \mathcal{T}_h} ||[u^l - P_h u^l]||_{0,\partial K}^2 \\
& + \sum_{l=1}^{L} w_l \sum_{K \in \mathcal{T}} ||u^l - P_h u^l||_{0,\partial K}^2.
\end{aligned}$$

Collecting terms and applying Lemma 2.3.4 shows that

$$\begin{aligned}\left(|||\{u^l\}-\{P_hu^l\}|||_h^{(1)}\right)^2 \lesssim & \sum_{l=1}^{L} w_l \sum_{K\in\mathcal{T}_h} ||u_K^l - P_K u_K^l||_{0,K}^2 \\ & + \sum_{l=1}^{L} w_l \sum_{K\in\mathcal{T}_h} ||u_K^l - P_K u_K^l||_{0,\partial K}^2 \\ \lesssim & \sum_{l=1}^{L} w_l \sum_{K\in\mathcal{T}_h} h_K^{2(k+1)} ||u_K^l||_{k+1,K}^2 \\ & + \sum_{l=1}^{L} w_l \sum_{K\in\mathcal{T}_h} h_K^{2(k+1)-1} ||u_K^l||_{k+1,K}^2 \\ \lesssim & h^{2(k+1)-1} \sum_{l=1}^{L} w_l ||u^l||_{k+1,X}^2,\end{aligned}$$

i.e.,

$$|||\{u^l\}-\{P_hu^l\}|||_h^{(1)} \lesssim h^{k+1/2}\left(\sum_{l=1}^{L} w_l ||u^l||_{k+1,X}^2\right)^{1/2}. \tag{2.32}$$

On the other hand, by the definition of $a_h(\cdot,\cdot)$, we have

$$a_h^{(1)}(\{u^l\}-\{P_hu^l\},\boldsymbol{v}_h) = \mathbb{I}_1 + \mathbb{I}_2. \tag{2.33}$$

where

$$\begin{aligned}\mathbb{I}_1 := & \sum_{l=1}^{L} w_l \sum_{K\in\mathcal{T}_h} \Big\{ \int_{\partial K} \boldsymbol{\omega}_l\cdot\boldsymbol{n}_K \, (\widehat{u^l - P_h u^l}) \, v_{h,K}^l d\sigma(\boldsymbol{x}) \\ & - \int_K (u_K^l - P_K u_K^l)\, \boldsymbol{\omega}_l\cdot\boldsymbol{\nabla} v_{h,K}^l \, dx \Big\}, \\ \mathbb{I}_2 := & \sum_{l=1}^{L} w_l \int_X \mu_t (u^l - P_h u^l) v_h^l dx - \sum_{l=1}^{L} w_l \int_X \mu_s \sum_{i=1}^{L} w_i g(\cdot, \boldsymbol{\omega}_l\cdot\boldsymbol{\omega}_i)(u^i - P_h u^i) v_h^l dx.\end{aligned}$$

Since $\boldsymbol{\omega}_l\cdot\boldsymbol{\nabla} v_{h,K}^l \in P_k(K)$, we know

$$\int_K (u_K^l - P_K u_K^l)\, \boldsymbol{\omega}_l\cdot\boldsymbol{\nabla} v_{h,K}^l \, dx = 0.$$

For each l between 1 and L, denote by $\mathcal{E}_h^{l,+}$ the set of all faces of $\mathcal{T}_h$ on ∂X_+^l. Using the Cauchy-Schwarz inequality and Lemma 2.3.4, we have

$$\begin{aligned}
|\mathbb{I}_1| &\lesssim \sum_{l=1}^{L} w_l \sum_{e\in\mathcal{E}_h^{l,+}} \int_e \boldsymbol{\omega}_l\cdot\boldsymbol{n}\,|u^l - P_h u^l||v_h^l| d\sigma(\boldsymbol{x}) \\
&\quad + \sum_{l=1}^{L} w_l \sum_{e\in\mathcal{E}_h^{i}} \int_e |\boldsymbol{\omega}_l\cdot\boldsymbol{n}_e|\,|u^l - P_h u^l||[v_h^l]| d\sigma(\boldsymbol{x}) \\
&\lesssim \sum_{l=1}^{L} w_l \Big[\sum_{e\in\mathcal{E}_h^{l,+}} \int_e |u^l - P_h u^l|^2 d\sigma(\boldsymbol{x})\Big]^{1/2} \Big[\sum_{e\in\mathcal{E}_h^{l,+}} \int_e \boldsymbol{\omega}_l\cdot\boldsymbol{n}\,|v_h^l|^2 d\sigma(\boldsymbol{x})\Big]^{1/2} \\
&\quad + \sum_{l=1}^{L} w_l \Big[\sum_{e\in\mathcal{E}_h^{i}} \int_e |u^l - P_h u^l|^2 d\sigma(\boldsymbol{x})\Big]^{1/2} \Big[\sum_{e\in\mathcal{E}_h^{i}} \int_e |\boldsymbol{\omega}_l\cdot\boldsymbol{n}_e|\,|[v_h^l]|^2 d\sigma(\boldsymbol{x})\Big]^{1/2} \\
&\lesssim h^{k+1/2}\Big(\sum_{l=1}^{L} w_l ||u^l||_{r,X}^2\Big)^{1/2} |||\boldsymbol{v}_h|||_h^{(1)}. \qquad (2.34)
\end{aligned}$$

By the definition of $\mathbb{I}_2$ we have

$$\begin{aligned}
|\mathbb{I}_2| &= \Big|\sum_{l=1}^{L} w_l \int_X \mu_t (u^l - P_h u^l) v_h^l dx - \sum_{l=1}^{L} w_l \int_X \mu_s \sum_{i=1}^{L} w_i g(\cdot, \boldsymbol{\omega}_l\cdot\boldsymbol{\omega}_i)(u^i - P_h u^i) v_h^l dx\Big| \\
&\le \sum_{l=1}^{L} w_l \int_X \mu_t |(u^l - P_h u^l)||v_h^l| dx \\
&\quad + \sum_{l=1}^{L} w_l \int_X \mu_s \sum_{i=1}^{L} w_i g(\cdot, \boldsymbol{\omega}_l\cdot\boldsymbol{\omega}_i)|u^i - P_h u^i||v_h^l| dx.
\end{aligned}$$

Applying the Cauchy-Schwarz inequality to the left term, and Lemma 2.3.1 to the right term we find that

$$\begin{aligned}
|\mathbb{I}_2| &\lesssim \sum_{l=1}^{L} w_l \left[\int_X |u^l - P_h u^l|^2 dx\right]^{1/2} \left[\int_x \mu_t^2 |v_h^l|^2 dx\right]^{1/2} \\
&\quad + \left[\sum_{l=1}^{L} w_l \int_X m\mu_s (u^l - P_h u^l)^2 dx\right]^{1/2} \left[\sum_{l=1}^{L} \int_X m\mu_s (v_h^l)^2\right]^{1/2}.
\end{aligned}$$

Combining terms yields

$$|\mathbb{I}_2| \lesssim \left[\sum_{l=1}^{L} w_l \int_X \left(u^l - P_h u^l\right)^2 dx\right]^{1/2} \left[\sum_{l=1}^{L} w_l \int_X \left(v_h^l\right)^2 dx\right]^{1/2}.$$

Remembering Lemma 2.3.4 shows that

$$|\mathbb{I}_2| \lesssim h^{k+1/2}\bigg(\sum_{l=1}^{L} w_l ||u^l||_{k+1,X}^2\bigg)^{1/2} |||\boldsymbol{v}_h|||_h^{(1)}.$$

This, together with (2.31)–(2.34), leads to the stated error estimate. □

Using an argument similar to the proof of the above theorem along with the stability result (2.20) and the definition (2.15), we can also derive an error estimate for the method (2.14), described as follows.

Theorem 2.3.7. *Under the assumption* (2.5)*, the discrete-ordinate DG method* (2.14) *admits the following error estimate:*

$$|||\{u^l\} - \boldsymbol{u}_h|||_h^{(2)} \lesssim \left(1 + \sqrt{c_p}\right) h^{k+1/2}\bigg(\sum_{l=1}^{L} w_l ||u^l||_{k+1,X}^2\bigg)^{1/2},$$

whenever the problem (2.6) *has a solution* $\{u^l\}$ *with the regularity* $u^l \in H^{k+1}(X)$*,* $1 \le l \le L$*.*

We now give error estimates between the solution u to the RTE and the solution $\{u^l\}$ to the semi-discretized problem (2.6). Define the errors

$$\varepsilon^l(\boldsymbol{x}) = u(\boldsymbol{x}, \boldsymbol{\omega}_l) - u^l(\boldsymbol{x}),\ 1 \le l \le L. \tag{2.35}$$

In order to ease presentation, we only consider the case for the Henyey-Greenstein phase function g of (1.5), though all the derivation may be carried out after straightforward modification for more general phase functions. Subtracting (2.6) from (1.9)

we see that, for all $1 \le l \le L$,

$$\begin{aligned}\boldsymbol{\omega}_l \cdot \nabla \varepsilon^l + \mu_t \varepsilon^l &= \mu_s \int_\Omega g(\boldsymbol{\omega}_l \cdot \boldsymbol{\omega})\, u(\cdot, \boldsymbol{\omega})\, d\sigma(\boldsymbol{\omega}) - \mu_s \sum_{i=1}^{L} w_i g(\boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i)\, u^i \\ &= \mu_s \sum_{i=1}^{L} w_i g(\boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i)\, \varepsilon^i + \eta^l, \end{aligned} \tag{2.36}$$

with η^l defined as

$$\eta^l = \mu_s \int_\Omega g(\boldsymbol{\omega}_l \cdot \boldsymbol{\omega})\, u(\cdot, \boldsymbol{\omega})\, d\sigma(\boldsymbol{\omega}) - \mu_s \sum_{i=1}^{L} w_i g(\boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i)\, u(\cdot, \boldsymbol{\omega}_i). \tag{2.37}$$

We now multiply (2.36) by $w_l \varepsilon^l$, and integrate over X to find

$$\begin{aligned} & w_l \int_X \boldsymbol{\omega}_l \cdot \nabla \varepsilon^l \varepsilon^l dx + w_l \int_X \mu_t (\varepsilon^l)^2 dx \\ & \quad = w_l \int_X \mu_s \sum_{i=1}^{L} w_i g(\boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i)\, \varepsilon^i \varepsilon^l dx + w_l \int_X \eta^l \varepsilon^l dx. \end{aligned} \tag{2.38}$$

Summing over $1 \le l \le L$ and rewriting the first term we have

$$\begin{aligned} & \frac{1}{2} \sum_{l=1}^{L} w_l \int_X \boldsymbol{\omega}_l \cdot \nabla (\varepsilon^l)^2 dx + \sum_{l=1}^{L} w_l \int_X \mu_t (\varepsilon^l)^2 dx \\ & \quad = \sum_{l=1}^{L} w_l \int_X \mu_s \sum_{i=1}^{L} w_i g(\boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i)\, \varepsilon^i \varepsilon^l dx + \sum_{l=1}^{L} w_l \int_X \eta^l \varepsilon^l dx. \end{aligned} \tag{2.39}$$

As $u^l(\boldsymbol{x}) = u(\boldsymbol{x}, \boldsymbol{\omega}_l) = 0$ on ∂X^l_- we use the Stokes theorem to find that

$$\begin{aligned} & \frac{1}{2} \sum_{l=1}^{L} w_l \int_{\partial X^l_+} \boldsymbol{\omega}_l \cdot \boldsymbol{n}\, (\varepsilon^l)^2 d\sigma(\boldsymbol{x}) + \sum_{l=1}^{L} w_l \int_X \mu_t (\varepsilon^l)^2 dx \\ & \quad - \sum_{l=1}^{L} w_l \int_X \mu_s \sum_{i=1}^{L} w_i g(\boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i)\, \varepsilon^i \varepsilon^l dx = \sum_{l=1}^{L} w_l \int_X \varepsilon^l \eta^l dx. \end{aligned}$$

Applying the estimate (2.22) to the equation above, we find

$$\frac{1}{2} \sum_{l=1}^{L} w_l \int_{\partial X^l_+} \boldsymbol{\omega}_l \cdot \boldsymbol{n}\, (\varepsilon^l)^2 d\sigma(\boldsymbol{x}) + c_0' \sum_{l=1}^{L} w_l \int_X (\varepsilon^l_h)^2 dx \le \sum_{l=1}^{L} w_l \int_X \varepsilon^l \eta^l dx.$$

Therefore,

$$\sum_{l=1}^{L} w_l \int_X (\varepsilon^l)^2 dx \lesssim \sum_{l=1}^{L} w_l \int_X (\eta^l)^2 dx \tag{2.40}$$

by means of the Hölder inequality and the Cauchy-Schwarz inequality.

Assume that the numerical quadrature (2.1) integrates exactly all polynomials of degree no more than n. Then, by Theorem 2.2.1 and the definition (2.37),

$$\begin{aligned} |\eta^l(\boldsymbol{x})| &= \mu_s(\boldsymbol{x}) \left| \int_\Omega g(\boldsymbol{\omega}_l \cdot \boldsymbol{\omega})\, u(\boldsymbol{x}, \boldsymbol{\omega})\, d\sigma(\boldsymbol{\omega}) - \sum_{i=1}^{L} w_i g(\boldsymbol{\omega}_l \cdot \boldsymbol{\omega}_i)\, u(\boldsymbol{x}, \boldsymbol{\omega}_i) \right| \\ &\lesssim c(r', g) n^{-r'} ||u(\boldsymbol{x}, \cdot)||_{r', \Omega}, \end{aligned}$$

where $r' > 1$ and $c(r', g)$ is a positive constant depending on r' and the phase function g. Combining (2.40) and this estimate gives

$$\sum_{l=1}^{L} w_l \int_X (\varepsilon^l)^2 dx \lesssim c(r', g) n^{-r'} \int_X ||u(\boldsymbol{x}, \cdot)||^2_{r', \Omega} dx.$$

This combined with Theorems 2.3.6 and 2.3.7 leads to the following results.

Corollary 2.3.8. *Suppose that* (2.5) *holds and the numerical quadrature* (2.1) *has a degree n of precision. Suppose further that $u^l \in H^{k+1}(X)$ for $1 \le l \le L$. If the phase function g appearing in* (1.3) *is taken as the Henyey-Greenstein function with $\eta \in (-1, 1)$, then the discrete-ordinate DG method* (2.13) *admits the following error estimate for $r' > 1$:*

$$\begin{aligned} \left(\sum_{l=1}^{L} w_l ||u(\cdot, \boldsymbol{\omega}_l) - u_h^l||^2_{0,X} \right)^{1/2} &\lesssim h^{k+1/2} \left(\sum_{l=1}^{L} w_l ||u^l||^2_{k+1,X} \right)^{1/2} \\ &\quad + c(r', g) n^{-r'} \left(\int_X ||u(\cdot, \cdot)||^2_{r', \Omega} dx \right)^{1/2}, \end{aligned}$$

whenever $u \in L^2(X, H^{r'}(\Omega))$.

Corollary 2.3.9. *Suppose that* (2.5) *holds and the numerical quadrature* (2.1) *has a degree n of precision. Suppose further that $u^l \in H^{k+1}(X)$ for $1 \le l \le L$. If the phase function g appearing in* (1.3) *is taken as the Henyey-Greenstein function with $\eta \in (-1, 1)$, then the discrete-ordinate DG method* (2.14) *admits the following error estimate for $r' > 1$:*

$$\left(\sum_{l=1}^{L} w_l ||u(\cdot, \boldsymbol{\omega}_l) - u_h^l||_{0,X}^2 \right)^{1/2} \lesssim \left(1 + \sqrt{c_p}\right) h^{k+1/2} \left(\sum_{l=1}^{L} w_l ||u^l||_{k+1,X}^2 \right)^{1/2} + c(r', g) n^{-r'} \left(\int_X ||u(\cdot, \cdot)||_{r',\Omega}^2 dx \right)^{1/2},$$

whenever $u \in L^2(X, H^{r'}(\Omega))$.

2.4 Numerical results

We first comment on the assumption (2.5). Recalling the normalization condition (1.4), we expect that as long as the numerical quadrature (2.2) is sufficiently accurate, the quantity $m(\boldsymbol{x})$ defined in (2.4) will be close to 1, and then the assumption (2.5) will follow from the condition (1.12). As an example, for the Henyey–Greenstein phase function, consider using the S_N quadratures for the numerical integration (cf. [14]). Due to the symmetry of the S_N quadratures, it is easy to see that for any $\eta \in (-1, 1)$, the value of $m = m(\boldsymbol{x})$ corresponding to η equals that corresponding to $-\eta$. Also, note that $m = 1$ for $\eta = 0$. In Table 2.1, we list the value of m for $\eta = 0.2$, 0.4, 0.6, 0.8, and a few choices of N.

Now we give several examples of solving the boundary value problem (1.9)–(1.10) using the discrete ordinate DG method. For the angular variable $\boldsymbol{\omega}$, we will write both $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3)^T$ and $\boldsymbol{\omega} = (\sin\theta\cos\psi, \sin\theta\sin\psi, \cos\theta)^T$. The spatial

	$\eta = 0.2$	$\eta = 0.4$	$\eta = 0.6$	$\eta = 0.8$
$N = 2$	1.004065	1.082299	1.627474	5.790811
$N = 4$	1.000013	1.002545	1.073823	2.216768
$N = 6$	1.000001	1.000355	1.016232	1.476457
$N = 8$	1.000000	1.000028	1.004487	1.249246
$N = 12$	1.000000	1.000001	1.000500	1.083949
$N = 16$	0.999999	1.000000	1.000082	1.033950

Table 2.1: Value of m for several choices of η and N

variable will be denoted as $\boldsymbol{x} = (x_1, x_2, x_3)^T$. We use the Henyey–Greenstein phase function defined in (1.5) for each example, and will just specify the value of η. We choose $\mu_t(\boldsymbol{x}) = \mu_t$ and $\mu_s(\boldsymbol{x}) = \mu_s$ to be constants. We report numerical values of the error $|||u - u_h|||_h$, defined by (2.17) for $c_p = 0$ and by (2.18) for $c_p > 0$, and the approximate $L^2(X \times \Omega)$ error

$$\|u - u_h\|_h = \left[\sum_{l=1}^{L} w_l \|u(\cdot, \boldsymbol{\omega}_l) - u_h^l\|_{0,X}^2 \right]^{\frac{1}{2}}. \tag{2.41}$$

In the first two examples, the spatial domain is the unit cube $X = (0, 1)^3$. For a positive integer n, we partition $\overline{X}$ into n^3 subcubes $\{X_i\}$, each with edge length $1/n$. Denote by S the set of all the centers and vertices of the subcubes. A mesh is generated by creating the Delaunay tesselation of the points in S. Denote by h the maximum length of the edges of the tetrahedron in the mesh; in this meshing scheme we have $h = \sqrt{2}/n$. A sample mesh (for $n = 1$) is shown in Figure 2.1. We choose

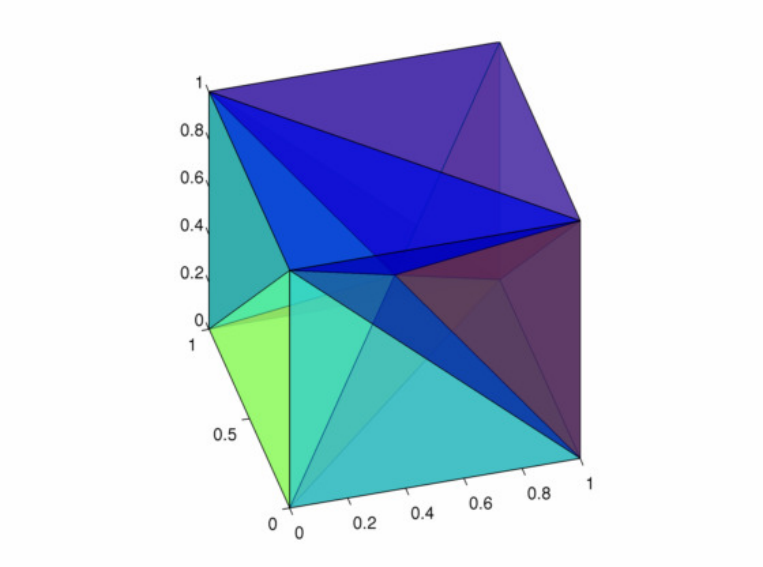

Figure 2.1: A sample mesh

the local polynomial degree $k = 1$. In the examples, if not stated otherwise, we use the S_4 quadrature, so that there are 24 different angular directions $\{\boldsymbol{\omega}_l\}_{l=1}^{24}$. In some of the examples, we use the more accurate S_{12} quadrature and there are 168 different angular directions $\{\boldsymbol{\omega}_l\}_{l=1}^{168}$. The error figures correspond to the parameter $c_p = 0.1$.

Example 2.4.1. We take $\sigma_t(\boldsymbol{x}) = 2$, $\sigma_s(\boldsymbol{x}) = 1$, and $\eta = 0$. With the right hand side function

$$f(\boldsymbol{x}, \boldsymbol{\omega}) = \pi\omega_1 \cos(\pi x_1)\sin(\pi x_2)\sin(\pi x_3) + \pi\omega_2 \sin(\pi x_1)\cos(\pi x_2)\sin(\pi x_3)$$
$$+ \pi\omega_3 \sin(\pi x_1)\sin(\pi x_2)\cos(\pi x_3) + \sin(\pi x_1)\sin(\pi x_2)\sin(\pi x_3),$$

the true solution is

$$u(\boldsymbol{x}, \boldsymbol{\omega}) = \sin(\pi x_1)\sin(\pi x_2)\sin(\pi x_3).$$

Errors of numerical solutions are shown in Tables 2.2, 2.3, and Figures 2.2,

h	$c_p = 0$	$c_p = .01$	$c_p = .1$	$c_p = 1$	$c_p = 10$
$\frac{\sqrt{2}}{2}$	1.341e-01	1.333e-01	1.313e-01	1.479e-01	1.849e-01
$\frac{\sqrt{2}}{4}$	3.726e-02	3.692e-02	3.603e-02	3.910e-02	4.248e-02
$\frac{\sqrt{2}}{8}$	9.835e-03	9.721e-03	9.411e-03	9.756e-03	1.008e-02
$\frac{\sqrt{2}}{16}$	2.514e-03	2.479e-03	2.393e-03	2.440e-03	2.486e-03

Table 2.2: $||u - u_h||$ for Example 2.4.1, S_4 quadrature

h	$c_p = 0$	$c_p = .01$	$c_p = .1$	$c_p = 1$	$c_p = 10$
$\frac{\sqrt{2}}{2}$	4.014e-01	1.641e-01	2.132e-01	2.809e-01	2.533e-01
$\frac{\sqrt{2}}{4}$	1.621e-01	4.824e-02	7.355e-02	9.524e-02	6.144e-02
$\frac{\sqrt{2}}{8}$	6.126e-02	1.431e-02	2.556e-02	3.195e-02	1.731e-02
$\frac{\sqrt{2}}{16}$	2.229e-02	4.369e-03	8.910e-03	1.101e-02	5.414e-03

Table 2.3: $|||u - u_h|||$ for Example 2.4.1, S_4 quadrature

2.3.

The numerical approximations show good convergence properties despite the use of relatively few angular nodes. This is likely due to the simple phase function g.

□

Example 2.4.2. As a second example, we use a non-constant phase function. The true solution is chosen to be

$$u(\boldsymbol{x}, \boldsymbol{\omega}) = 10\omega_3 \sin(\pi x_1) \sin(\pi x_2) \sin(\pi x_3).$$

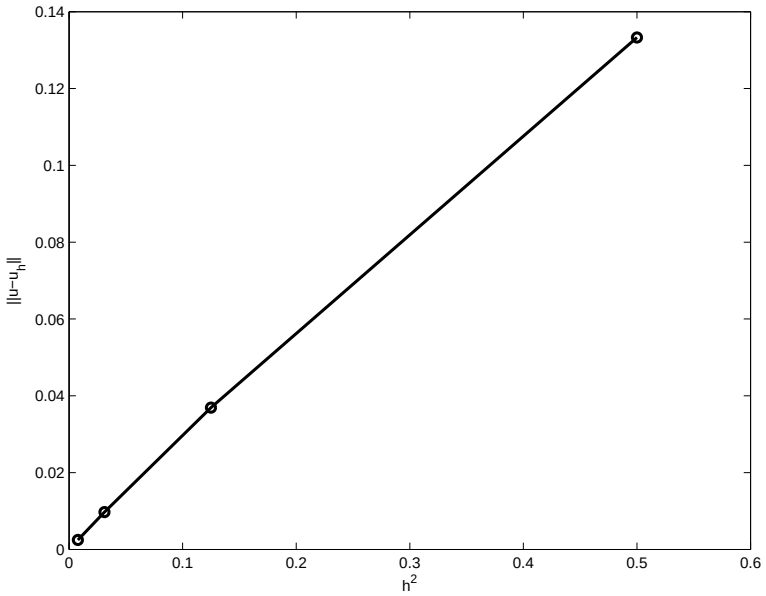

Figure 2.2: $||u - u_h||$ for Example 2.4.1, S_4 quadrature

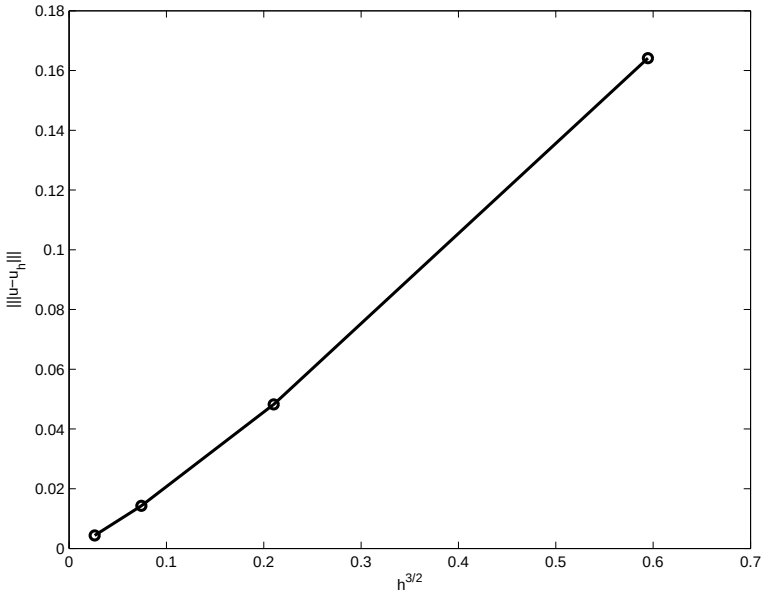

Figure 2.3: $|||u - u_h|||$ for Example 2.4.1, S_4 quadrature

The corresponding right hand side function is

$$\begin{aligned} f(\boldsymbol{x},\boldsymbol{\omega}) = {} & 10\pi\omega_3^2 \sin(\pi x_1)\sin(\pi x_2)\cos(\pi x_3) + 10\pi\omega_2\omega_3 \sin(\pi x_1)\cos(\pi x_2)\sin(\pi x_3) \\ & + 10\pi\omega_1\omega_3 \cos(\pi x_1)\sin(\pi x_2)\sin(\pi x_3) \\ & + 10\left(\mu_t\omega_3 - \eta\mu_s\cos\theta\right)\sin(\pi x_1)\sin(\pi x_2)\sin(\pi x_3), \end{aligned}$$

where μ_s, μ_t are the scattering and absorption parameters, and η is the degree of anisotropy. We take $\mu_t = 3$ and $\mu_s = 1$.

Initially, we set $\eta = 0.1$, the numerical results are reported in Tables 2.4, 2.5, and Figures 2.4, 2.5. We compare these results with the results from using S_{12} quadrature for the numerical solution, in order to investigate the effects of quadrature order on the accuracy of the numerical solution. The results for using the S_{12} quadrature are given in Tables 2.6, 2.7 and Figures 2.6, 2.7. Values marked as N/A are not available due to the large amount of computational resources required. Comparing the results from S_4 quadrature and S_{12} quadrature, we conclude that the numerical solution errors in Tables 2.4 and 2.6 are mainly due to the spatial discretization.

For $\eta = 0.5$, the numerical results with the S_4 quadrature are given in Tables 2.8, 2.9, and Figures 2.8, 2.9. The numerical solution errors, especially the error $\|u - u_h\|_h$, for $h = \sqrt{2}/16$ are larger than expected, indicating that the error from S_4 quadrature may have become a significant part of the total error. This is expected, as for large values of η the phase function g becomes nearly singular. We see that numerical results with S_{12} quadrature, as given in Tables 2.10, 2.11, and Figures 2.10, 2.11, show significant improvement. Therefore, for values of η near one, a very large

h	$c_p = 0$	$c_p = .01$	$c_p = .1$	$c_p = 1$	$c_p = 10$
$\frac{\sqrt{2}}{2}$	7.420e-01	7.379e-01	7.237e-01	8.024e-01	9.649e-01
$\frac{\sqrt{2}}{4}$	2.097e-01	2.080e-01	2.027e-01	2.200e-01	2.417e-01
$\frac{\sqrt{2}}{8}$	5.598e-02	5.539e-02	5.367e-02	5.581e-02	5.801e-02
$\frac{\sqrt{2}}{16}$	1.440e-02	1.421e-02	1.374e-02	1.403e-02	1.433e-02

Table 2.4: $||u - u_h||$ for Example 2.4.2, $\eta = .1$, S_4 quadrature

h	$c_p = 0$	$c_p = .01$	$c_p = .1$	$c_p = 1$	$c_p = 10$
$\frac{\sqrt{2}}{2}$	2.257e+00	9.227e-01	1.204e+00	1.603e+00	1.390e+00
$\frac{\sqrt{2}}{4}$	9.180e-01	2.723e-01	4.177e-01	5.454e-01	3.539e-01
$\frac{\sqrt{2}}{8}$	3.497e-01	8.147e-02	1.462e-01	1.837e-01	9.975e-02
$\frac{\sqrt{2}}{16}$	1.279e-01	2.504e-02	5.119e-02	6.346e-02	3.122e-02

Table 2.5: $|||u - u_h|||$ for Example 2.4.2, $\eta = .1$, S_4 quadrature

h	$c_p = 0$	$c_p = .01$	$c_p = .1$	$c_p = 1$	$c_p = 10$
$\frac{\sqrt{2}}{2}$	7.428e-01	7.392e-01	7.260e-01	8.030e-01	9.643e-01
$\frac{\sqrt{2}}{4}$	2.120e-01	2.101e-01	2.035e-01	2.199e-01	2.418e-01
$\frac{\sqrt{2}}{8}$	5.743e-02	5.664e-02	5.400e-02	N/A	N/A
$\frac{\sqrt{2}}{16}$	1.496e-02	1.469e-02	1.385e-02	N/A	N/A

Table 2.6: $||u - u_h||$ for Example 2.4.2, $\eta = .1$, S_{12} quadrature

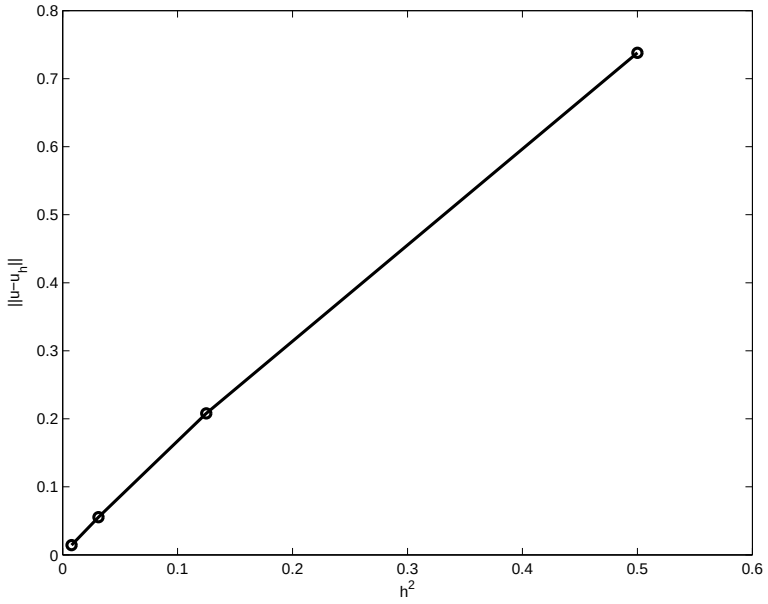

Figure 2.4: $||u - u_h||$ for Example 2.4.2, $\eta = .1$, S_4 quadrature

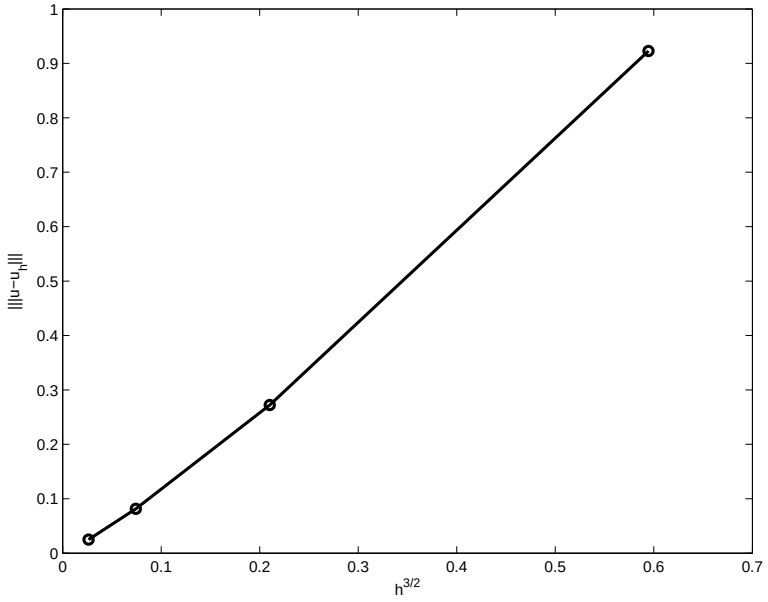

Figure 2.5: $|||u - u_h|||$ for Example 2.4.2, $\eta = .1$, S_4 quadrature

h	$c_p = 0$	$c_p = .01$	$c_p = .1$	$c_p = 1$	$c_p = 10$
$\frac{\sqrt{2}}{2}$	2.270e+00	9.164e-01	1.192e+00	1.591e+00	1.376e+00
$\frac{\sqrt{2}}{4}$	9.273e-01	2.749e-01	4.173e-01	5.447e-01	3.536e-01
$\frac{\sqrt{2}}{8}$	3.554e-01	8.362e-02	1.469e-01	N/A	N/A
$\frac{\sqrt{2}}{16}$	1.307e-01	2.605e-02	5.163e-02	N/A	N/A

Table 2.7: $|||u - u_h|||$ for Example 2.4.2, $\eta = .1$, S_{12} quadrature

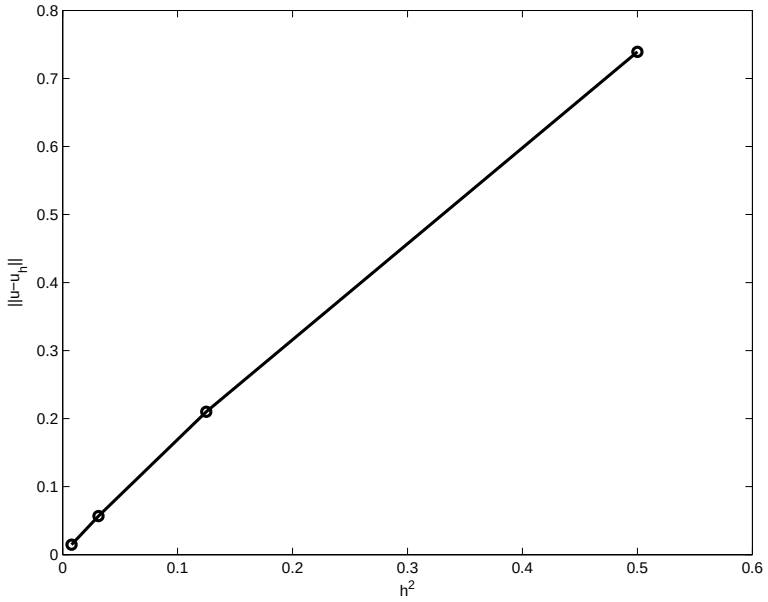

Figure 2.6: $||u - u_h||$ for Example 2.4.2, $\eta = .1$, S_{12} quadrature

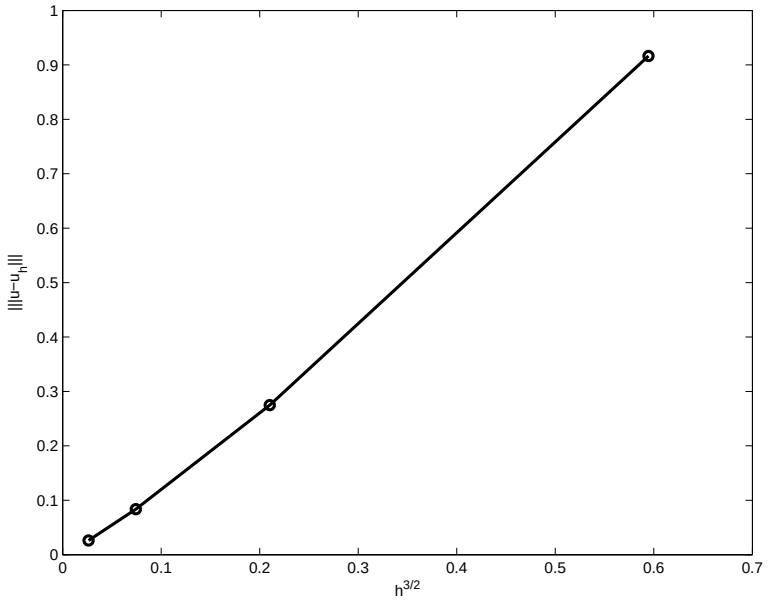

Figure 2.7: $|||u - u_h|||$ for Example 2.4.2, $\eta = .1$, S_{12} quadrature

number of angular nodes will be required for an accurate numerical approximation.

□

Example 2.4.3. As a final numerical example, we choose the spatial domain to be an approximate cylinder (it is actually polyhedral to aid discretization) of unit radius with height 2. We set $\mu_t = 2$, $\mu_s = 1$ and $f(\boldsymbol{x}, \boldsymbol{\omega}) = \chi_B(\boldsymbol{x})$, where B is approximately a ball in the center of the cylinder with radius .5. The Henyey-Greenstein phase function is used with scattering parameter $\eta = .5$. We begin with an initial coarse cylinder, and refine it successively. In the initial mesh, there are 201 elements, and $h \approx 1.06$. With each refinement, the value of h is decreased by a factor of $1/2$ and the number of mesh elements is increased by a factor of 8. The quadrature rule used in this example is obtained by constructing a finite element mesh of the sphere as in

h	$c_p = 0$	$c_p = .01$	$c_p = .1$	$c_p = 1$	$c_p = 10$
$\frac{\sqrt{2}}{2}$	7.441e-01	7.399e-01	7.256e-01	8.084e-01	9.796e-01
$\frac{\sqrt{2}}{4}$	2.118e-01	2.101e-01	2.048e-01	2.229e-01	2.452e-01
$\frac{\sqrt{2}}{8}$	6.724e-02	6.674e-02	6.531e-02	6.753e-02	6.979e-02
$\frac{\sqrt{2}}{16}$	4.128e-02	4.122e-02	4.106e-02	4.127e-02	4.1466e-02

Table 2.8: $||u - u_h||$ for Example 2.4.2, $\eta = .5$, S_4 quadrature

h	$c_p = 0$	$c_p = .01$	$c_p = .1$	$c_p = 1$	$c_p = 10$
$\frac{\sqrt{2}}{2}$	2.275e+00	9.277e-01	1.211e+00	1.609e+00	1.402e+00
$\frac{\sqrt{2}}{4}$	9.253e-01	2.774e-01	4.223e-01	5.494e-01	3.581e-01
$\frac{\sqrt{2}}{8}$	3.550e-01	9.610e-02	1.552e-01	1.911e-01	1.123e-01
$\frac{\sqrt{2}}{16}$	1.383e-01	5.726e-02	7.272e-02	8.186e-02	6.032e-02

Table 2.9: $|||u - u_h|||$ for Example 2.4.2, $\eta = .5$, S_4 quadrature

h	$c_p = 0$	$c_p = .01$	$c_p = .1$	$c_p = 1$	$c_p = 10$
$\frac{\sqrt{2}}{2}$	7.480e-01	7.443e-01	7.316e-01	8.119e-01	9.783e-01
$\frac{\sqrt{2}}{4}$	2.130e-01	2.111e-01	2.045e-01	2.210e-01	2.424e-01
$\frac{\sqrt{2}}{8}$	5.758e-02	5.678e-02	5.414e-02	5.591e-02	5.807e-02
$\frac{\sqrt{2}}{16}$	1.498e-02	1.470e-02	1.387e-02	N/A	N/A

Table 2.10: $||u - u_h||$ for Example 2.4.2, $\eta = .5$, S_{12} quadrature

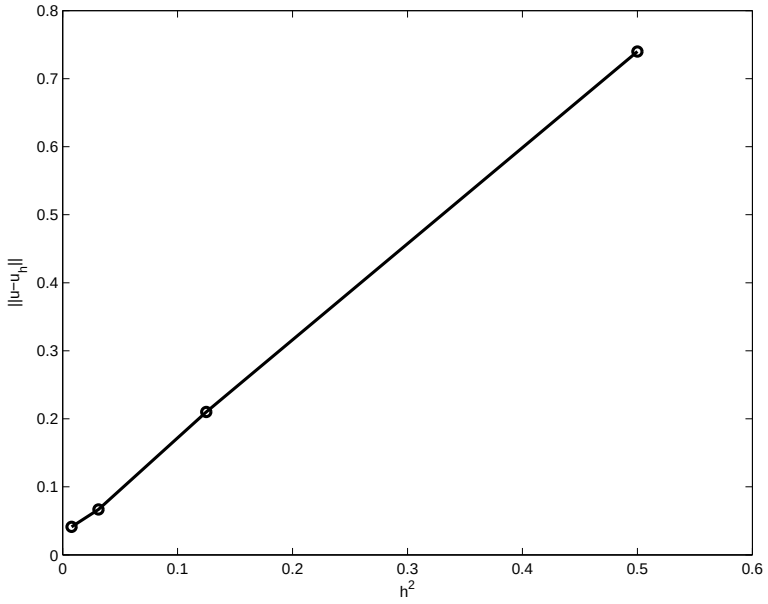

Figure 2.8: $||u - u_h||$ for Example 2.4.2, $\eta = .5$, S_4 quadrature

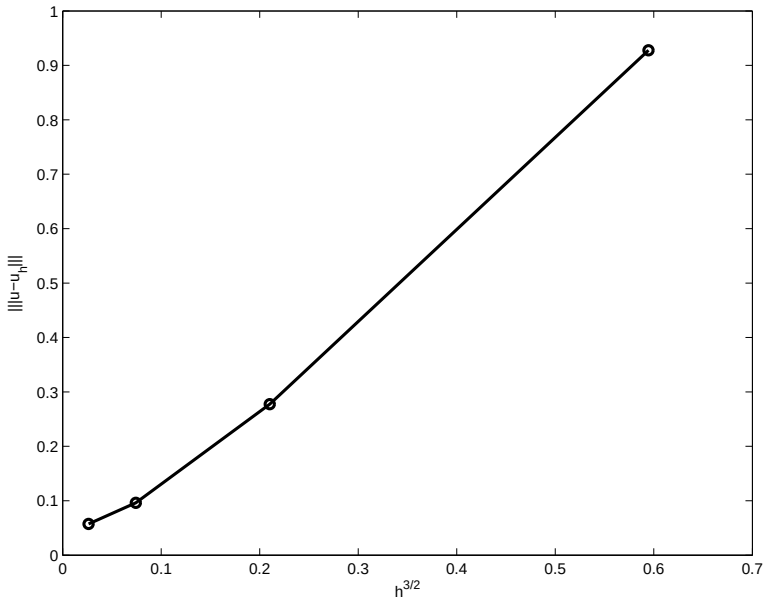

Figure 2.9: $|||u - u_h|||$ for Example 2.4.2, $\eta = .5$, S_4 quadrature

h	$c_p = 0$	$c_p = .01$	$c_p = .1$	$c_p = 1$	$c_p = 10$
$\frac{\sqrt{2}}{2}$	2.280e+00	9.207e-01	1.196e+00	1.593e+00	1.384e+00
$\frac{\sqrt{2}}{4}$	9.308e-01	2.760e-01	4.186e-01	5.455e-01	3.537e-01
$\frac{\sqrt{2}}{8}$	3.562e-01	8.385e-02	1.472e-01	1.838e-01	9.977e-02
$\frac{\sqrt{2}}{16}$	1.309e-01	2.609e-02	5.169e-02	N/A	N/A

Table 2.11: $|||u - u_h|||$ for Example 2.4.2, $\eta = .5$, S_{12} quadrature

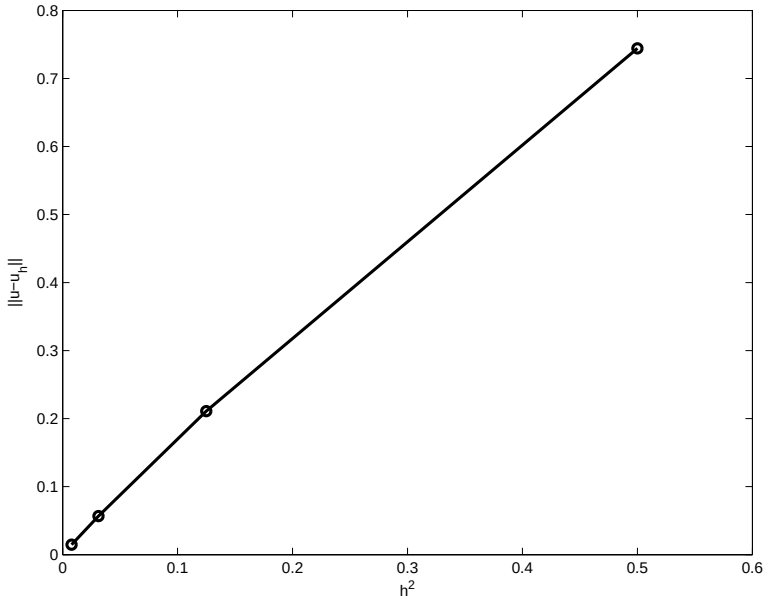

Figure 2.10: $||u - u_h||$ for Example 2.4.2, $\eta = .5$, S_{12} quadrature

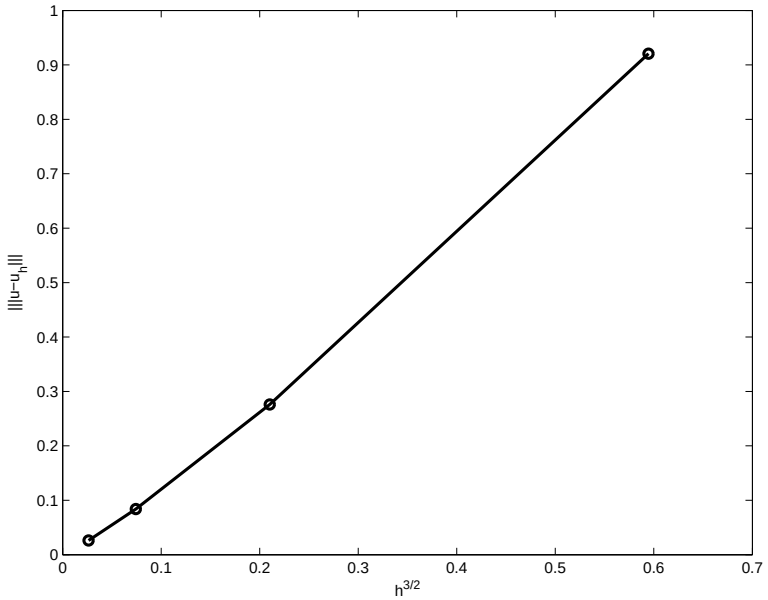

Figure 2.11: $|||u - u_h|||$ for Example 2.4.2, $\eta = .5$, S_{12} quadrature

[4], and requiring that all piecewise linear polynomials on this mesh are integrated exactly. The numerical results are found in Tables 2.12, 2.13, 2.4.3 and Figures 2.12, 2.13 and 2.14.

We note that as in the previous example, the expected convergence order in h is only achieved when enough angular nodes are used.

□

Taking the previous results in to consideration, we see that when the numerical quadrature is sufficiently accurate, the convergence order with respect to meshsize is

$$|||u - u_h|||_h = O(h^{1.5}),$$

and

$$\|u - u_h\|_h = O(h^2).$$

h	$c_p = 0$	$c_p = .01$	$c_p = .1$
1.069	5.379e-02	5.453e-02	6.491e-02
.5346	2.220e-02	2.235e-02	2.631e-02
.2673	9.541e-03	9.498e-03	1.077e-02

Table 2.12: $||u - u_h||$ for Example 2.4.3, 18 angular nodes

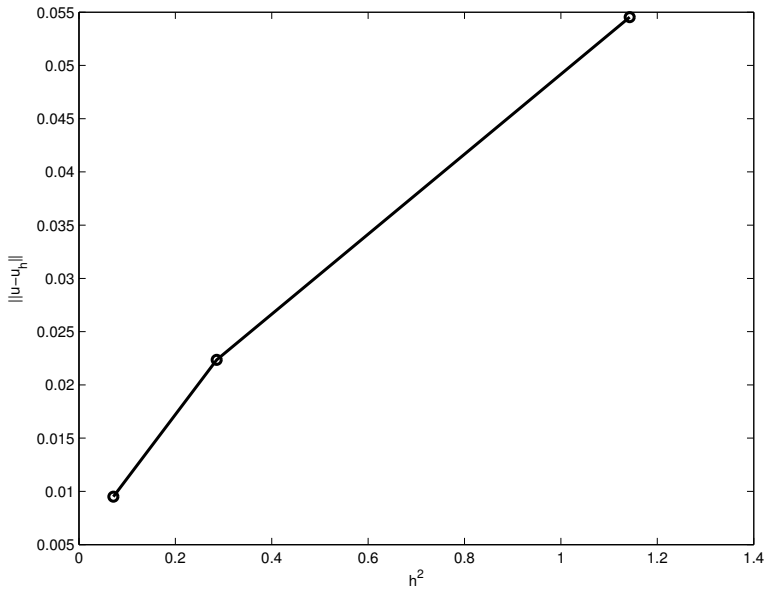

Figure 2.12: $||u - u_h||$ for Example 2.4.3, 18 angular nodes

h	$c_p = 0$	$c_p = .01$	$c_p = .1$
1.069	5.092e-02	5.147e-02	5.980e-02
.5346	1.776e-02	1.784e-02	2.066e-02
.2673	3.100e-03	3.130e-03	4.316e-03

Table 2.13: $||u - u_h||$ for Example 2.4.3, 66 angular nodes

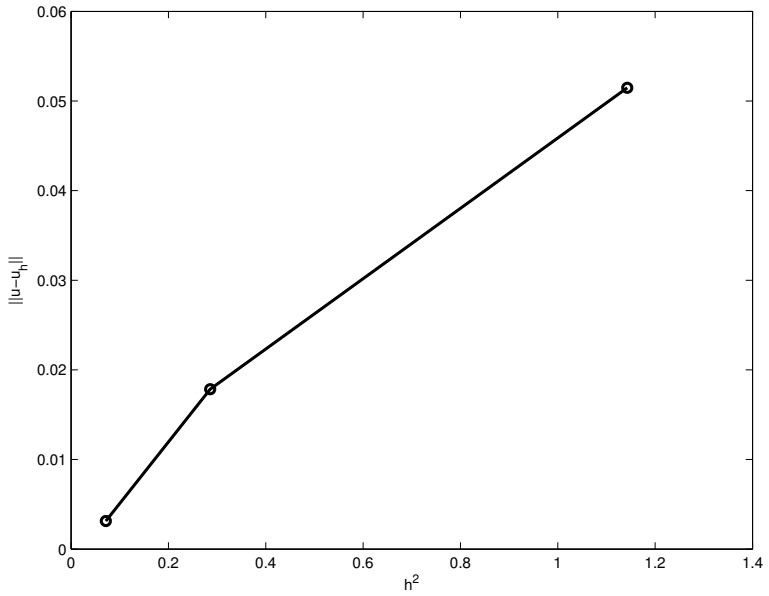

Figure 2.13: $||u - u_h||$ for Example 2.4.3, 66 angular nodes

h	$c_p = 0$	$c_p = .01$	$c_p = .1$
1.069	5.113e-02	5.171e-02	6.008e-02
.5346	1.420e-02	1.439e-02	1.761e-02
.2673	3.048e-03	3.113e-03	4.354e-03

Table 2.14: $||u - u_h||$ for Example 2.4.3, 258 angular nodes

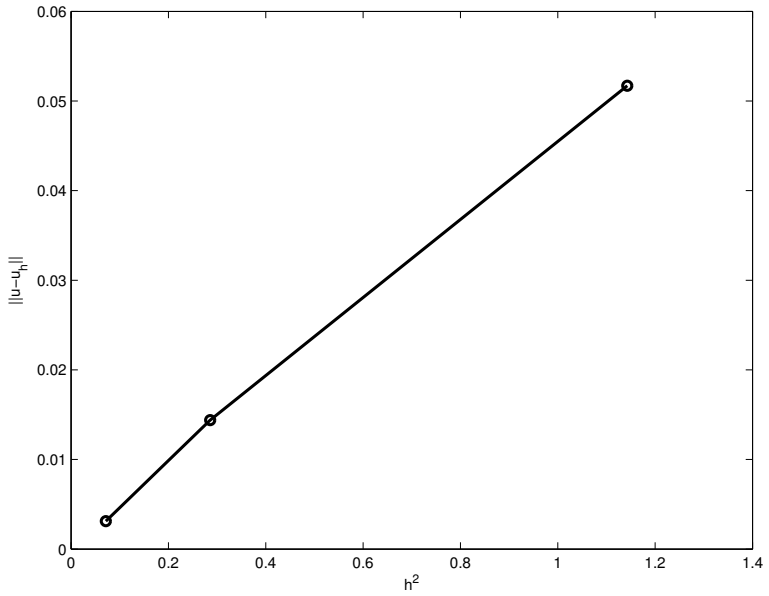

Figure 2.14: $||u - u_h||$ for Example 2.4.3, 258 angular nodes

Not surprisingly, inclusion of the stabilization terms in the bilinear form leads to improvement in solution accuracy if the penalty parameter c_p is chosen properly. In the examples, $c_p = 0.01$ and $c_p = 0.1$ are good choices. However, the inclusion of the penalty term leads to a higher degree of coupling in the discrete system of equations. As the size of the penalty term increases, the iteration method used takes more iterations to converge. Therefore, we advocate the use of the discrete-ordinate discontinuous Galerkin method with either small or no penalty parameter.

CHAPTER 3
GENERALIZED FOKKER-PLANCK EQUATION

3.1 The generalized Fokker-Planck equation

Certain applications, for example inverse problems related to medical imaging, may require the RTE to be solved several times in order to arrive at a solution. The RTE is an integro-differential equation of five independent variables. The high dimensionality and presence of an integral term make it very difficult to solve quickly in practice. In some cases, we may be able to approximate the RTE with an equation that is simpler to solve and is appropriate for the problem domain. In this section, we consider one approximation to the RTE that is valid in the case that the phase function g is highly forward peaked. This approximation is called the generalized Fokker-Planck equation (GFPE).

In tissue scattering is highly forward peaked, i.e. after scattering photons are likely to be traveling in a direction similar to their direction of travel prior to scattering ([36]). For the Henyey-Greenstein phase function this means that η is near unity. In this circumstance a simplification to the RTE can be introduced through Taylor's expansion of $u(\boldsymbol{x}, \hat{\boldsymbol{\omega}})$ for $\hat{\boldsymbol{\omega}}$ about $\boldsymbol{\omega}$ ([36]). The result is the Fokker-Planck equation

$$\forall(\boldsymbol{x}, \boldsymbol{\omega}) \in X \times \Omega,$$

$$\boldsymbol{\omega}\cdot\boldsymbol{\nabla} u(\boldsymbol{x}, \boldsymbol{\omega}) + \mu_a(\boldsymbol{x})\, u(\boldsymbol{x}, \boldsymbol{\omega}) = \frac{\mu_{tr}(\boldsymbol{x})}{2}\, \Delta^* u(\boldsymbol{x}, \boldsymbol{\omega}) + f(\boldsymbol{x}, \boldsymbol{\omega}). \tag{3.1}$$

Here Δ^* is the Laplace-Beltrami operator (1.3). $\mu_{tr} = \mu_s(1-\eta)$ is called the transport

cross section and the degree of anisotropy η is defined by

$$\eta(\boldsymbol{x}) = 2\,\pi \int_{-1}^{1} t\, g(\boldsymbol{x}, t)\, dt.$$

The eigenvalues of the Laplace-Beltrami operator are $-n(n+1)$ and the associated eigenfunctions may be taken as spherical harmonics $Y_{n,m}$, $0 \leq n < \infty$, $1 \leq m \leq 2n+1$ from Section 1.3 (cf. [36]). The eigenvalues for integral operator S with g taken to be the Henyey-Greenstein phase function are η^n and the associated eigenfunctions are also $Y_{n,m}$. We see that although the two operators have the same eigenfunctions, the integral operator S is bounded whereas the differential operator on the right side of (3.1) is unbounded. To address this issue, Leakeas and Larsen [38] propose to approximate the RTE by

$$\forall(\boldsymbol{x}, \boldsymbol{\omega}) \in X \times \Omega,$$

$$\boldsymbol{\omega}\cdot\boldsymbol{\nabla} u(\boldsymbol{x}, \boldsymbol{\omega}) + \mu_a(\boldsymbol{x})\, u(\boldsymbol{x}, \boldsymbol{\omega}) = \mu_s(\boldsymbol{x})\, \beta\Delta^*(I - \alpha\,\Delta^*)^{-1} u(\boldsymbol{x}, \boldsymbol{\omega}) + f(\boldsymbol{x}, \boldsymbol{\omega}). \tag{3.2}$$

The eigenvalues for the differential operator on the right-hand side of (3.2) will again be the spherical harmonics $Y_{n,m}$. The parameters α and β are chosen to make certain eigenvalues match the corresponding eigenvalues of the operator S. In this thesis we consider a generalized Fokker-Planck equation (GFPE) of the form

$$\forall(\boldsymbol{x}, \boldsymbol{\omega}) \in X \times \Omega,$$

$$\boldsymbol{\omega}\cdot\boldsymbol{\nabla} u(\boldsymbol{x}, \boldsymbol{\omega}) + \mu_t(\boldsymbol{x})\, u(\boldsymbol{x}, \boldsymbol{\omega}) = \mu_s(\boldsymbol{x})\, (I - \alpha(\boldsymbol{x})\,\Delta^*)^{-1} u(\boldsymbol{x}, \boldsymbol{\omega}) + f(\boldsymbol{x}, \boldsymbol{\omega}), \tag{3.3}$$

$$u(\boldsymbol{x}, \boldsymbol{\omega}) = u_{\text{in}}(\boldsymbol{x}, \boldsymbol{\omega}) \text{ on } \Gamma_-. \tag{3.4}$$

Here, $\alpha(\boldsymbol{x})$ is a positively-valued function of $\boldsymbol{x}$. For the Henyey-Greenstein phase function (1.5) we choose $\alpha = (1-\eta)/(2\eta)$ a constant. This choice of α makes the

eigenvalues corresponding to $n = 0$ and $n = 1$ match the eigenvalues of S, as well as the limit as $n \to \infty$.

Introduce a function

$$w(\boldsymbol{x}, \boldsymbol{\omega}) = (I - \alpha(\boldsymbol{x})\,\Delta^*)^{-1} u(\boldsymbol{x}, \boldsymbol{\omega}). \tag{3.5}$$

Then the equation (3.3) can be rewritten as

$$\forall (\boldsymbol{x}, \boldsymbol{\omega}) \in X \times \Omega,$$

$$\boldsymbol{\omega} \cdot \boldsymbol{\nabla} u(\boldsymbol{x}, \boldsymbol{\omega}) + \mu_t(\boldsymbol{x})\, u(\boldsymbol{x}, \boldsymbol{\omega}) = \mu_s(\boldsymbol{x})\, w(\boldsymbol{x}, \boldsymbol{\omega}) + f(\boldsymbol{x}, \boldsymbol{\omega}). \tag{3.6}$$

3.2 Well-posedness of the GFPE

When studying any new equation, it is important to consider its well-posedness before studying any numerical methods. In this section, we show that the GFPE problem (3.3)–(3.4) has a unique solution and the solution is stable with respect to the source function and the boundary condition as long as the assumptions (1.12)–(1.13) are satisfied. Further, we prove a positivity property which is required for the model (3.3)–(3.4) to be physically meaningful.

Throughout the rest of this section, we assume the domain X is convex in order to simplify notation. All arguments presented may be extended to generalized convexity condition introduced in Section 1.4. Recalling the notation from Section 1.4 we see that for each $\boldsymbol{\omega} \in \Omega$ and each $\boldsymbol{z} \in X_{\boldsymbol{\omega}}$, $X_{\boldsymbol{\omega},\boldsymbol{z}}$ is the line segment

$$X_{\boldsymbol{\omega},\boldsymbol{z}} = \{\boldsymbol{z} + s\boldsymbol{\omega} \mid s \in (s_-, s_+)\},$$

where $s_\pm = s_\pm(\boldsymbol{\omega}, \boldsymbol{z})$ depend on $\boldsymbol{\omega}$ and $\boldsymbol{z}$, and $\boldsymbol{x}_\pm = \boldsymbol{z} + s_\pm \boldsymbol{\omega}$ are the intersection points of the line $\{\boldsymbol{z} + s\boldsymbol{\omega} \mid s \in \mathbb{R}\}$ with ∂X.

We show well-posedness in the form of two theorems.

Theorem 3.2.1. *Under the assumptions* (1.12) *and* (1.13)*, the problem* (3.3)–(3.4) *has a unique solution* $u \in H^{1,2}(X \times \Omega)$

Proof. We begin by rewriting the GFPE (3.3)–(3.4) as a fixed-point problem. We will write $s_\pm$ instead of $s_\pm(\boldsymbol{\omega}, \boldsymbol{z})$ wherever there is no danger for confusion. The equation (3.6) may be rewritten as

$$\begin{aligned} &\frac{\partial}{\partial s} u(\boldsymbol{z} + s\boldsymbol{\omega}, \boldsymbol{\omega}) + \mu_t(\boldsymbol{z} + s\boldsymbol{\omega})\, u(\boldsymbol{z} + s\boldsymbol{\omega}, \boldsymbol{\omega}) \\ &\qquad = \mu_s(\boldsymbol{z} + s\boldsymbol{\omega})\, w(\boldsymbol{z} + s\boldsymbol{\omega}, \boldsymbol{\omega}) + f(\boldsymbol{z} + s\boldsymbol{\omega}, \boldsymbol{\omega}). \end{aligned} \tag{3.7}$$

Upon multiplication by the integrating factor

$$e^{\int_{s_-}^{s} \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\, ds}$$

we obtain

$$\begin{aligned} &\frac{\partial}{\partial s} \left(e^{\int_{s_-}^{s} \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\, ds} u(\boldsymbol{z} + s\boldsymbol{\omega}, \boldsymbol{\omega}) \right) \\ &\qquad = e^{\int_{s_-}^{s} \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\, ds} \Big(\mu_s(\boldsymbol{z} + s\boldsymbol{\omega})\, w(\boldsymbol{z} + s\boldsymbol{\omega}, \boldsymbol{\omega}) + f(\boldsymbol{z} + s\boldsymbol{\omega}, \boldsymbol{\omega}) \Big). \end{aligned} \tag{3.8}$$

Integrating this equation from s_- to s yeilds

$$\begin{aligned} &e^{\int_{s_-}^{s} \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\, ds} u(\boldsymbol{z} + s\boldsymbol{\omega}, \boldsymbol{\omega}) - u_{\text{in}}(\boldsymbol{z} + s_-\boldsymbol{\omega}, \boldsymbol{\omega}) \\ &\qquad = \int_{s_-}^{s} e^{\int_{s_-}^{t} \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\, ds} \left(\mu_s(\boldsymbol{z} + t\boldsymbol{\omega})\, w(\boldsymbol{z} + t\boldsymbol{\omega}, \boldsymbol{\omega}) + f(\boldsymbol{z} + t\boldsymbol{\omega}, \boldsymbol{\omega}) \right) dt. \end{aligned}$$

Therefore,

$$u = Au + F, \tag{3.9}$$

where

$$Au(\boldsymbol{z}+s\boldsymbol{\omega},\boldsymbol{\omega}) = \int_{s_-}^{s} e^{-\int_t^s \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\,ds}\mu_s(\boldsymbol{z}+t\boldsymbol{\omega})\,w(\boldsymbol{z}+t\boldsymbol{\omega},\boldsymbol{\omega})\,dt,$$

$$F(\boldsymbol{z}+s\boldsymbol{\omega},\boldsymbol{\omega}) = e^{-\int_{s_-}^s \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\,ds}u_{\text{in}}(\boldsymbol{z}+s_-\boldsymbol{\omega},\boldsymbol{\omega}) + \int_{s_-}^{s} e^{-\int_t^s \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\,ds}f(\boldsymbol{z}+t\boldsymbol{\omega},\boldsymbol{\omega})\,dt.$$

Reversing the above procedure, we can derive (3.7) from (3.9).

Now that the GFPE problem is converted to a fixed point problem, we show that the operator A is a contraction on $L^2(X\times\Omega)$. Notice that

$$\begin{aligned}
&\int_{s_-}^{s_+} \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\,|Au(\boldsymbol{z}+s\boldsymbol{\omega},\boldsymbol{\omega})|^2\,ds \\
&\quad= \int_{s_-}^{s_+} \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\left|\int_{s_-}^{s} e^{-\int_t^s \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\,ds}\mu_s(\boldsymbol{z}+t\boldsymbol{\omega})\,w(\boldsymbol{z}+t\boldsymbol{\omega},\boldsymbol{\omega})dt\right|^2 ds \\
&\quad\le \int_{s_-}^{s_+} \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\left(\int_{s_-}^{s} e^{-\int_t^s \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\,ds}\mu_s(\boldsymbol{z}+t\boldsymbol{\omega})\,dt\right) \\
&\qquad\cdot\left(\int_{s_-}^{s} e^{-\int_t^s \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\,ds}\mu_s(\boldsymbol{z}+t\boldsymbol{\omega})\,|w(\boldsymbol{z}+t\boldsymbol{\omega},\boldsymbol{\omega})|^2\,dt\right)ds.
\end{aligned}$$

Set

$$\kappa = \sup_{\boldsymbol{x}\in X}\frac{\mu_s(\boldsymbol{x})}{\mu_t(\boldsymbol{x})}, \tag{3.10}$$

and note that assumption (1.12) implies

$$\kappa < 1.$$

Since

$$\begin{aligned}
\int_{s_-}^{s} e^{-\int_t^s \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\,ds}\mu_s(\boldsymbol{z}+t\boldsymbol{\omega})\,dt &\le \kappa\int_{s_-}^{s} e^{-\int_t^s \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\,ds}\mu_t(\boldsymbol{z}+t\boldsymbol{\omega})\,dt \\
&= \kappa\left(1-e^{-\int_{s_-}^s \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\,ds}\right) \\
&< \kappa,
\end{aligned}$$

we have

$$\begin{aligned}
&\int_{s_-}^{s_+} \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\,|Au(\boldsymbol{z}+s\boldsymbol{\omega},\boldsymbol{\omega})|^2\,ds \\
&\qquad \le \kappa \int_{s_-}^{s_+} \mu_t(\boldsymbol{z}+s\boldsymbol{\omega}) \int_{s_-}^{s} e^{-\int_t^s \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\,ds} \mu_s(\boldsymbol{z}+t\boldsymbol{\omega})\,|w(\boldsymbol{z}+t\boldsymbol{\omega},\boldsymbol{\omega})|^2\,dt\,ds \\
&\qquad = \kappa \int_{s_-}^{s_+} \mu_s(\boldsymbol{z}+t\boldsymbol{\omega})\,|w(\boldsymbol{z}+t\boldsymbol{\omega},\boldsymbol{\omega})|^2 \left(\int_t^{s_+} e^{-\int_t^s \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\,ds} \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\,ds \right) dt.
\end{aligned}$$

Since

$$\int_t^{s_+} e^{-\int_t^s \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\,ds} \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\,ds = 1 - e^{-\int_t^{s_+} \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\,ds} < 1,$$

we obtain

$$\begin{aligned}
\int_{s_-}^{s_+} \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\,|Au(\boldsymbol{z}+s\boldsymbol{\omega},\boldsymbol{\omega})|^2\,ds &\le \kappa \int_{s_-}^{s_+} \mu_s(\boldsymbol{z}+t\boldsymbol{\omega})\,|w(\boldsymbol{z}+t\boldsymbol{\omega},\boldsymbol{\omega})|^2\,dt \\
&\le \kappa^2 \int_{s_-}^{s_+} \mu_t(\boldsymbol{z}+t\boldsymbol{\omega})\,|w(\boldsymbol{z}+t\boldsymbol{\omega},\boldsymbol{\omega})|^2\,dt.
\end{aligned}$$

Integrating the above inequality first with respect to $\boldsymbol{z} \in X_{\boldsymbol{\omega}}$ and then with respect to $\boldsymbol{\omega} \in \Omega$, we have thus proved the inequality

$$\|\mu_t^{1/2} Au\|_{L^2(X\times\Omega)} \le \kappa\, \|\mu_t^{1/2} w\|_{L^2(X\times\Omega)}. \tag{3.11}$$

Recall that by definition (3.5) the function w must satisfy

$$(I - \alpha\Delta^*)\, w = u \quad \text{in } X \times \Omega.$$

Therefore for a.e. $\boldsymbol{x} \in X$ $w(\boldsymbol{x},\cdot) \in H^1(\Omega)$ and

$$\int_\Omega (wv + \alpha\, \boldsymbol{\nabla}^* w \cdot \boldsymbol{\nabla}^* v)\, d\sigma(\boldsymbol{\omega}) = \int_\Omega u\, v\, d\sigma(\boldsymbol{\omega}) \quad \forall\, v \in H^1(\Omega). \tag{3.12}$$

Since $\alpha = \alpha(\boldsymbol{x}) > 0$, for any given $u(\boldsymbol{x},\cdot) \in L^2(\Omega)$, this problem has a unique solution $w(\boldsymbol{x},\cdot) \in H^1(\Omega)$ by the Lax-Milgram Lemma (cf. [8, p. 336], [24]). Thus, we may

take $v(\boldsymbol{\omega}) = w(\boldsymbol{x}, \boldsymbol{\omega})$ in (3.12) to see that

$$\int_\Omega \left(|w|^2 + \alpha \, |\boldsymbol{\nabla}^* w|^2 \right) \, d\sigma(\boldsymbol{\omega}) = \int_\Omega u \, w \, d\sigma(\boldsymbol{\omega}).$$

Therefore

$$\int_\Omega \left(|w|^2 + 2\alpha \, |\boldsymbol{\nabla}^* w|^2 \right) d\sigma(\boldsymbol{\omega}) \leq \int_\Omega |u|^2 d\sigma(\boldsymbol{\omega}). \tag{3.13}$$

In particular,

$$\int_\Omega |w|^2 d\sigma(\boldsymbol{\omega}) \leq \int_\Omega |u|^2 d\sigma(\boldsymbol{\omega}).$$

Thus,

$$\|\mu_t^{1/2} w\|_{L^2(X\times\Omega)} \leq \|\mu_t^{1/2} u\|_{L^2(X\times\Omega)}. \tag{3.14}$$

Combining (3.11) and (3.14), we see that the operator $A : L^2(X \times \Omega) \to L^2(X \times \Omega)$ satisfies

$$||\mu_t A u||_{L^2(X\times\Omega)} \leq \kappa ||\mu_t u||_{X\times\Omega}.$$

Thus, A is contractive with respect to the weighted norm $\|\mu_t^{1/2} v\|_{L^2(X\times\Omega)}$. Note that from (1.12) we have

$$c_0 ||v||_{L^2(X\times\Omega)} \leq ||\mu_t v||_{L^2(X\times\Omega)} \leq ||\mu_t||_{L^\infty(X\times\Omega)} ||v||_{L^2(X\times\Omega)}$$

Therefore, the weighted norm is equivalent to the standard L^2 norm and an application of the Banach fixed-point theorem (cf. [8, p. 209] or [57]) implies that (3.9) has a unique solution $u \in L^2(X \times \Omega)$. Consideration of (3.6) shows that $\boldsymbol{\omega} \cdot \boldsymbol{\nabla} u(\boldsymbol{x}, \boldsymbol{\omega}) \in L^2(X \times \Omega)$. Thus, the solution $u \in H^{1,2}(X \times \Omega)$. □

Theorem 3.2.2. *Let u be the solution to* (3.3)–(3.4)*. Under the assumptions* (1.12) *and* (1.13)*, u is Lipschitz continuous with respect to the source function f and the*

boundary condition u_{in}*: For given* $f_1, f_2 \in L^2(X \times \Omega)$ *and* $u_{\text{in},1}, u_{\text{in},2} \in L^2(\Gamma_-)$*, let* $u_1 = u(f_1, u_{\text{in},1}), u_2 = u(f_2, u_{\text{in},2}) \in H^{1,2}(X \times \Omega)$ *be the corresponding solutions of the problems*

$$\boldsymbol{\omega}\cdot\boldsymbol{\nabla} u_1 + \mu_t u_1 = \mu_s (I - \alpha\,\Delta^*)^{-1} u_1 + f_1 \quad \text{in } X \times \Omega,$$

$$u_1 = u_{\text{in},1} \quad \text{on } \Gamma_-$$

and

$$\boldsymbol{\omega}\cdot\boldsymbol{\nabla} u_2 + \mu_t u_2 = \mu_s (I - \alpha\,\Delta^*)^{-1} u_2 + f_2 \quad \text{in } X \times \Omega,$$

$$u_2 = u_{\text{in},2} \quad \text{on } \Gamma_-,$$

we have the bound

$$\|u_1 - u_2\|_{H^{1,2}(X\times\Omega)} \le c\left[\|u_{\text{in},1} - u_{\text{in},2}\|_{L^2(\Gamma_-)} + \|f_1 - f_2\|_{L^2(X\times\Omega)}\right] \tag{3.15}$$

for some constant c *depending only on* μ_t*,* μ_s*, and* X*.*

Proof. From (3.9),

$$\begin{aligned}\|\mu_t^{1/2} u\|_{L^2(X\times\Omega)} &\le \|\mu_t^{1/2} A u\|_{L^2(X\times\Omega)} + \|\mu_t^{1/2} F\|_{L^2(X\times\Omega)} \\ &\le \kappa\,\|\mu_t^{1/2} u\|_{L^2(X\times\Omega)} + c\left[\|u_{\text{in}}\|_{L^2(\Gamma_-)} + \|f\|_{L^2(X\times\Omega)}\right].\end{aligned}$$

So we have the bound

$$\|u\|_{L^2(X\times\Omega)} \le c\left[\|u_{\text{in}}\|_{L^2(\Gamma_-)} + \|f\|_{L^2(X\times\Omega)}\right]. \tag{3.16}$$

Using the equation (3.6),

$$\boldsymbol{\omega}\cdot\boldsymbol{\nabla} u = \mu_s w - \mu_t u + f.$$

Then we can improve the bound (3.16) to the following

$$\|u\|_{H^{1,2}(X\times\Omega)} \leq c\left[\|u_{\text{in}}\|_{L^2(\Gamma_-)} + \|f\|_{L^2(X\times\Omega)}\right], \tag{3.17}$$

finishing the proof. □

In order for the problem (3.3)–(3.4) to be accurate, it should have some physical significance. Here we provide one result demonstrating that the solution u is positive whenever the source f and inflow condition u_{in} are both positive. This property is required for the model to be physically meaningful.

Proposition 3.2.3. *Assume* (1.12)–(1.13). *If* $f \geq 0$ *a.e. in* $X \times \Omega$ *and* $u_{\text{in}} \geq 0$ *a.e. on* Γ_-, *then* $u \geq 0$ *a.e. in* $X \times \Omega$.

Proof. By (3.9) and (3.2) the solution u satisfies

$$u = Au + F$$

with $||\mu_t A||_{L^2(X\times\Omega)} < 1$. Thus by the geometric series theorem (cf. [8])

$$u = (I - A)^{-1}F = \sum_{j=0}^{\infty} A^j F.$$

Since F is defined as

$$F(\boldsymbol{z} + s\boldsymbol{\omega}, \boldsymbol{\omega}) = e^{-\int_{s_-}^{s} \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\,ds} u_{\text{in}}(\boldsymbol{z} + s_-\boldsymbol{\omega}, \boldsymbol{\omega}) + \int_{s_-}^{s} e^{-\int_{t}^{s} \mu_t(\boldsymbol{z}+s\boldsymbol{\omega})\,ds} f(\boldsymbol{z} + t\boldsymbol{\omega}, \boldsymbol{\omega})\,dt.$$

we see that the assumptions imply that $F \geq 0$ a.e. in $X \times \Omega$. We now show that $z \geq 0$ a.e. in $X \times \Omega$ implies $w \geq 0$ a.e. in $X \times \Omega$ where w is the solution of (3.12) with u replaced by z. Since $F \geq 0$ this this implies $A^j F \geq 0$ a.e. in $X \times \Omega$ for any $j \geq 0$. This in turn gives that $u = \sum_{j=0}^{\infty} A^j F \geq 0$.

In (3.12), take $v = w^- = \min(w, 0)$ to obtain

$$\int_\Omega (|w^-|^2 + \alpha\,|\boldsymbol{\nabla}^* w^-|^2)\,d\sigma(\boldsymbol{\omega}) = \int_\Omega z\,w^- d\sigma(\boldsymbol{\omega}) \le 0.$$

Hence, $w^- = 0$, i.e., $w \ge 0$ a.e. in $X \times \Omega$. □

3.3 An iteration method and its convergence

Approximating the solution of systems of PDE using finite element methods generally reduces to solving large linear systems of equations. Once the systems become large enough it becomes necessary to develop efficient solution methods. In this section, we study an iteration method and its convergence. Though the method is studied only at the continuous level, the same results hold for the discretized system of equations.

Let $w^{(0)}$ be an initial guess, e.g., we may take $w^{(0)} = 0$. Then, for $n = 1, 2, \cdots$, define $u^{(n)}$ and $w^{(n)}$ as follows:

$$\boldsymbol{\omega}\cdot\boldsymbol{\nabla} u^{(n)} + \mu_t u^{(n)} = \mu_s w^{(n-1)} + f \quad \text{in } X \times \Omega, \tag{3.18}$$

$$u^{(n)} = u_{\text{in}} \quad \text{on } \Gamma_-, \tag{3.19}$$

$$w^{(n)} = (I - \alpha\,\Delta^*)^{-1} u^{(n)}. \tag{3.20}$$

Define the iteration errors

$$e_u^{(n)} = u - u^{(n)}, \qquad e_w^{(n)} = w - w^{(n)},$$

and assume (1.12) and (1.13) are valid. Subtracting (3.18) from (3.3) we see that the

iteration errors must satisfy

$$\boldsymbol{\omega}\cdot\boldsymbol{\nabla} e_u^{(n)} + \mu_t e_u^{(n)} = \mu_s e_w^{(n-1)} \quad \text{in } X\times\Omega, \tag{3.21}$$

$$e_u^{(n)} = 0 \quad \text{on } \Gamma_-, \tag{3.22}$$

$$e_w^{(n)} = (I - \alpha\,\Delta^*)^{-1} e_u^{(n)}. \tag{3.23}$$

We have the following theorem regarding the size of the iteration errors.

Theorem 3.3.1. *Under the assumptions* (1.12) *and* (1.13)

$$\|\mu_t^{1/2} e_u^{(n)}\|_{L^2(X\times\Omega)} \le \kappa^n \,\|\mu_t^{1/2} e_u^{(0)}\|_{L^2(X\times\Omega)}.$$

Further,

$$\|e_u^{(n)}\|_{H^{1,2}(X\times\Omega)} \le c\kappa^n ||\mu_t^{1/2} e_u^{(0)}||_{L^2(X\times\Omega)}.$$

Proof. Similar to (3.11), we have

$$\|\mu_t^{1/2} e_u^{(n)}\|_{L^2(X\times\Omega)} \le \kappa \,\|\mu_t^{1/2} e_w^{(n-1)}\|_{L^2(X\times\Omega)}.$$

Similar to (3.14), we have

$$\|\mu_t^{1/2} e_w^{(n-1)}\|_{L^2(X\times\Omega)} \le \|\mu_t^{1/2} e_u^{(n-1)}\|_{L^2(X\times\Omega)}.$$

Thus,

$$\|\mu_t^{1/2} e_u^{(n)}\|_{L^2(X\times\Omega)} \le \kappa \,\|\mu_t^{1/2} e_u^{(n-1)}\|_{L^2(X\times\Omega)},$$

and so

$$\|\mu_t^{1/2} e_u^{(n)}\|_{L^2(X\times\Omega)} \le \kappa^n \|\mu_t^{1/2} e_u^{(0)}\|_{L^2(X\times\Omega)} \to 0 \quad \text{as } n\to\infty, \tag{3.24}$$

as desired. Additionally, we have

$$\|\mu_t^{1/2} e_w^{(n)}\|_{L^2(X\times\Omega)} \le \|\mu_t^{1/2} e_u^{(n)}\|_{L^2(X\times\Omega)} \le \kappa^n ||\mu_t^{1/2} e_u^{(0)}||_{L^2(X\times\Omega)}. \tag{3.25}$$

From (3.21),

$$\boldsymbol{\omega}\cdot\boldsymbol{\nabla} e_u^{(n)} = \mu_s e_w^{(n-1)} - \mu_t e_u^{(n)}.$$

Therefore,

$$||\boldsymbol{\omega}\cdot\nabla e_u^{(n)}||_{L^2(X\times\Omega)} \leq ||\mu_s e_w^{(n-1)}||_{L^2(X\times\Omega)} + ||\mu_t e_u^{(n)}||_{L^2(X\times\Omega)}. \tag{3.26}$$

Using the definition of κ and the fact that $\mu_t \in L^\infty(X)$ gives

$$||\boldsymbol{\omega}\cdot\nabla e_u^{(n)}||_{L^2(X\times\Omega)} \leq c\kappa^n ||\mu_t^{1/2} e_u^{(0)}||_{L^2(X\times\Omega)}. \tag{3.27}$$

Combining (3.24) and (3.27) yields the estimate

$$||e_u^{(n)}||_{H^{1,2}(X\times\Omega)} \leq c\kappa^n ||\mu_t^{1/2} e_u^{(0)}||_{L^2(X\times\Omega)}, \tag{3.28}$$

which completes the proof. □

The same convergence statement holds for the iteration method applied to the discretization of the GFPE introduced in the next section. The iteration method is used in numerical examples in Section 3.5.

3.4 Discretizations

In this section we describe a numerical method for solving the GFPE (3.3). First, consider the discretization of a general form of the forward problem,

$$\forall(\boldsymbol{x},\boldsymbol{\omega}) \in X\times\Omega,$$

$$\boldsymbol{\omega}\cdot\boldsymbol{\nabla} u(\boldsymbol{x},\boldsymbol{\omega}) + \mu_t(\boldsymbol{x})\, u(\boldsymbol{x},\boldsymbol{\omega}) = \mu_s(\boldsymbol{x})\, w(\boldsymbol{x},\boldsymbol{\omega}) + f(\boldsymbol{x},\boldsymbol{\omega}), \tag{3.29}$$

$$(I - \alpha(\boldsymbol{x})\,\Delta^*)\, w(\boldsymbol{x},\boldsymbol{\omega}) = u(\boldsymbol{x},\boldsymbol{\omega}), \tag{3.30}$$

$$u(\boldsymbol{x},\boldsymbol{\omega}) = u_{\text{in}}(\boldsymbol{x},\boldsymbol{\omega}), \quad (\boldsymbol{x},\boldsymbol{\omega}) \in \Gamma_-. \tag{3.31}$$

Choose a set of nodes $\{\boldsymbol{\omega}_l\}_{l=1}^L$ on the unit sphere Ω. Let W^h be a finite element space with a nodal basis $\{\phi_l^\omega(\boldsymbol{\omega})\}_{l=1}^L$ corresponding to the nodes $\{\boldsymbol{\omega}_l\}_{l=1}^L$. By a nodal basis we refer to the property

$$\phi_l^\omega(\boldsymbol{\omega}_m) = \delta_{lm}, \quad 1 \le l, m \le L.$$

Let $\overline{X} = \cup_{K\in\mathcal{T}^h} K$ be a finite element partition of $\overline{X}$ and let U^h be a corresponding finite element space. Denote by $\{\phi_i^x(\boldsymbol{x})\}_{i=1}^I$ a basis of U^h. The numerical solution of the problem (3.29)–(3.30) is expressed as follows:

$$u_h(\boldsymbol{x},\boldsymbol{\omega}) = \sum_{i=1}^{I}\sum_{l=1}^{L} u_h^{il}\phi_i^x(\boldsymbol{x})\,\phi_l^\omega(\boldsymbol{\omega}), \tag{3.32}$$

$$w_h(\boldsymbol{x},\boldsymbol{\omega}) = \sum_{i=1}^{I}\sum_{l=1}^{L} w_h^{il}\phi_i^x(\boldsymbol{x})\,\phi_l^\omega(\boldsymbol{\omega}), \tag{3.33}$$

For convenience, we will also use the following notation:

$$\begin{aligned} u_h^l(\boldsymbol{x}) &= \sum_{i=1}^{I} u_h^{il}\phi_i^x(\boldsymbol{x}), \quad u_h^i(\boldsymbol{\omega}) = \sum_{l=1}^{L} u_h^{il}\phi_l^\omega(\boldsymbol{\omega}),\\ w_h^l(\boldsymbol{x}) &= \sum_{i=1}^{I} w_h^{il}\phi_i^x(\boldsymbol{x}), \quad w_h^i(\boldsymbol{\omega}) = \sum_{l=1}^{L} w_h^{il}\phi_l^\omega(\boldsymbol{\omega}). \end{aligned} \tag{3.34}$$

As in the previous section, we begin by developing a weak formulation for (3.29). Denote

$$V^h = U^h \times W^h.$$

Then multiplying (3.29) by $v \in H^{1,2}(X\times\Omega)$ and integrating over an arbitrary element of the mesh K we arrive at

$$\int_K (\boldsymbol{\omega}\cdot\boldsymbol{\nabla} uv + \mu_t uv)\,dx = \int_K (\mu_s wv + fv)\,dx.$$

Integrating by parts yields

$$\int_{\partial K} \boldsymbol{\omega} \cdot \boldsymbol{\eta}_K uv d\sigma(x) - \int_K u\boldsymbol{\omega} \cdot \boldsymbol{\nabla} v dx + \int_K \mu_t uv dx = \int_K (\mu_s wv + fv)\, dx.$$

Let $\{w_i\}$ be the associated weights for any quadrature rule using the nodes $\boldsymbol{\omega}_i$. Then employing the quadrature rule to integrate the above equation over Ω yields

$$\begin{aligned}
&\sum_{l=1}^{L} w_l \left[\int_{\partial K} \boldsymbol{\omega}_l \cdot \boldsymbol{\eta}_K u^l v^l d\sigma(x) - \int_K u^l \boldsymbol{\omega} \cdot \boldsymbol{\nabla} v^l d\boldsymbol{x} + \int_K \mu_t u^l v^l dx \right] \\
&\quad = \sum_{l=1}^{L} w_l \left[\int_K \left(\mu_s w^l v^l + f^l v^l \right) dx \right].
\end{aligned}$$

Here we use $u^l(\boldsymbol{x}) = u(\boldsymbol{x}, \boldsymbol{\omega}_l)$. We then define the numerical solution $u_h \in V^h$ by requiring

$$\begin{aligned}
&\forall v_h \in V^h, \\
&\sum_{l=1}^{L} w_l \Big[\int_{\partial K} \boldsymbol{\omega}_l \cdot \boldsymbol{\eta}_K \widehat{u_h}(\boldsymbol{x}, \boldsymbol{\omega}_l) v_h(\boldsymbol{x}, \boldsymbol{\omega}_l) d\sigma(x) - \int_K u_h(\boldsymbol{x}, \boldsymbol{\omega}_l) \boldsymbol{\omega}_l \cdot \boldsymbol{\nabla} v_h(\boldsymbol{x}, \boldsymbol{\omega}_l) d\boldsymbol{x} \\
&\quad + \int_K \mu_t(\boldsymbol{x}) u_h(\boldsymbol{x}, \boldsymbol{\omega}_l) v_h(\boldsymbol{x}, \boldsymbol{\omega}_l) dx \Big] \\
&\quad = \sum_{l=1}^{L} w_l \int_K \left[\mu_s(\boldsymbol{x}) w_h(\boldsymbol{x}, \boldsymbol{\omega}_l) v_h(\boldsymbol{x}, \boldsymbol{\omega}_l) + f(\boldsymbol{x}, \boldsymbol{\omega}_l) v_h(\boldsymbol{x}, \boldsymbol{\omega}_l) \right] dx. \qquad (3.35)
\end{aligned}$$

As in the previous section $\widehat{u_h}$ is the so-called numerical flux and we may take

$$\widehat{u_h}(\boldsymbol{x}, \boldsymbol{\omega}_l) = \begin{cases} u_{\text{in}}^h(\boldsymbol{x}, \boldsymbol{\omega}_l), & \text{if } (\boldsymbol{x}, \boldsymbol{\omega}_l) \in \Gamma_-, \\ \lim\limits_{\varepsilon \to 0+} u_h(\boldsymbol{x} - \varepsilon \boldsymbol{\omega}_l), & \text{otherwise,} \end{cases}$$

where $u_{\text{in}}^h(\boldsymbol{x}, \boldsymbol{\omega}_l)$ is a finite element approximation of $u_{\text{in}}(\boldsymbol{x}, \boldsymbol{\omega}_l)$ from the finite element space V^h.

As the expressions on both sides of the equality (3.35) are linear in v_h and the

space V_h is spanned by $\{\phi_i^x(\boldsymbol{x})\phi_j^\omega(\boldsymbol{\omega})\}$ we may equivalently require

$$
\begin{aligned}
&1 \le i \le I, 1 \le j \le L,\\
&\sum_{l=1}^{L} w_l \Big[\int_{\partial K} \boldsymbol{\omega}_l \cdot \boldsymbol{\eta}_K \widehat{u_h}(\boldsymbol{x},\boldsymbol{\omega}_l)\phi_i^x(\boldsymbol{x})\phi_j^\omega(\boldsymbol{\omega}_l) d\sigma(x) - \int_K u_h(\boldsymbol{x},\boldsymbol{\omega}_l)\boldsymbol{\omega}_l \cdot \boldsymbol{\nabla}\phi_i^x(\boldsymbol{x})\phi_j^\omega(\boldsymbol{\omega}_l) dx \\
&\quad + \int_K \mu_t(\boldsymbol{x}) u_h(\boldsymbol{x},\boldsymbol{\omega}_l)\phi_i^x(\boldsymbol{x})\phi_j^\omega(\boldsymbol{\omega}_l) dx \Big] \\
&\quad = \sum_{l=1}^{L} w_l \int_K \left[\mu_s(\boldsymbol{x}) w_h(\boldsymbol{x},\boldsymbol{\omega}_l)\phi_i^x(\boldsymbol{x})\phi_j^\omega(\boldsymbol{\omega}_l) + f(\boldsymbol{x},\boldsymbol{\omega}_l)\phi_i^x(\boldsymbol{x})\phi_j^\omega(\boldsymbol{\omega}_l) \right] dx. \qquad (3.36)
\end{aligned}
$$

Consideration of condition (3.4) shows us that (3.36) is equivalent to

$$
\begin{aligned}
&1 \le i \le I, 1 \le j \le L,\\
&\int_{\partial K} \boldsymbol{\omega}_j \cdot \boldsymbol{\eta}_K \widehat{u_h}(\boldsymbol{x},\boldsymbol{\omega}_j)\phi_i^x(\boldsymbol{x}) d\sigma(x) - \int_K u_h(\boldsymbol{x},\boldsymbol{\omega}_j)\boldsymbol{\omega}_j \cdot \boldsymbol{\nabla}\phi_i^x(\boldsymbol{x}) d\boldsymbol{x} \\
&\quad + \int_K \mu_t(\boldsymbol{x}) u_h(\boldsymbol{x},\boldsymbol{\omega}_j)\phi_i^x(\boldsymbol{x}) dx \\
&\quad = \int_K \left[\mu_s(\boldsymbol{x}) w_h(\boldsymbol{x},\boldsymbol{\omega}_j)\phi_i^x(\boldsymbol{x}) + f(\boldsymbol{x},\boldsymbol{\omega}_j)\phi_i^x(\boldsymbol{x}) \right] dx. \qquad (3.37)
\end{aligned}
$$

Rewriting using the notation of (3.34) leads to the formulation

$$
\begin{aligned}
&1 \le i \le I, 1 \le j \le L,\\
&\int_{\partial K} \boldsymbol{\omega}_j \cdot \boldsymbol{\eta}_K \widehat{u_h^j}(\boldsymbol{x})\phi_i^x(\boldsymbol{x}) d\sigma(x) - \int_K u_h^j(\boldsymbol{x})\boldsymbol{\omega}_j \cdot \boldsymbol{\nabla}\phi_i^x(\boldsymbol{x}) d\boldsymbol{x} \\
&\quad + \int_K \mu_t(\boldsymbol{x}) u_h^j(\boldsymbol{x})\phi_i^x(\boldsymbol{x}) dx = \int_K \left[\mu_s(\boldsymbol{x}) w_h^j(\boldsymbol{x})\phi_i^x(\boldsymbol{x}) + f^j(\boldsymbol{x})\phi_i^x(\boldsymbol{x}) \right] dx. \qquad (3.38)
\end{aligned}
$$

A weak formulation of (3.30) is find w such that $w(\boldsymbol{x},\cdot) \in H^1(\Omega)$ for a.e. $\boldsymbol{x} \in X$ and

$$
\int_\Omega (wv + \alpha\, \boldsymbol{\nabla}^* w \cdot \boldsymbol{\nabla}^* v)\, d\sigma(\boldsymbol{\omega}) = \int_\Omega u\, v\, d\sigma(\boldsymbol{\omega}) \quad \forall\, v \in H^1(\Omega). \qquad (3.39)
$$

We require the approximation to w, $w_h \in W^h$, satisfy

$$\begin{aligned} &\forall v_h \in V^h, \\ &\int_\Omega \left(w_h v_h + \alpha \boldsymbol{\nabla}^* w_h \boldsymbol{\nabla}^* v_h\right) d\sigma(\boldsymbol{\omega}) = \int_\Omega u_h v_h d\sigma(\boldsymbol{\omega}). \end{aligned} \tag{3.40}$$

Equivalently,

$$\begin{aligned} &1 \le i \le I, 1 \le j \le L, \\ &\int_\Omega \left[\sum_{m=1}^{I} w_h^m(\boldsymbol{\omega})\phi_m^x(\boldsymbol{x})\phi_i^x(\boldsymbol{x})\phi_j^\omega(\boldsymbol{\omega}) + \alpha \sum_{m=1}^{I} \phi_m^x(\boldsymbol{x})\boldsymbol{\nabla}^* w_h^m(\boldsymbol{\omega})\phi_i^x(\boldsymbol{x})\boldsymbol{\nabla}^*\phi_j^\omega(\boldsymbol{\omega})\right] d\sigma(\boldsymbol{\omega}) \\ &\quad = \int_\Omega \sum_{m=1}^{I} u_h^m(\boldsymbol{\omega})\phi_m^x(\boldsymbol{x})\phi_i^x(\boldsymbol{x})\phi_j^\omega(\boldsymbol{\omega}) d\sigma(\boldsymbol{\omega}). \end{aligned} \tag{3.41}$$

Rearranging gives

$$\begin{aligned} &1 \le i \le I, 1 \le j \le L, \\ &\sum_{m=1}^{I} \phi_m^x(\boldsymbol{x})\phi_i^x(\boldsymbol{x}) \int_\Omega \left[w_h^m(\boldsymbol{\omega})\phi_j^\omega(\boldsymbol{\omega}) + \alpha\boldsymbol{\nabla}^* w_h^m(\boldsymbol{\omega})\boldsymbol{\nabla}^*\phi_j^\omega(\boldsymbol{\omega})\right] d\sigma(\boldsymbol{\omega}) \\ &\quad = \sum_{m=1}^{I} \phi_m^x(\boldsymbol{x})\phi_i^x(\boldsymbol{x}) \int_\Omega u_h^m(\boldsymbol{\omega})\phi_j^\omega(\boldsymbol{\omega}) d\sigma(\boldsymbol{\omega}). \end{aligned} \tag{3.42}$$

Linear independence of the ϕ_i^x gives that

$$\begin{aligned} &1 \le i \le I, 1 \le j \le L, \\ &\int_\Omega \left[w_h^i(\boldsymbol{\omega})\phi_j^\omega(\boldsymbol{\omega}) + \alpha\boldsymbol{\nabla}^* w_h^i(\boldsymbol{\omega})\boldsymbol{\nabla}^*\phi_j^\omega(\boldsymbol{\omega})\right] d\sigma(\boldsymbol{\omega}) \\ &\quad = \int_\Omega u_h^i(\boldsymbol{\omega})\phi_j^\omega(\boldsymbol{\omega}) d\sigma(\boldsymbol{\omega}). \end{aligned} \tag{3.43}$$

Combining (3.38) and (3.43) gives a complete description of the discrete-ordinate discontinuous Galerkin method.

3.5 Numerical examples

In this section we present some preliminary numerical examples on solving the GFPE. The purpose is to demonstrate the method described in section (3.4), the iteration method (3.3) and to form a hypothesis on the order of convergence of the discrete ordinate discontinuous Galerkin method applied to the GFPE.

First we consider solving a problem of the type (3.30):

$$w(\boldsymbol{\omega}) - \alpha\Delta^* w(\boldsymbol{\omega}) = f(\boldsymbol{\omega}), \quad \boldsymbol{\omega} \in \Omega, \tag{3.44}$$

where $\alpha \geq 0$ and $f \in L^2(\Omega)$ are given. The corresponding weak formulation is

$$w \in H^1(\Omega): \quad \int_\Omega (w\, v + \alpha\, \boldsymbol{\nabla}^* w \cdot \boldsymbol{\nabla}^* v)\, d\sigma(\boldsymbol{\omega}) = \int_\Omega f\, v\, d\sigma(\boldsymbol{\omega}) \quad \forall\, v \in H^1(\Omega). \tag{3.45}$$

We follow [4] to formulate a Galerkin method where the basis functions are piecewise linear on triangular elements of the mesh, and bilinear on rectangular elements of the mesh. The partition of the closed domain $G = [0, 2\pi] \times [0, \pi]$ for the variables ψ and θ in [4] involves two parameters: n_ψ and n_θ, where n_ψ is usually taken to be 4, 6 or 8 and n_θ is even. Initially, the domain G is divided n_ψ times in the ψ-direction and n_θ times in the θ-direction. The generated rectangles at $\theta = 0, \pi$ remain unchanged, corresponding to the triangles on the unit sphere at the poles. Define $h_\theta = \pi/n_\theta$. An edge with $\theta = k\, h_\theta \leq \pi/2$, $1 \leq k \leq n_\theta/2$, is equally split into $(k-1)$ parts. Then the generated nodes are connected properly and the generated mesh is reflected with respect to the line $\theta = \pi/2$. Figure 3 of [4] shows representative partitions of G and the corresponding isotropic triangulation of the unit sphere Ω. We refer the reader to [4] for details of this method, and give one brief example to

n_θ	4	8	16	32	64	128
e	1.465321	0.465694	0.143430	0.041513	0.010879	0.002756

Table 3.1: $L^2(\Omega)$ error for Example 3.5.1

demonstrate the correctness of implementation.

Example 3.5.1. Set $\alpha = 1$ and choose the function f so that the solution w of the problem (3.44) is

$$w(\psi, \theta) = \sin(2\pi x(\psi, \theta)) \sin(\pi y(\psi, \theta)).$$

In Table 3.1, we report the $L^2(\Omega)$ error

$$e = \left(\int_\Omega |w - w^h|^2 d\sigma(\boldsymbol{\omega}) \right)^{\frac{1}{2}}$$

for $n_\psi = 8$ and several values of n_θ. We observe a numerical convergence order of 2. The numerical solution with $n_\theta = 128$ is depicted in Figure 3.1 whereas the true solution is shown in Figure 3.2. □

The second example concerns solving the GFPE.

Example 3.5.2. This is an example of using (3.38)–(3.43) to solve the problem (3.29)–(3.30). The spatial domain is $X = (0, 1)^3$.

Let $\alpha = 1$, $\mu_t = 2$, and $\mu_s = 1$. We choose the functions f and u_{in} so that the true solution to (3.29)–(3.30) is

$$u(\boldsymbol{x}, \boldsymbol{\omega}) = -\frac{3}{2} \cos^2 \psi \left\{ \cos^2 \psi \left[1 + 7 \cos(2\theta) \right] + 8 \sin^2 \psi \right\} \cdot \sin^2 \theta \, \sin(\pi x_1) \sin(\pi x_2) \sin(\pi x_3). \tag{3.46}$$

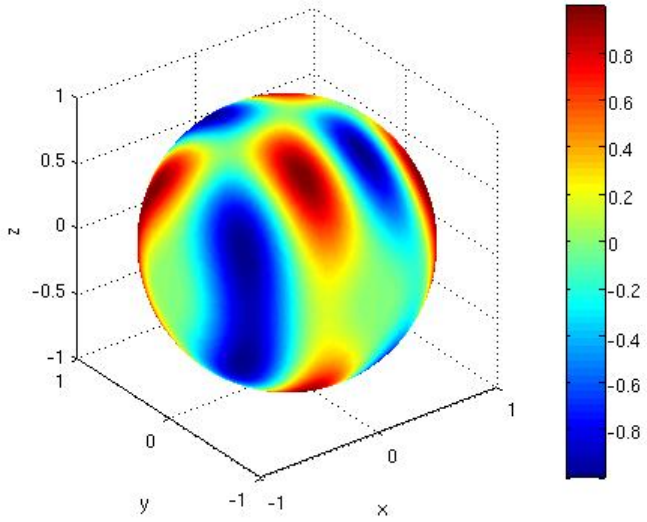

Figure 3.1: Example 3.5.1: Numerical solution with $n_\theta = 128$

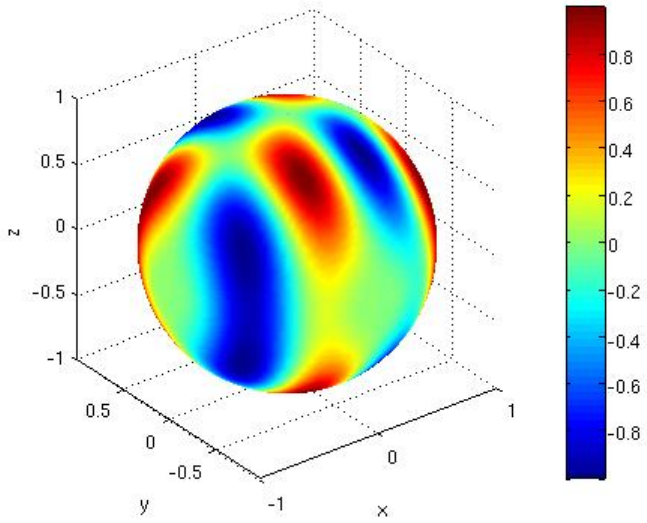

Figure 3.2: Example 3.5.1: True solution

h	$n_\theta = 4$	order	$n_\theta = 8$	order	$n_\theta = 12$	order	$n_\theta = 16$	order
$\frac{\sqrt{2}}{2}$	0.3884		0.3461		0.3432		0.3413	
$\frac{\sqrt{2}}{4}$	0.1174	1.7257	0.1003	1.7866	0.0990	1.7936	0.0983	1.7954
$\frac{\sqrt{2}}{8}$	0.0509	1.2063	0.0300	1.7427	0.0275	1.8496	0.0270	1.8666
$\frac{\sqrt{2}}{16}$	N/A	N/A	0.0150	1.0008	0.0093	1.5688	0.0078	1.7869

Table 3.2: Example 3.5.2: L^2 error for several values of h and n_θ with $n_\psi = 6$

For a positive integer n, we partition $\overline{X}$ into n^3 subcubes $\{X_i\}$, each with edge length $1/n$. Denote by S the set of all the centers and vertices of the subcubes. A mesh is generated by creating the Delaunay tesselation of the points in S. Denote by h the maximum length of the edges of the tetrahedron in the mesh; $h = \sqrt{2}/n$. The local polynomial degree $k = 1$.

Table 3.5.2 gives the error

$$e = \left\{ \frac{4\pi}{L} \sum_{l=1}^{L} ||u_h^l - u(\cdot, \boldsymbol{\omega}_l)||^2_{L^2(X)} \right\}^{\frac{1}{2}}$$

for $n_\psi = 6$ and several values of h and n_θ. When $h = \frac{\sqrt{2}}{16}$ and $n_\theta = 4$, the iteration method used fails to converge; the corresponding entry in the table is marked N/A. The column "order" is for the quantity $\text{order}_i = (\log e_i - \log e_{i-1})/(\log h_i - \log h_{i-1})$, where e_i is the error defined above corresponding to the mesh-size h_i. For small n_θ (e.g., $n_\theta = 4$), notice a deterioration in the convergence order as h decreases from $\sqrt{2}/4$ to $\sqrt{2}/8$. The deterioration phenomenon disappears for the chosen values of h when the value of n_θ is increased. For very small values of h, to achieve a numerical

h	$n_\theta = 4$	$n_\theta = 8$	$n_\theta = 16$
1.069	5.632e-02	5.507e-02	5.547e-02
.5346	2.034e-02	1.688e-02	1.694e-02
.2673	4.914e-03	3.749e-03	3.595e-03

Table 3.3: Example 3.5.3: $||u - u_h||$ for different values of h and n_θ

convergence order around the expected value 2, both n_θ and n_ψ need to be large enough.

For reference, for $n_\psi = 6$, with $n_\theta = 4$ there are 18 directional nodes, with $n_\theta = 8$ there are 98, with $n_\theta = 12$ there are 218, and with $n_\theta = 16$ there are 386. □

Example 3.5.3. In this example the domain X is an approximate cone of unit radius and height 2. The spatial meshing scheme is identical to that of Example 2.4.3. The angular meshing scheme is identical to the previous example, except that $n_\phi = 4$ rather than 6, again the parameter n_θ is varied. The source term is $f(\boldsymbol{x}, \boldsymbol{\omega}) = \chi_B(\boldsymbol{x})$, where χ_B is the indicator function of an approximate ball in the center of X with radius .5. We set $\mu_t = 2$, and $\mu_s = 1$. We use $\eta = .9$, and $\alpha = (1 - \eta)/(2\eta)$. Results are found in Table 3.3 and Figures 3.3, 3.4, and 3.5. As in the previous example, we see that when the numerical quadrature is sufficiently accurate, the numerical method has order 2 in the spatial mesh size with respect to the $L^2(X \times \Omega)$ norm.

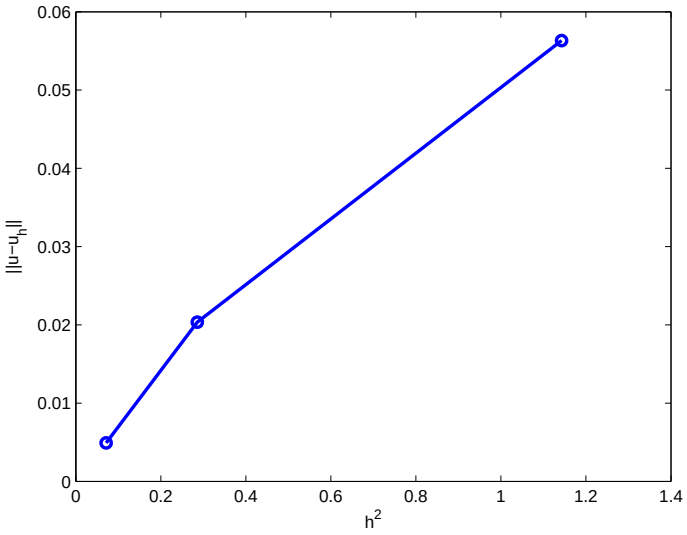

Figure 3.3: $||u - u_h||$ for Example 3.5.3, $n_\theta = 4$

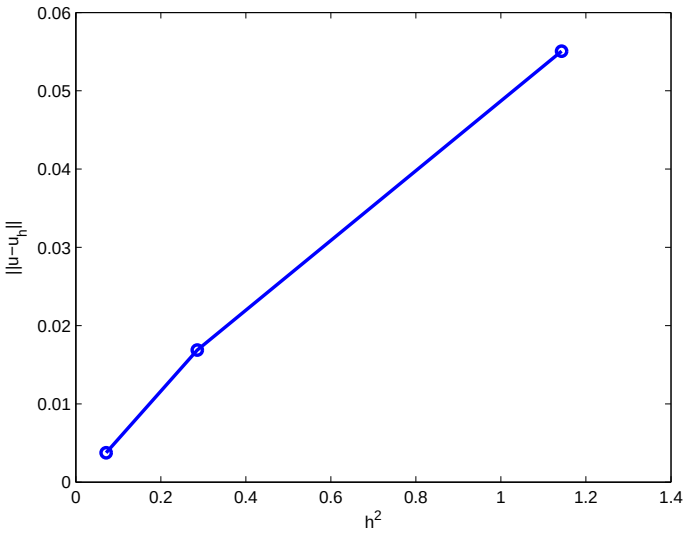

Figure 3.4: $||u - u_h||$ for Example 3.5.3, $n_\theta = 8$

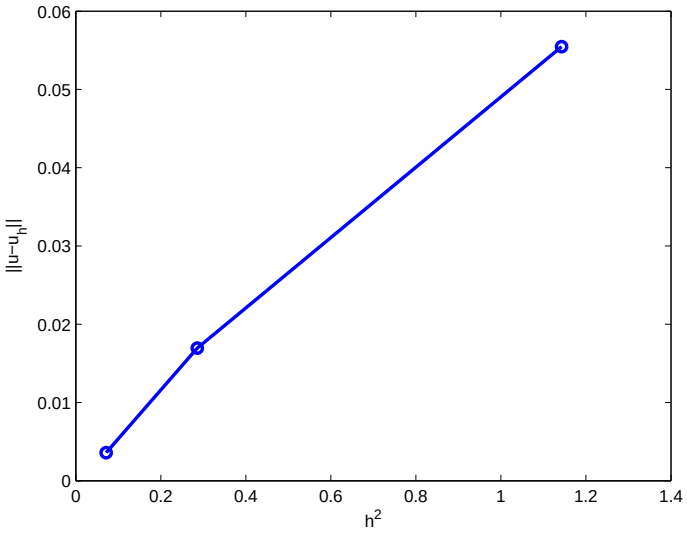

Figure 3.5: $||u - u_h||$ for Example 3.5.3, $n_\theta = 16$

CHAPTER 4
CONCLUDING REMARKS AND FURTHER WORK

In this thesis we provide a numerical method for solving the radiative transport equation and a generalized Fokker-Planck equation. Stability and error estimates are provided for the discrete-ordinate discontinuous Galerkin method introduced for the RTE. For the GFPE, well-posedness of the problem is shown and a discrete-ordinate discontinuous Galerkin method is introduced. For both problems numerical results are given demonstrating the performance of the methods.

There are several topics that warrant further investigation related to this work. In Chapter 3 we introduce an iteration method for solving the GFPE. This iteration method is based upon the source iteration method that is commonly used when solving the RTE ([1]). When applied to the discretized system of equations, both iteration methods resemble Jacobi iteration. In 2009, Gao and Zhao ([27]) introduced an alternative to source iteration that amounts to Gauss-Seidel iteration for block matrices. The natural next step in this direction is development of successive over relaxation methods for these problems. Initial numerical experiments indicate that subject to proper choice of the acceleration parameter ω the SOR scheme will converge, and will reduce the number of iterations required for convergence by up to $\%\,50$.

One of the main advantages of using a discontinuous Galerkin method is that conforming meshes are not required, i.e. hanging nodes pose no problem [18]. Additionally, it is easy to formulate DG methods in which the numerical solution has different polynomial degree on neighboring elements [18]. These properties make

adaptive mesh refinement much simpler than adaptive mesh refinement for continuous finite element methods. In order to implement an adaptive refinement algorithm, we must first develop some a posteriori error estimate. In recent years several researches (cf. [2, 32, 31]) have published results on a posteriori error estimation for hyperbolic problems. We may modify these results to suit our needs in the future.

Discontinuous Galerkin methods tend to be very good candidates for parallel implementation ([18, 20]). In particular, subject to the choice of numerical flux, it may be possible to find the numerical solution on several elements simultaneously. Such is the case for both the discrete-ordinate DG method for the RTE and for the GFPE. Initial investigation leads us to believe that the methods introduced in this thesis can be implemented with very high parallel efficiency even on hundreds or thousands of processors. With this in mind, it may be profitable to investigate so-called GPU computing for the parallel implementation of the discrete-ordinate DG methods presented in this thesis.

The work presented here on efficient solution of the RTE is part of a larger work for solving the inverse problem with applications in biomedical imaging. Briefly, the inverse problem may be to reconstruct the source function, scattering parameter, absorption parameter, or some combination of these given data from experiment (cf. [30, 53]). Upon discretization, this problem is generally reduced to a minimization problem over a high dimensional space whose objective function involves solving the forward RTE with fixed parameters. Although the cost of evaluating the objective function is high, the most expensive part of naive algorithms is gradient evaluation

and/or finding a descent direction. Typically, an algorithm may require one objective function evaluation for each entry in the gradient; this quickly becomes quite restrictive. In 1998, Dorn (cf. [21, 22]) introduced the so-called transport-backtransport method, in which a descent direction for the objective function may be found by a single solution of an adjoint problem. In the future, we may develop efficient numerical methods for solving the adjoint problem and apply the results to solving the inverse problem.

REFERENCES

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

[2] S. Adjerid, K. Devine, J. Flaherty, and L. Krivodonova. A posteriori error estimation for discontinuous Galerkin solutions of hyperbolic problems. *Computer Methods in Applied Mechanics and Engineering*, 191:1097–1112, 2002.

[3] A. Agoshkov. *Boundary Value Problems for Transport Equations*. Birkhäuser, Boston, 1998.

[4] T. Apel and C. Pester. Clement-type interpolation on spherical domains—interpolation error estimates and application to a posteriori error estimation. *IMA J. Numer. Anal.*, 25:310–336, 2005.

[5] D. N. Arnold, F. Brezzi, B. Cockburn, and L. D. Marini. Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM J. Numer. Anal.*, 39:1749–1779, 2002.

[6] M. Asadzadeh and A. Kadem. Chebyshev spectral–S_N method for the neutron transport equation. *Comput. Math. Appl.*, 52:509–524, 2006.

[7] K. Atkinson. Numerical integration on the sphere. *J. Austral. Math. Soc. Series B*, 23:332–347, 1982.

[8] K. Atkinson and W. Han. *Theoretical Numerical Analysis: A Functional Analysis Framework*. Springer, New York, third edition, 2009.

[9] G. Avtandilov, A. Dembo, O. Komardin, P. Lazarev, M. Paukshto, L. Shkolnik, and O. Zayratiyants. Human tissue analysis by small-angle x-ray scattering. *Journal of Applied Crystallography*, 33:511–514, 2000.

[10] G. Bal and A. Tamasan. Inverse source problems in transport equations. *SIAM J. Math. Anal.*, 39:57–76, 2007.

[11] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*. Springer, New York, third edition, 2008.

[12] F. Brezzi, B. Cockburn, L.D. Marini, and E. Süli. Stabilization mechanisms in discontinuous Galerkin finite element methods. *Comput. Methods Appl. Mech. Engrg.*, 195:3293–3310, 2006.

[13] F. Brezzi, L. D. Marini, , and E. Süli. Discontinuous Galerkin methods for first-order hyperbolic problems. *Math. Models Methods Appl. Sci.*, 14:1893–1903, 2004.

[14] B. G. Carlson and K. D. Lathrop. Transport theory—the method of discrete ordinates. In H. Greenspan, C. N. Kelber, and D. Okrent, editors, *Computing Methods in Reactor Physics*, pages 171–266. Gordon and Breach Science Publishers, New York, 1968.

[15] B. Chang, T. Manteuffel, S. McCormick, J. Ruge, and B. Sheehan. Spatial multigrid for isotropic neutron transport. *SIAM J. Sci. Comput.*, 29:1900–1917, 2007.

[16] D. Chapman, W. Thomlinson, R. E. Johnston, D. Washburn E. Pisano, N. Gmür, Z. Zhong, R. Menk, F. Arfelli, and D. Sayers. Diffraction enhanced x-ray imaging. *Physics in Medicine and Biology*, 42:2015–2025, 1997.

[17] P. G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, Amsterdam, 1978.

[18] B. Cockburn. Discontinuous Galerkin methods. *ZAMM Z. Angew. Math. Mech.*, 83:731–754, 2003.

[19] B. Cockburn. Discontinuous Galerkin methods for computational fluid dynamics. In E. Stein, R. de Borst, and T. J. R. Hughes, editors, *Encyclopedia of Computational Mechanics*, volume 3, pages 91–127. John Wiley & Sons, 2004.

[20] B. Cockburn, G. E. Karniadakis, and C.W. Shu, editors. *Discontinuous Galerkin methods: Theory, computation and applications*, volume 11 of *Lecture Notes in Computational Science in Engineering*. Springer, 2000.

[21] O. Dorn. A transport-backtransport method for optical tomography. *Inverse Probl.*, 14:1107–1130, 1998.

[22] O. Dorn. Scattering and absorption transport sensitivity functions for optical tomography. *Opt. Express*, 13:492–506, 2000.

[23] P. Edström. A fast and stable solution method for the radiative transfer problem. *SIAM Review*, 47:447–468, 2005.

[24] L. C. Evans. *Partial Differential Equations*. American Mathematical Society, 1998.

[25] M. Fernández, J. Keyriläinen, R. Serimaa, M. Torkkeli, M.-L. Karjalainen-Lindsberg, M. Leidenius, K. von Smitten, M. Tenhunen, S. Fiedler, A. Bravin, T. M. Weiss, and P. Suortti. Human breast cancer in vitro: matching histopathology with small-angle x-ray scattering and diffraction enhanced x-ray imaging. *Physics in Medicine and Biology*, 50:2991–3006, 2005.

[26] W. Freeden, T. Gervens, and M. Schreiner. *Constructive Approximation on the Sphere with Applications to Geomathematics*. Oxford University Press, Oxford, 1998.

[27] H. Gao and H. Zhao. A fast forward solver of radiative transfer equation. *Transport Theory and Statistical Physics*, 38:149–192, 2009.

[28] S. J. Glick. Breast CT. *Annual Review of Biomedical Engineering*, 9:501–526, 2007.

[29] S. J. Glick, S. Thacker, X. Gong, and B. Liu. Evaluating the impact of x-rays spectral shape on image quality in flat-panel CT breast imaging. *Medical Physics*, 34:5–24, 2007.

[30] W. Han, J. A. Eichholz, J. Huang, and J. Lu. RTE based bioluminescence tomography: a theoretical study. *Inverse Problems in Science & Engineering*, in press.

[31] R. Hartmann and P. Houston. Adaptive discontinuous Galerkin finite element methods for nonlinear hyperbolic conservation laws. *SIAM J. Sci. Comput.*, 24:979–1004, 2002.

[32] R. Hartmann and P. Houston. Adaptive discontinuous Galerkin finite element methods for the compressible Euler equations. *Journal of Computational Physics*, 182:508–532, 2002.

[33] L. G. Henyey and J. L. Greenstein. Diffuse radiation in the Galaxy. *Astrophys. J*, 93:70–83, January 1941.

[34] K. Hesse and I. H. Sloan. Cubature over the sphere S^2 in Sobolev spaces of arbitrary order. *J. Approx. Theory*, 141:118–133, 2006.

[35] K. Kerlikowske, D. Grady, S. M. Rubin, C. Sandrock, and V. L. Ernster. Efficacy of screening mammography—a meta analysis. *JAMA-Journal of the American Medical Association*, pages 149–154, 1995.

[36] A. D. Kim and J. B. Keller. Light propagation in biological tissue. *J. Opt. Soc. Am. A*, 20:92–98, 2003.

[37] A. D. Kim and M. Moscoso. Chebyshev spectral methods for radiative transfer. *SIAM J. Sci. Comput.*, 23:2074–2094, 2002.

[38] C. L. Leakeas and E. W. Larsen. Generalized Fokker-Planck approximations of particle transport with highly forward-peaked scattering. *Nucl. Sci. Eng.*, 137:236–250, 2001.

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

[40] E. E. Lewis and W. F. Miller. *Computational Methods of Neutron Transport*. John Wiley & Sons, New York, 1984.

[41] R. A. Lewis, K. D. Rogers, C. J. Hall, E. Towns-Andrews E, S. Slawson, A. Evans, S. E. Pinder, I. O. Ellis, C. R. Boggis, A. P. Hufton, and D. R. Dance. Breast cancer diagnosis using scattered x-rays. *Journal of Synchrotron Radiation*, 7:348–352, 2000.

[42] E. Machorro. Discontinuous Galerkin finite element method applied to the 1-d spherical neutron transport equation. *J. Comput. Phys.*, 223:67–81, 2007.

[43] T. MacRobert. *Spherical Harmonics*. Pergamon Press, third edition, 1967.

[44] M. F. Modest. *Radiative Heat Transfer*. Academic Press, second edition, 2003.

[45] A. Momose, T. Takeda, Y. Itai, and K. Hirano. Phase-contrast x-ray computed tomography for observing biological soft tissues. *Nature Medicine*, 2:473–475, 1996.

[46] M. Muttarak, S. Pojchamarnwiputh, and B. Chaiwun. Breast cancer in women under 40 years: Preoperative detection by mammography. *Annals Academy of Medicine Singapore*, 32:433–437, 2003.

[47] F. Natterer and F. Wübbeling. *Mathematical Methods in Image Reconstruction*. SIAM, Philadelphia, 2001.

[48] F. Pfeiffer, M. Bech, O. Bunk, P. Kraft, E. F. Eikenberry, Ch. Brönnimann, C. Grünzweig, and C. David. Hard-x-ray dark-field imaging using a grating interferometer. *Nature Materials*, 7:134–137, 2008.

[49] F. Pfeiffer, O. Bunk, C. David, M. Bech, G. Le Duc, A. Bravin, and P. Cloetens. High-resolution brain tumor visualization using three-dimensional x-ray phase contrast tomography. *Physics in Medicine and Biology*, 52:6923–6930, 2007.

[50] E. D. Pisano, C. Gatsonis, E. Hendrick, M. Yaffe, J. K. Baum, S. Acharyya, E. F. Conant, L. L. Fajardo, L. Bassett, C. D'Orsi, R. Jong, and M. Rebner. Diagnostic performance of digital versus film mammography for breast-cancer screening. *New England Journal of Medicine*, 353:1773–1783, 2005.

[51] I. Pucci-Minafra, C. Luparello, M. Andrioloand L. Basiricò, A. Aquino, and S. Minafra. A new form of tumour and fetal collagen that binds laminin. *Biochemistry*, 32:7421–7427, 1993.

[52] W. H. Reed and T. R. Hill. Triangular mesh methods for the neutron transport equation. Technical Report LA-UR-73-479, Los Alamos National Laboratory, 1973.

[53] K. Ren, G. Bal, and A. Heilscher. Frequency domain optical tomography based on the equation of radiative transfer. *SIAM J. Sci. Comput.*, 28:1463–1489, 2006.

[54] A. H. Stroud. *Approximate Calculation of Multiple Integrals*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1971.

[55] G. E. Thomas and K. Stamnes. *Radiative Transfer in the Atmosphere and Ocean*. Cambridge University Press, 1999.

[56] W. Zdunkowski, T. Trautmann, and A. Bott. *Radiation in the Atmosphere: A Course in Theoretical Meteorology*. Cambridge University Press, 2007.

[57] E. Zeidler. *Nonlinear Functional Analysis and its Applications. I: Fixed-point Theorems*. Springer-Verlag, New York, 1985.